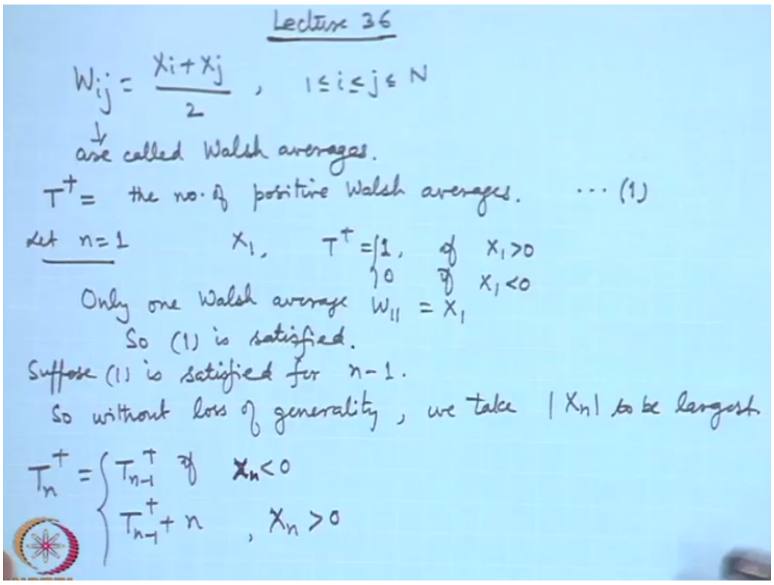
Statistical Methods for Scientists and Engineers
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology – Kharagpur

Lecture – 36
Nonparametric Methods - IX

In the previous lecture, I have shown that how the signed ranks, that is the ranks of absolute values are the modulus of xi can be used to create a test for the nonparametric location problems. Let us consider further the quantities which are called Walsh averages.

(Refer Slide Time: 00:42)



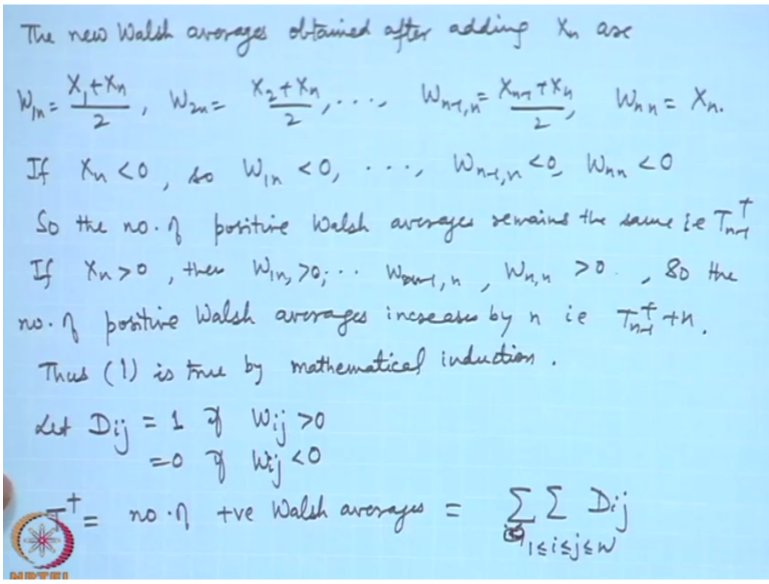
Wij, this is defined as the average of two of the observations xi and xj. So this wij these are called Walsh averages. Then if I consider T+ which I define as a Wilcoxon signed rank statistics this is actually the number of positive Walsh averages. Let me call it statement number one. Let us look at proof of this say suppose I take n = 1 that means only one observation is there. So that means it is only x1 so T+ that will be one if x1 > 0 and it is = 0 if x1 is < 0 and here only one Walsh average is there, only one Walsh average wii = x1.

So if it is positive so then it is T+ = 1 and it is negative then it is = 0. So one is satisfied for n = 1. Suppose this statement one is satisfied for n - 1. Now we have to add x in there and that means that we have to consider the modulus of xn. So without our loss of generality we take it to be largest. Suppose it is not largest then we can consider another permutation of that in which it will become the largest. Since we are assuming that one is satisfied for n – 1 so whatever permutation we take in that permutation also it will be true therefore without loss of generality.

