Statistical Methods for Scientists and Engineers

Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology – Kharagpur

Lecture - 28
Non-parametric Methods – I

Friends, now we will be starting the new topic called non-parametric methods. Now if you remember my terminology in the previous lectures for example when we were discussing parametric methods we always started with the statement like saying let X1, X2, Xn be a random sample from a probability from a population with probability distribution and we called it P theta.

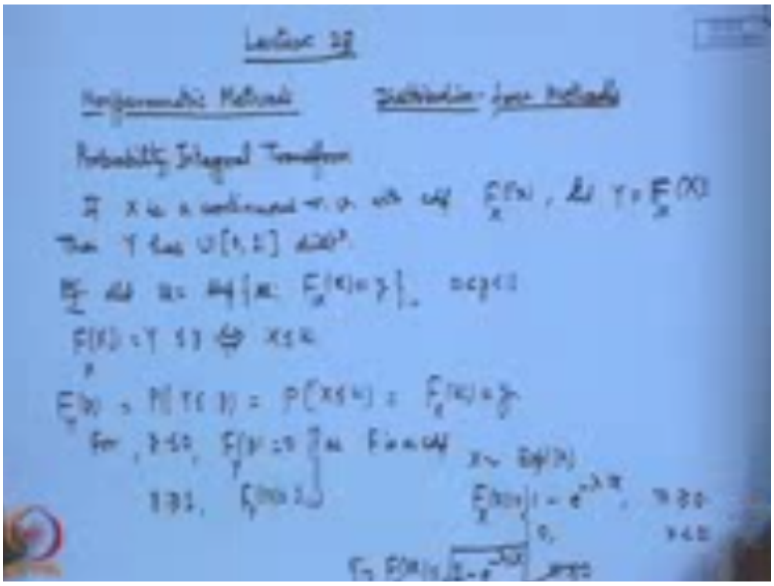
That means we knew that the distribution is of the forum of a normal distribution with parameters of square, a Poisson distribution with parameter lambda with binomial distribution with parameters in NP or a gamma distribution with parameters say RM lambda. Pareto distribution, Weibull distribution so these are all called parametric model. Because basically what we are saying is that we are able to pin point the probability model for our phenomenon or the population under its study.

But and therefore the method that we are derived they were very specific to that for example unbiased estimation, minimum variance and by estimation we had the concept of Mean squared error of the estimators, we had the concept of admissible minimax estimators. We had also the testing problems in which we conceded a parametric model say normal distribution so we are testing with the value = 0 or = 1 or the theta is <= theta 0 or theta > theta 0, etc.

So, all the procedures that we are developed where under the assumption that we are having certain distribution for our population under this. But many times it happens that either the data is insufficient for fitting of a distribution or it is too volatile to actually fit a distribution. In that case we may need the methods which are under the assumption that only we may have some general assumption like a continuous distribution, etc.

And its specific form is not assumed. So such methods they are called distribution free methods or the methods of non-parametric statistics.

(Refer Slide Time: 02:29)



So distribution free. So basically what happens that whatever methods we derived they will be free from the distributional assumption model, distribution model assumption. For example, we will not say that it is a normal distribution or it is a Poisson distribution or it is a binomial distribution, etc. In order, these methods are also quite old after, right after the Fisherian era that means 1930's Neyman-Pearson theory.

Abraham Wald came and through his efforts mainly the non-parametric methods Wald–Wolfowitz, etc. they started developing and especially there was a Wald on the non-parametric methods. So these methods are quite old and in 1960's by Hajek and there are other people who develop these methods So we will be coming across and then also Kolmogorov-Smirnov they develop the powerful method which was also free from the distribution assumptions.

So because of that these methods are gain popularity. Now with the advent of computer oriented procedures these methods are easy to apply. One thing one can understand that since we are having less assumption about the model of the population that in general the methods will be slightly less powerful than the methods when we have the information on the type of the distribution.

But that is expected because if we have more information assume then your method should be more powerful. The primary building block of non-parametric methods is the following results that is called probability integral transform. Briefly, I mentioned about this thing in my distribution theory that means when we were discussing the distribution of a function of random variable.

So this particular result is actually a building block of the methods which are developed for the non-parametric statistics. So I will state the following result. If X is a continuous random variable with some cdf say F of x. Let us define Y = F, x of X that means in place of this small x I replace by capital X. So this becomes a function of the random variable because see it is like this.

Let me take an example, suppose I am considering say x following exponential distribution with parameter lambda. Then what is your capital effects? That is 1 – e to the power – lambda x for x >= to 0 it is = 0 for x < 0. In these if I define Y = Fx then that is = 1-e to the power – lambda x basically for x >= 0 which has true of course because x is a positive random variable so this will be this function basically.

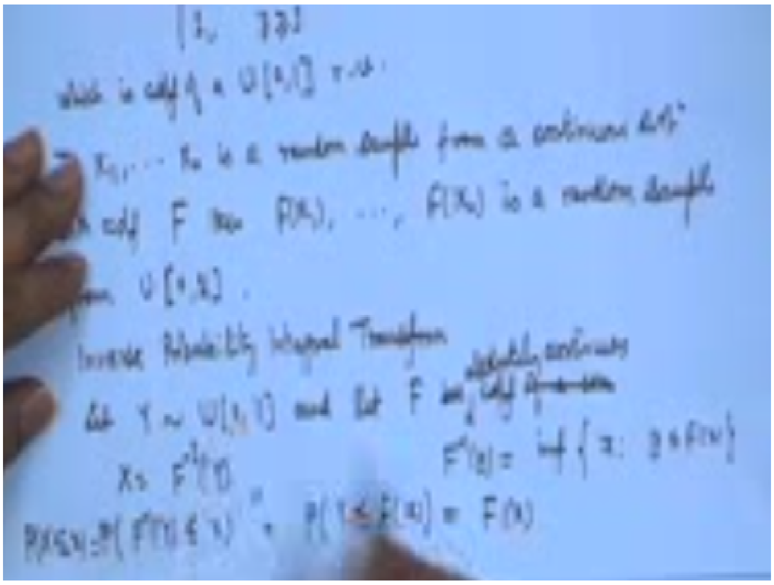
So then if I look at the distribution of that then Y has a uniform distribution on the interval 0 to 1. Now this is a very strong or you can say powerful result but because what it says that from any distribution I can achieve at least in the continuous case I can go to the uniform distribution and together with this we will give inverse of this it becomes a very powerful tool for the simulation. Let me look at the proof of this.

Let us define say, u = Fx of x such that Fx = Fx X = y for 0 < y < 1. So if I consider F of x = y <= to y, then this is equivalent to saying that x is <= u. So if I consider say Fy y that is = probability of Y <= y that is same as probability of x <= to u that is = Fx of u. Fx of u is nothing but y. So what we are getting that if I am considering a y a point between 0 to 1 then capital Y <= y is equivalent to x <= u.

Because u is the supreme of this set of values for which Fx = Y. So probability of Y <= y is same as probability of x < or equal to u so that is Fx u and that is = Y. So we have proved that for y between 0 to 1 Fy = Y and of course if I consider say Y to be <= 0 suddenly F (y) will be 0 why? because f is a cdf. So cdf takes values between 0 to 1 and for Y >= 1 Fy y that will be = 1.

So these 2 statements are valid because F is a cdf. All the values of Y Fy are between 0 to 1 only. So what we have proved? So we have proved that Fy, let me write it.

(Refer Slide Time: 08:46)



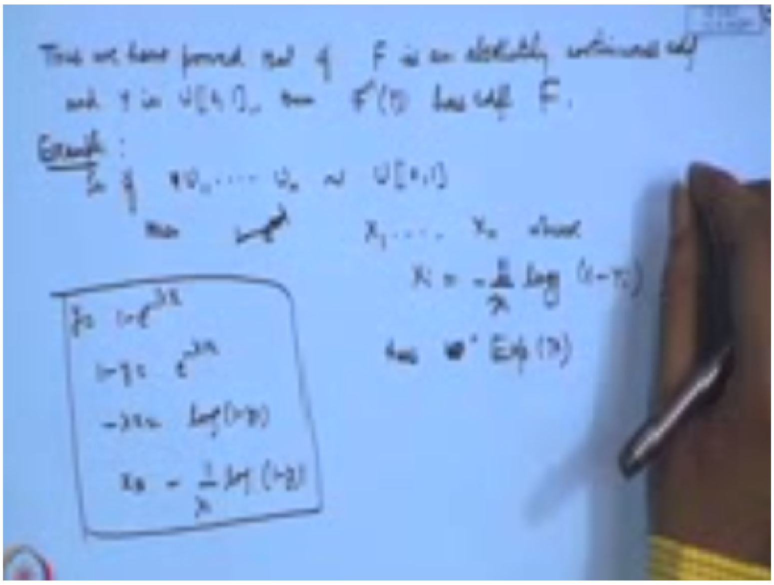
Thus we have shown that Fy = 0 for Y <= to 0 it is = Y, for 0 < Y < 1 it is = 1 for Y >= to 1. Which is cdf of a uniform 01 random number. So we have proved that if X is a continuous random variable with cdf Fx then y has a that means Fx in place of small x I replace by capital X so it becomes a function, so this random variable will have a uniform distribution.

So that means if I have a random sample if X1, X2, Xn is a random sample from a continuous distribution with cdf say capital F, then F of X1, F of X2, F of Xn is a random sample from uniform 01. So this is a very powerful result and in fact many of the methodologies of non-parametric statistics will be based on this result. Of course we are interested in the whether the converse of this result is also true.

That we consider inwards probability integral transform. So, let us look at say Y following uniform 01 and let F be cdf of a so basically we assume it to be absolutely continuous. Let F be absolutely continuous cdf, okay. Then define say X = F inverse Y. Let us look at, so how do you define inverse? F inverse of a function is defined as in all of the set of all access for which Y is <= Fx.

So if I consider probability of F inverse Y <= say x that is probability of x <= x. So that is = probability of Y <= Fx. But what is this thing? Y is uniform 01so this has to be simply F (x). So what we are saying? Cdf of capital x is Fx. So what we are proving is that if Y has this, so let me write this result here in the form of a theorem.

(Refer Slide Time: 12:30)



Proved that if F is an absolutely continuous cdf and y is uniform 01 then F inverse Y has cdf capital F. These 2 results taken together are also the, you can say building blocks of the simulation procedures. Because in simulation we have to generate random samples from some given distribution with cdf capital F.

So what we do the procedures have been develop to generate pseudo- random numbers basically that is the uniformly distributed numbers on from 1 to some upper mount which is defined by the upper limit of the largest one in a computer program. What we do? We divide by that upper mount so you get uniformly distributed random numbers between 0 to 1. Now F is then own distribution then you can consider F inverse of that.