So without loss of generality we take modulus of xn to be the largest. So now let us consider tn + 1 so I have put subscript here just to denote that it is based on n observations. So it is tn - 1 + if xn is negative because that will not add to the T+. T+ is the sum of the ranks of the positive one so if it is negative then it will not add and if it is plus then since it is I am assuming to be the largest its rank is n so that will be added here.

Now what are the new Walsh averages when I am adding nth observation then the new Walsh averages that will be coming.

(Refer Slide Time: 04:01)

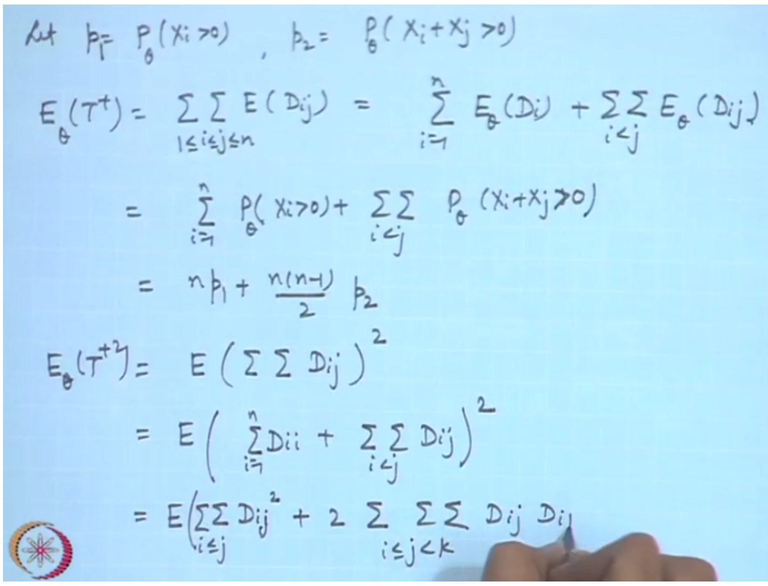


The new Walsh averages obtained after adding xn. They are x1 + xn/2 that is w1n, w2n = x2 + xn/2 and so on. Wn - 1 n that is xn - 1 + xn/2 and wnn that is xn. Now if xn < 0 I have assumed that modulus xn is the largest that means in absolute value it is the largest so if this largest absolute value if it is negative then whatever be x1, x2, xn - 1 ultimately they will make it negative so w1n it will become all of them become and of course wnn is negative.

So, the number of positive Walsh averages remains the same. On the other hand, that is tn - 1 +. If xn is positive, if it is positive and since it is the largest in magnitude therefore it will make whether it is positive or negative it will make all of them to be positive. Then w1 and this is positive and so on. W2n, wn -1, n, wnn they are all positive. So the number of positive Walsh averages increases by n that is tn - 1 + n.

So, thus one that is T+ is the number of positive Walsh averages. So I can call it as a theorem here which I have proved now by using induction, by mathematical induction this result is true all the time. Now based on this, let us define the indicator function dij = 1 if Walsh average is positive it is = 0 if Walsh average is negative. Of course = 0 we do not have to consider because of the assumption of the continuity of the random variables the probability of wij = 0 will be zero. So then T+ is the number of positive Walsh averages and that is actually equal to double summation Dij I = 1 <= j, 1 <= I <= j, <= n I can write here.

(Refer Slide Time: 08:00)



If I consider P as the probability of xi > 0 when true median value is theta, P2 is the probability under the true median value theta of xi + xj being positive. So I call these values P1, P2. Then in terms of P1 and P2 we can write expectation of T+ that = double summation expectation of Dij 1 <= I <= j <= n that is equal to expectation of Di that means for the ones which are j = I and then those terms for which it is less.