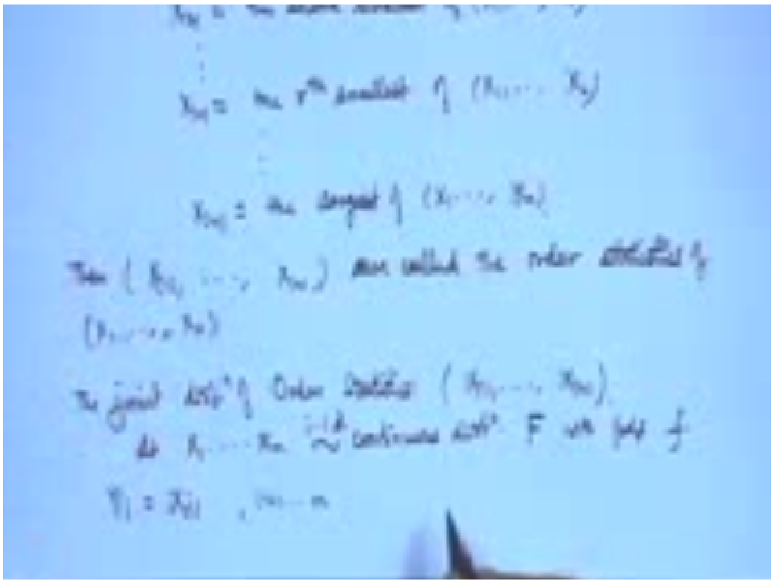
Then we will get those random sample from that particular distribution. Let us consider say this example of exponential distribution. So if we consider, so let me consider this as an example if say U1, U2, Un is a random sample from uniform 01 then 1 – e to the power – lambda. So you calculate the inverse of this actually if I write the Y = 1 – e to the power – lambda X this is F. So 1 – Y = e to the power – lambda x.

So – lambda x = log off 1 – y. So x = - 1 by lambda log off 1 – Y. So this is the inverse of this. So then if I consider the random variables X1, X2, Xn where Xi = 1 by lambda log of 1 – Yi. Then this has exponential lambda distribution. So this method helps us to generate random samples from various distributions at least for them for which this F inverse in a closed form.

If F inverse is not in a closed form, then of course the method becomes difficult. For example, if we look at the normal distribution but then we look at some other transformation because many distributions are related to each other and for example normal distribution can be generated through some other transformation so there are methods which are available for that. May be in one of the classes I will briefly touch up on the simulation part also.

Now let us, the next building blocks of the non-parametric methods they are called order statistics.

(Refer Slide Time: 16:13)



These order statistics also I discussed earlier in distribution theory. As you know that if I consider let X1, X2, Xn be a random sample. We define X1 to be the smallest of X1, X2, Xn. Then X2 we define to be the second smallest of X1, X2, Xn and in a similar way Xr is the rth smallest of X1, X2, Xn and so on and finally X1 is the largest of X1, X2, Xn. Then we call this X1, X2, Xn as the order statistics of X1, X2, Xn.

If we assume that random variables are continuous then the probability of any of the X1 being X2 being = will be 0. In that case we can study the distribution theory or the properties of the distribution of this. Let me firstly consider the joint distribution of X1, X2, Xn. If you remember in the part when I was discussing the distribution theory and we discuss the distribution the joint distribution of the functions of random variable.

We have considered this distribution. So that means if I consider say let me consider the general case the joint distribution of order statistics. Before going to the derivation let me also talk about the use of this order statistic from a practical point of view because I am giving this course for the scientist and engineers who are, who may not be knowing exactly the statistician they may be simply using these methods.

So what is the use of this? So many a times it happens that we are not interested in observations as they are given to us or as they arise. Rather we may be interested in a particular form of that R you can saying ordering of that. Let us look at some very straight forward examples. Suppose there is a testing of the strength of certain material or certain brand of certain instrument or something like that.

If we are considering certain brand of a certain instrument, then we may like to use that one which has the largest life time. So suppose the life times have been recorded then I am not interested in each of them rather I am interested only in the largest one. In a similar way suppose there is selection of the best candidate then I may be interested in say for example I have 2 positions.

So, out of 10 candidates who appear the scores are given for those 10 people. I may be looking at X 9 with a bracket and X 10 rather than looking at all of them I may be interested only in the best 2 that is the largest 2. Similarly, in some other cases I may be interested in the minimum. For example, an item with the lowest price or the 3 items with the lowest price in a set of say 10 items.

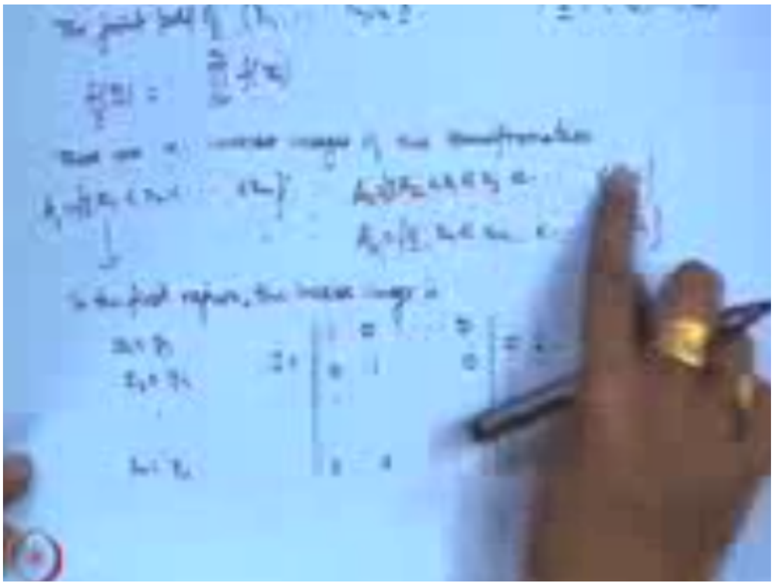
So, if you consider for example certain sports events say gymnastics or say figure skating and so on where scores are given by the judges. Then we look at the total scores and then we look at we rank them according to the largest to the lowest. And then the top 2, top 4 like that top 3 they are considered actually. That means these order statistics are always indirectly playing role in the real life in whatever physical application we are actually making use of them.

Therefore, in non-parametric methods when we are having the parametric methods then certainly we go by the averages and so on and so the distribution theory is nice. But when we have we do not have too much knowledge about the form of the distribution then we use the order statistics. So, studying the distribution of order statistics is one of the primer you can say our primary concern in the non-parametric methods.

So, the joint distribution of order statistics let us consider X1, X 2, Xn. We use the notations say let X1, X2, Xn be a random sample from a continuous distribution, okay. So, F, so this is a random sample that means I am at this stage assuming this to be independent and identically distributed and pdf then let us assume say small f. And we use this notation for that. We want the joint distribution of Y1, Y2, Yn, okay.

We know that in order to obtain this we have to first write down the joint pdf of X1, X2, Xn and then we apply the transformation.

(Refer Slide Time: 22:49)

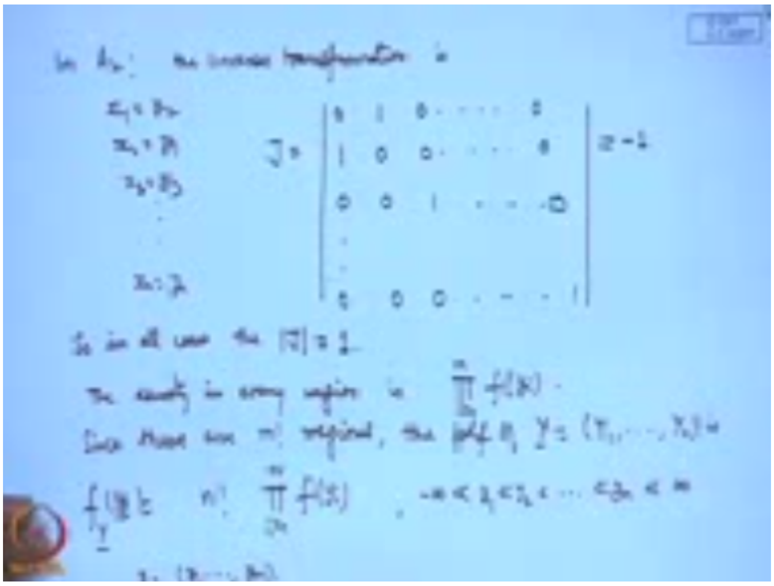


The joint distribution or the joint pdf of X1, X2, Xn, so that is simply written as product of F x i. i = 1 to n, okay. You can also write it as f(x) where x is this vector and small x is x1, x2, xn. Now this transformation when we make this transformation y i = x i, we have to write down the inverse transformation and the Jacobean. So, there are n factorial inverse images of this transformation.

For example, you may have X1 < X2 < Xn you may have x to < X1 < X3 < Xn and so on. You may have xn < xn – 1 < x 1. So, if we call these regions as say A1 that is x such that this happens A2 is x such that this is true and so on An factorial x such that this happens. Let us consider the Jacobean in one case. So the inverse transformation in the first region the inverse image is X1 = Y1 X2 = Y2 Xn = Yn.

So, Jacobean of the transformation that is del X1 by del Y1 that is 1 del X1 by del Y2 is 0 and so on. Del X2 y del Y1 0 del X2 Y2 del Y2 1and so on which is nothing but a determinant of the identity matrix that is equal to 1. Now what happens in the second region say for example A2.

(Refer Slide Time: 25:36)



In A2 the inverse image if you write then you will have X1 = Y2, X2 = Y1, X3 = Y3 and so on Xn = Yn. See, if we consider the Jacobean del X1 by del Y1 is 0 del X1 by del Y2 is 1 and then there are 0's. Here del X2 by del Y1 is 1 this = 0 0 0 0 1 which is nothing but determinant of the identity matrix with the first and second rows interchanged other rows are the same.

Or you can say first and second columns interchanged. So, this value will be equal to minus 1. Subsequently, if we consider other transformations for example if you look at this one here it is totally reversed. So, it is also obtained, so it will become 0 0 0 1. So, it will become simply all the rows and columns are changed here. Like this is going to the last one this is going to the second last this is going to the first one.

So, if that is happening then the determinant will be simply - 1 to the power n because n transformations in the initial one will be there. So, if you look at the determinant they are either + 1 or - 1 in every case. So, if I consider the absolute value it is going to be plus one. So, in all cases the absolute value of the determinant are the Jacobean is going to be plus one.

Now look at the density, the density in every region is nothing but so here it is interesting. You look at this one. Here all the terms are coming Fx 1, Fx 2, Fx n. When you consider the inverse transformation in the first region it is Fy 1, Fy 2, Fy n. In the second one it will be Fy 2, Fy 1, Fy n. In the last region it will become Fy n, FY 2, Fy n – 1 Fy 1. That means all the time the n terms are coming but they may be in any order ultimately it is the product.