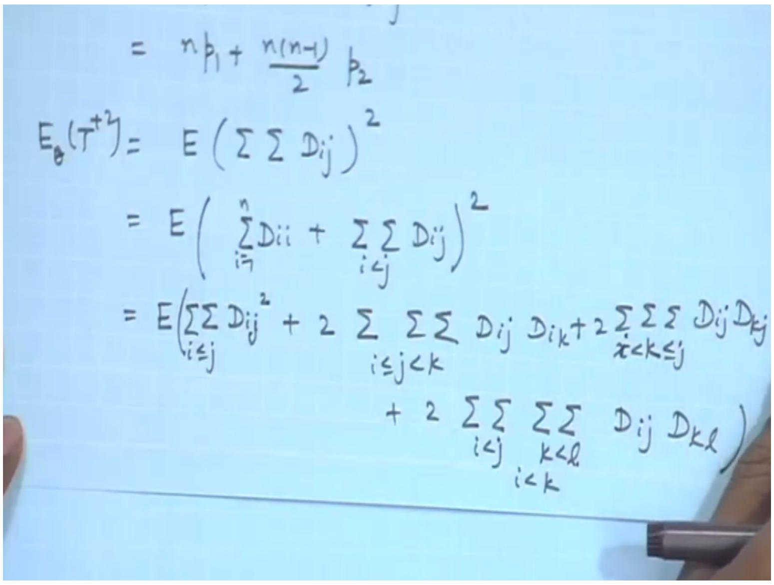
So this is = sigma probability of xi > 0 under theta and in the second one it is equal to probability of xi + xj > 0. Because actually I have defined in terms of directly the value 1 and 0 only for positive and negative therefore this is simply this. So it is actually becoming n * P1 + n * n - 1/2 P2. So under the alternative that means if median is any other value theta then the expectation of T+ will be in terms of this.

Similarly, if we look at the variance of this so for variance we need the expectation of this square so that is expectation of double summation Dij whole square. Now this you can write as expectation of summation Dii + summation Dij so this is I = 1 to n and here I < j. So this is equal to now let us expand these terms. So it is becoming Dii square and here what are the terms that we will be getting?

See all of the terms will be coming here and then there will be a crossed product also. So we can express it like this the terms for which so basically square of all the terms will be coming here because square of this and square of this so I can put double summation I <= j and this I can make ij here. So cross product terms will be of two types one is from here and one is from here. In this one you can consider ii and ij kind of terms.

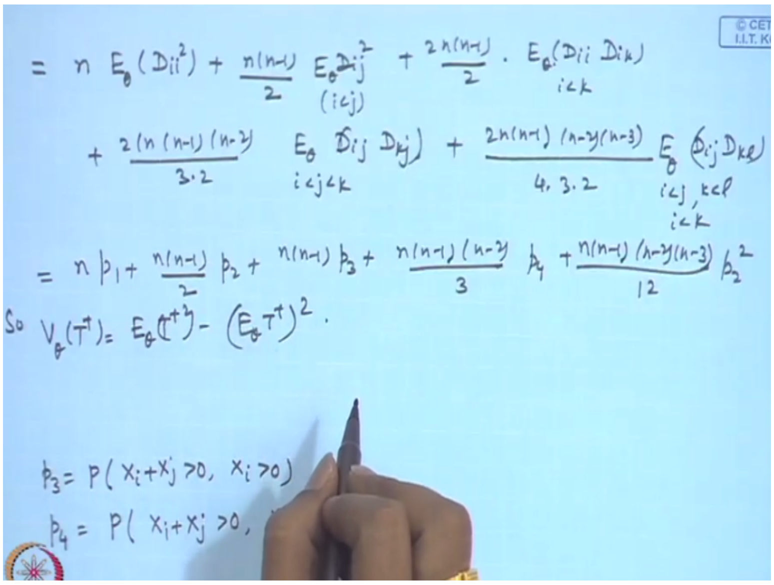
Here you can see that the terms will be ij and may be well first one may be same so ik. Then there can be term in which second is the common and then it can be one which is all are different so we can put it like this. I <= j < k, Dij, Dik.

(Refer Slide Time: 11:47)



Then plus we can write I < k or = j, Dij, Dkj and then we can also write two times I < j, k, l, I < k, Dij, Dkl. So these many terms will be coming when I squared this. So now if I look at expectation, so expectation here and in each of them it can be applied let us look at this.