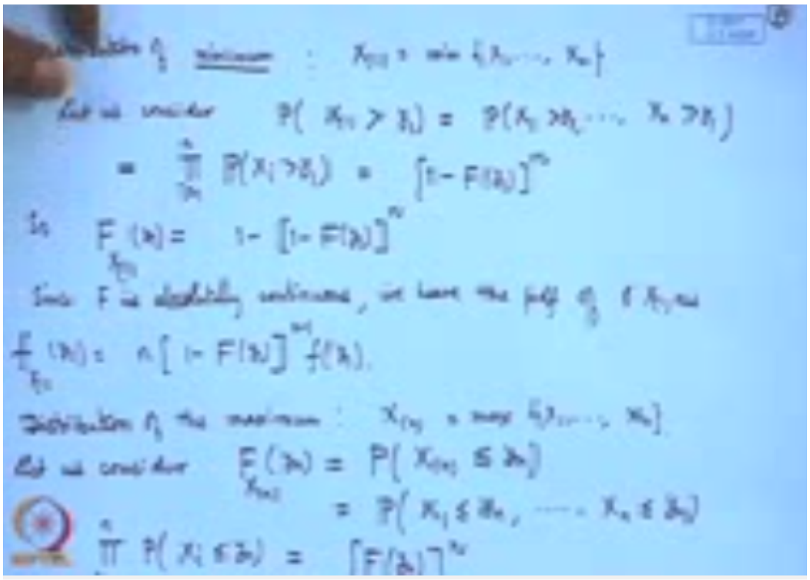
So, what you are getting product of f (yi). i = 1 to n and you have to sum up all these things n factorial times. Since there are n factorial regions the pdf of Y1, Y2, Yn is n factorial times product of f (yi) i = 1 to n. At the same time, you have Y1 < Y2 < Yn. In beginning I have not assumed any interval then it will be from - infinity to infinity.

If there is a sub interval for example if it is a uniform distribution on the interval 0 to 1 then this will be 0 to 1. If it is some a to b it will be like that if it is 0 to infinity, then this will come 0 to upper limit infinity so like that. There can be any regions here. So, this is the joint probability density of the order statistics. Once we have the joint density we can derive the density of the particular choice.

For example, if I want the density of second one, if I want the density of the third one if I want the density of the largest if you want the density of smallest, etc. we will have to evaluate the integral with respect to other variables. For example, if I want for Y1 then up to Y2 to Yn I have to integrate if I want for Y2 then Y1, Y3 and so on I have to integrate which can be done I will show you a systematic method for this can be obtained.

However, the case of the smallest and the largest can be done directly also because that is much more state forward based on the representation.

(Refer Slide Time: 30:38)



So, let me show you that things separately and then we look at that distribution of some middle one that is rth order statistics say where r can be 2, 3 and so on. So, let us consider say distribution of minimum that is x 1. It is the minimum of X1, X2, Xn. Let us consider say probability of say X1 > some values say Y1. Then it is = probability of X1 > y and so on Xn > Y1.

Because if the minimum is > Y1 then each of X1, X2, Xn has to be > Y1. At the same time each of X1, X2, Xn is > Y1 then the minimum has to be > Y1. So, this event and this event they are equivalent. Now this X1, X2, Xn are independently distributed so this can be written as the product of probabilities Xi > Y1 for i = 1 2.

Now each of Xi is having the same that of capital F. So, this is nothing but 1 – F of Y1 whole to the power n. So, what we have obtained? If we consider the cdf of X1 then it is = to 1 – 1 – F of Y1 to the power n. so, this is a general expression for the cumulative distribution function of the smallest order statistics and here you can easily see that I have not made any other assumption other than the form of the cdf as taking to be capital F.

There is no other assumption. I am simply taking X1, X2, Xn to be a random sample therefore they are having the same distribution F and they are independent, so this joint probability becomes equal to the product of individual probabilities. Since X1, X2, Xn they are continuous, so capital F is absolutely continuous function.

Therefore, it is differentiable almost everywhere and I can consider the pdf of since F is absolutely continuous. We have the pdf of X1 as, so if you differentiate with n times 1 – F of Y1 to the power n – 1 then derivative of this that is small f of Y1. So, we are able to derive the distribution of the smallest. Similarly, you can consider the distribution of the largest. That is Xn that is = maximum of X1, X2, Xn.

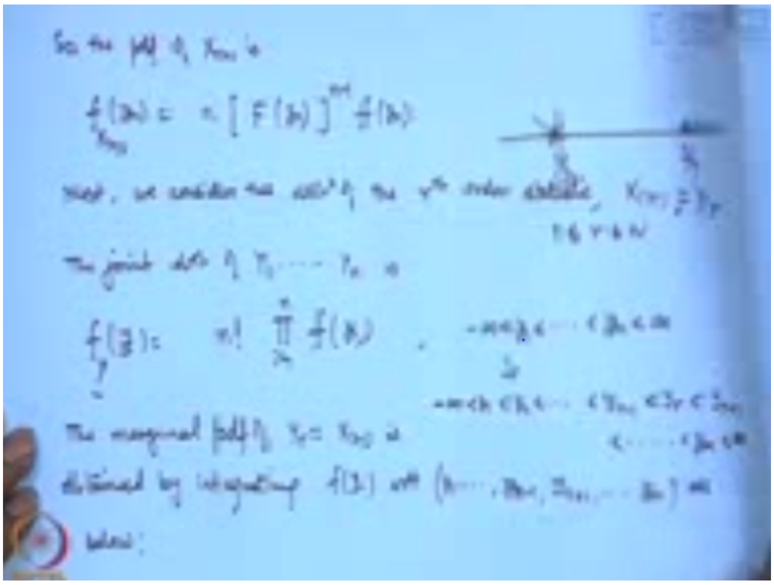
So, here let us consider again F (fx) n of Yn that is probability of Xn <= say Yn. Once again we utilize the definition of the order statistics we are saying that the maximum is <= to Yn. This is exactly equivalent to probability of X1 <= Yn and so on Xn <= Yn. Because if the maximum is <= Yn then individually each of X1, X2, Xn will be <= to Yn.

At the same time if each of X1, X2, Xn is <= Yn then the maximum is also going to be <= Yn. Once again as I applied the argument of independent and identically distributed random variables we can have this = product of probability of Xi <= Yn i =1 to n. And each of Xi has the same cdf F, so this is simply becoming F of Yn to the power n.

So, this is very interesting. The cdf of the largest is nothing but obtain from the original cdf by taking power n. So, this is very, very interesting in the minimum case it was becoming 1 – to the power n and then 1 - of that. And here it is the straight forward the same thing just raised to the power n. Now once again if capital F is the absolutely continuous function because F is a continuous random, x is a continuous random variable.

That means these random variables came from a continuous population. Therefore, it is differentiable almost everywhere and the pdf,

(Refer Slide Time: 36:20)



So, the pdf of Xn is obtained by derivative of this. So, it will become n times F (Yn) to the power n – 1 and then the derivative of this. So, I have been able to successfully obtain the distribution of the minimum and the maximum of the observations in a random sample. So, as I mentioned to you the typical applications are in selections when we want to choose the cheapest item we want to have the item with the maximum longevity and so on.

So, in almost all physical applications in economics, in industry, in social sciences, etc. everywhere we are having this thing holding here. So, that means we can actually derive the distribution of the minimum and maximum and we can utilize that. So, you can think in this particular fashion. For example, I have found the maximum now I am once we have selected that thing then we will be really using that again and again.

So, what is the distribution of that? For example if it is life what is the average life or what is the variability of the life? For example if it is the cost, so I have chosen the one item with the least cost then how that least cost is varying over time. So, all these things are of real interest in the for scientists and engineers in various disciplines. Including people working in economics or sociology where we are ranking the people according to some other kind of features.

Ranking of the items, ranking of the instruments and so on. Then more general thing would be that rather than looking at only the minimum and maximum we may be looking at any particular position. So, once again why that is of importance? For example in the usual parametric methods we generally consider the mean of the observations. So, mean of the observations is coming out we are able to obtain the distribution in most of the cases.

But when we are considering the form of the distribution not known then studying the distribution of the mean becomes very difficult. Except in the cases that where we assume that form and then we consider large sample theories, so we can consider central limit theorem.

But then again there are drawbacks of using the sample mean in sense of having less robustness in the sense that if there are wild fluctuation there are extreme observations then the sample mean is more affected. Then one may be interested in the median, median of the observations. So, median means suppose I am having odd number of observations then it is the middle.

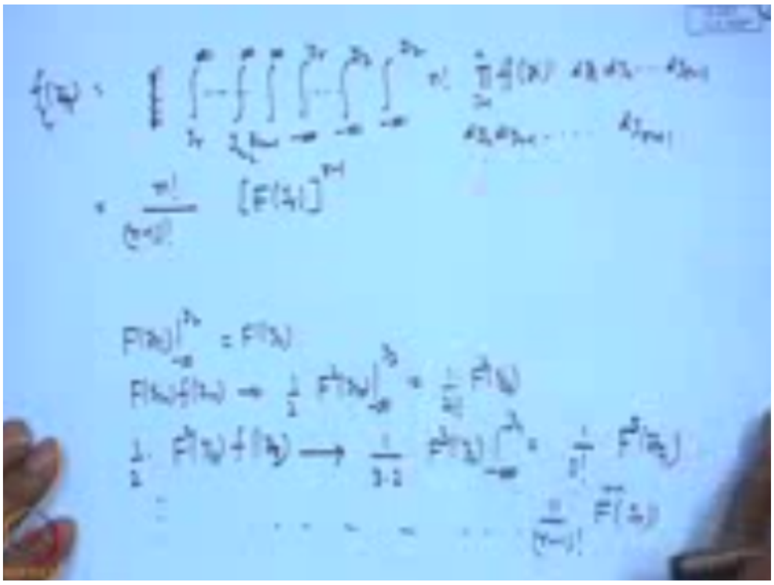
If I am considering even number of observations then it is between the middle 2 or you can take the average of the middle 2. That means in place of the largest or the smallest I may be interested in the distribution of some other order statistics. That means I may be interested in X3, X4. For example, I may be looking at a particular position. So, for example I may be interested in say one by 4th or I am interested in 3 by 4th that may be our cut off.

So, what are the distributions of this point what is the distribution of this point and so on. So, in general we want the distribution of the r^{th} order statistics. So, let us discuss that thing. So, next we consider the distribution of the r^{th} order statistics Xr. So, there of course your r is between 1 and n. If we go by our usual theory then what we have to do we have actually obtained the joint distribution of X1, X2, Xn.

So, Y1, Y2, Yn I use this notation. They are the orders statistics. From here I integrate the all the variables except the Yr here. So, we have to develop an algorithm for that. Let us consider the joint distribution, we call it y r. So, the joint distribution of Y1, Y2, Yn that we wrote n factorial times product of f Yi, i = 1 to n where. Now I write this region in a more elaborate way.

- infinity < Y1 < Y2 < so on y r – 1 < y r < y r + 1 < and so on y n < infinity. So, here except Yr I have to integrate all others. So, we can consider like this. The marginal pdf of Yr that = Xr is obtained by integrating Fy with respect to Y1, Y2, Yr – 1 y r + 1 and so on y n as below.

(Refer Slide Time: 42:51)



That is f (Yr) that is = to n factorial times, okay let me write it here. If we integrate with respect to Y1 then it will be from - infinitive to Y2. If we integrate then Y2 then it will be from – infinity to y 3 and so on up to Yr – 1 this will be integrated from – infinity to Yr. So, this is n factorial product of f (Yi) i = to 1 to n Dy 1, Dy 2, Dy r – 1. Now then let us look at integration of Yn. Yn will be integrated from Yn – 1 to infinity.

So, this will be from Yn – 1 to infinity. Then next one will be Yn – 1 then it will be from Yn – 2 to infinity and so on. Then ultimately y r + 1 will be from y r to infinity, okay. So, this is d y n, d y n – 1 and so up to d y r + 1. Now let us look at the integration part here. First one is integration of Fy 1. So, if we integrate Fy 1 we will get capital FY 1 and we integrate from - infinity to y 2.

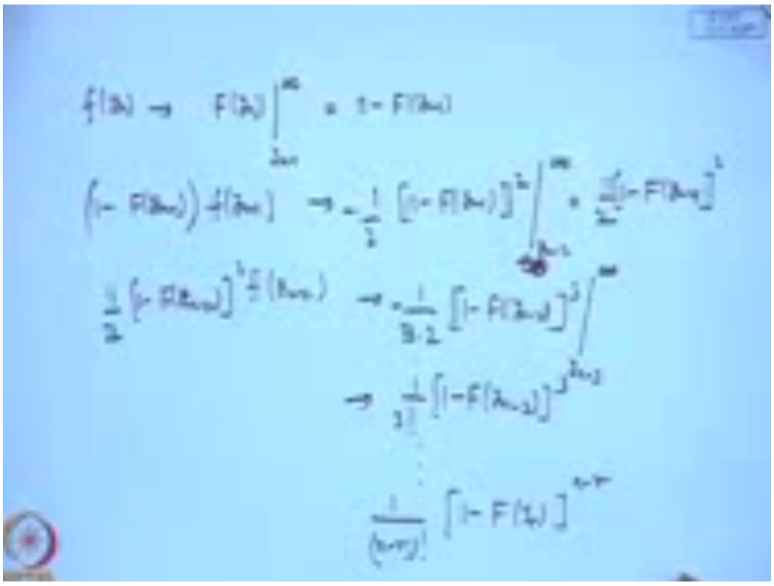
At Y2 this will becomes simply Y2 and this will become f (Y2) - f (-infinity) that is 0. So, it will be simply f (Y2). At the next stage then I have when I am integrating with respect to Y2 then I have to integrate f of Y2 into small f (y2). The integral of this will give me 1 by 2 f square y 2 from - infinity to y 3. So, this is simply becoming 1 by 2 f square y 3.

So, at the third stage then I have. 1 by 2 f square Y3 f (Y3) then I have to integrate with respect to Y3. Then this will give me 1 by 3 into 2 f cube y 3 from – infinity to Y4 which I can say 1 by 3 factorial f cube y 4. So, you can see this I call 1 by 2 factorial. So, I am getting a pattern here, so that means if I continue like this and go up to y r – 1 then the final integral will give me 1 by r the last one will give me f (Yr).

So, this will be r – 1 and this will be r -1 factorial because when we are getting 4 then here term is one less than it is 3 then one term is less here. So, at the when we do up to r then I will get f (Yr) here and here it will be r – 1 and r – 1 factorial. So that is one part here. So, I am getting n factorial divided by r – 1 factorial f (Yr) to the power r – 1. Now let us look at the other terms here.

The next terms will be coming small f (Yn).

(Refer Slide Time: 46:55)



So, the small f of Yn that will give me capital f (Yn) and the integral is now from Yn – 1 to infinity the integral for Yn is from Yn – 1 to infinity. So, this is Yn – 1 to infinity. Now f (infinity) is 1, so it is becoming 1 – f of Yn – 1. So, now at the next stage I will have 1 – f (Yn) – 1 multiplied by small f (Yn) – 1. So, this if I integrate I will get 1 by 2 1 – f (Yn) - 1 a square with a - sign.

So, this is from - infinity to now, sorry not from - infinity it is from now Yn – 2 to infinity. So, that is Yn -2 to infinity. So, this is giving me see at infinity this is becoming 0. So, this will become 1 – f (Yn) -2 whole square 1 by 2. So, now at the next stage I have 1 – Fy n – 2 x square f of Yn -2 and when this will be integrated I will get 1 by 3 into 2 1 – f (Yn) – 2 cube with a minus sign and this will be from Yn – 3 to infinity.

So, when I put the value this will give me 1 by 3 factorial 1 – Fy n – 3 cube. Like that I have to go up to Yr term here. So the last term will give me 1 by now you see if it is n – 3 here I am getting 3 factorial. So, if it is 1 – f (Yr) then the power will become n – r and here it will become n – r factorial. So, this will be the term which will be left out after integrating up to y r + 1.

So, let me substitute it here and I get here n – r factorial 1 – f (Yr) to the power n – r and I am left with the corresponding term f (Yr) here. So, we are able to derive the probability density function of the r^{th} order statistics which we can also write in that form of beta function because this is gamma r gamma n – r + 1 and this will become sum of that that is gamma of n – see if you add this n – r + r – 1 n – 1, so that is gamma of that.

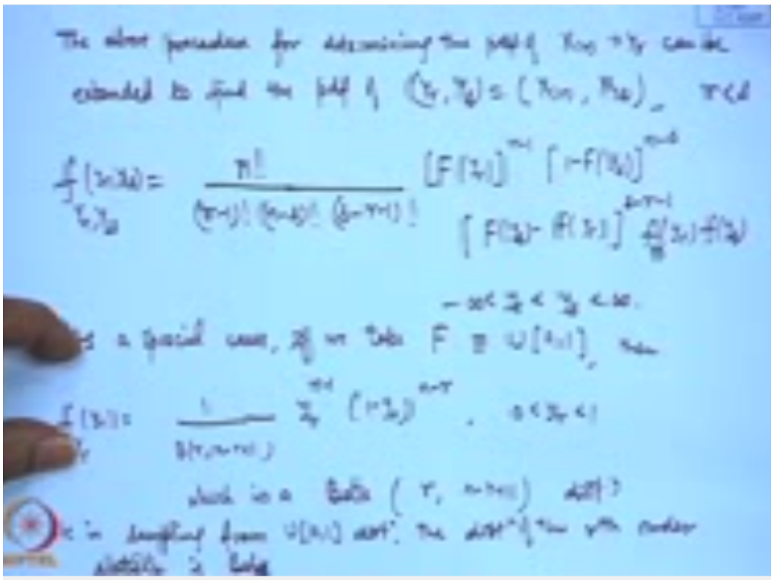
So, it is becoming basically beta of r n – r + 1. F (Yr) to the power r – 1 1- F Yr to the power n – r F (Yr) – infinity < Yr < infinity. So, we are able to actually derive the distribution of the r^{th} order statistics. Now this approach is very interesting because if I consider 2 here in place of 1. Suppose I want for first and second or I want for second and 4th that means in general I want for r^{th} and s^{th} order statistics.

Now then this procedure that I have given it will be applicable there also because then you will have to integrate up to r – 1 suppose r is < s. Then on the right hand side you have to integrate s + 1 s + 2 up to Sn and in between r and s you will have to integrate r + 1 up to s – 1. Now I have already given you the procedure that what term will become because if you are doing you will actually get this term that is 1 by r – 1 factorial f of y r r – 1.

Then you are doing up to s that is from s + 1 onwards you will get n –s factorial 1 – f (Ys) n – s. Now when you look at the terms between Yr and Ys that is Yr + 1 and so on then you are integrating in the range. So, that range will give you then simply r – s – 1 factorial, sorry not r –s. I am taking s to be > r.

So, it will be s – r – 1 factorial and the difference is coming here upper and lower limit in the these case one limit was coming out to be 0 but in that case both the limits will be coming. So, you will get f (Ys) – f of Yr to the power s – r – 1. So, this procedure can be extended. So, let me mention that thing.

(Refer Slide Time: 52:29)



The above procedure for determining the pdf (Xr) = Yr can be extended to find the pdf of say y r y s that is = to x r x s. Let us take a r < s. So, then this will give me f (Yr) Ys that is = 1 by r – 1 factorial, sorry this will be n factorial divided by r – 1 factorial and you will have n – s factorial as I mentioned and in the middle terms you will have s – r – 1 factorial.

And then you will have the f (Yr) to the power r – 1 1 – F (Ys) to the power n – s then you will have F (Ys) – F (Yr) to the power s – r – 1 F (Yr) F (Ys) – infinity < Yr < Ys < infinity. You can also look at say the distributions of 3 suppose I say r s t where r is < s < t. In that case you will have r – 1 then you will have n – t factorial then you will have s – r – 1 factorial then you will have t – s – 1 factorial and similar terms will come here also.

So, this procedure that I have given it is giving you an insight into the calculation for such things. So, that is very useful and you will be able to actually obtain the distributions of various type of such quantities here. In particular if we consider say any k order statistics where k is < n then that can also be obtained. Only thing is you have to write down that like if I say M1, M2, Mk where M1 is < m 2 < Mk.

So, entire thing can be written in algorithmic way that means the first one will become f of say y M1 to the power M1 – 1 then you will have f (My m – M2 – f (Ym) 1 to the power m 2 – m 1 – 1 like that all the terms will come there. So, this is a very useful method. I will be demonstrating you for example you can find out the distribution of the median the distribution of the range, etc.

But in particular let us see if we takes a uniform distribution then what do you get. Suppose I take F to be the uniform distribution as a special case if we take F to be uniform 0 1 then you will get say F (Yr) that is simply 1 by beta r n – r + 1 and here you will get imply Yr and here 1 – Yr. So, that is very interesting. This is nothing but which is a beta Rn – r +1 distribution. So, this is a known form. So, here it is very interesting here.

That if I consider a special case here then we are getting a beta distribution in the uniform case. So, this form of course it is used at many places and this is also given as an independent derivation of beta distribution. Although we have mentioned that beta distribution arises in various practical applications but this is also given as one application that means in sampling from uniform 0 1 distribution the distribution of the r^{th} order statistics is beta.

So, this is very interesting result here. I will be next talking about the movements of the orders statistics then as you can easily see that since we are not assuming a functional form for capital F that means we do not know exactly what form is there. That is difficulty in getting exact expressions. So, we will talk about that then we talk about the approximations then we talk about the expressions for the distribution of the order statistics.

Using this we will define something called an empirical distribution function which will be used as an estimate of capital F and we will use the properties of that. So, in the following lectures we will be continuing this theory.