(Refer Slide Time: 12:40)



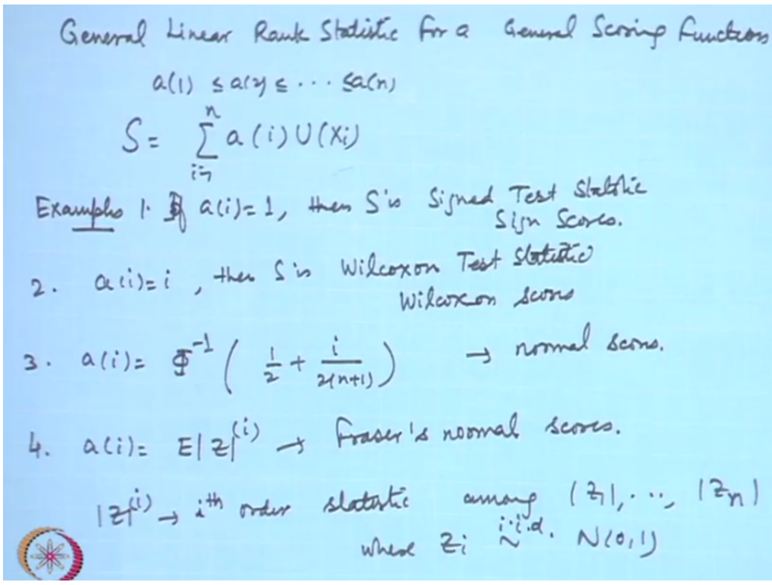
So this expectation is = n times expectation of say Dii square plus n * n - 1/2 expectation of Dij square plus here of course I < j + twice n * n - 1/2 expectation of Dii, Dik where I < k. Of course all of them are under the assumption that the median is theta plus twice n * n - 1 * n - 2 3 *2 expectation of Dij, Dkj. This is for I < j < k twice n * n - 1 * n - 2 * n - 3/4 * 3 * T to expectation of Dij Dkl here I < j, k < l and I < k.

So if you combine all these terms here now some additional probabilities will be coming, Earlier I defined P1 and P2 but now because this one will involve some "Voice not clear" probabilities let me write it here. If I define say P3 = probability of say xi + xj > 0 as well as xi > 0. Similarly, if I define P4 = probability of xi + xj > 0, xi + xk > 0. So this term is then becoming n * p1 + n * n1 - 1/2 P2 + n * n - 1 P3 n * n - 1 * n - 2/3 P4 + n * n - 1 * n -2 * n - 3/12 P 2 square.

So once we have the expectation of T+ square expectation T+ is already there so we have the expression for the variance of T+ also so that is basically this term - expectation theta T+ whole square. So I am not writing the full expression here it is simply the repetition. So in case the median value is some theta then also we have been able to derive the moments etc of the T+ here.

If you look at the nature of this statistics that we have used here it is defined as the ranks of the positive once and then uxi. So these are some functions of xi or some functions of modulus xi so we can consider various choices here and with various choices one can consider the general scoring function and we call it a linear rank statistics.

(Refer Slide Time: 16:36)

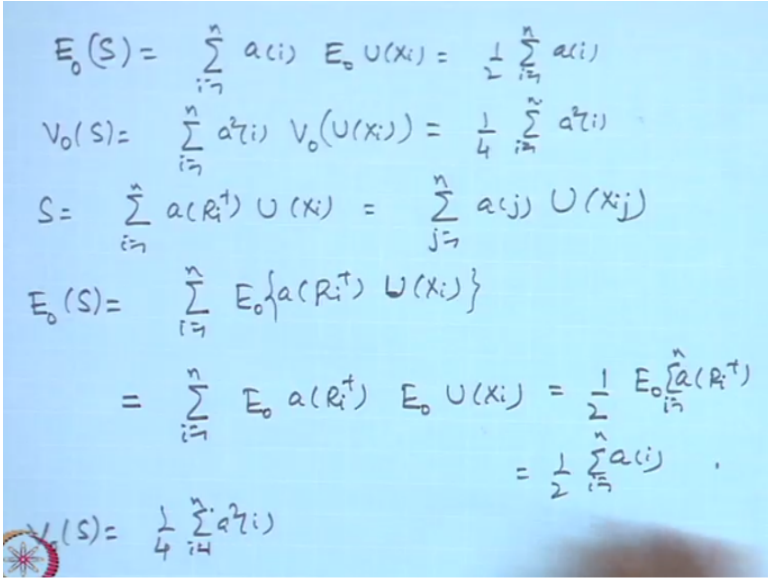


So general linear rank statistics for a general scoring function. So in general we consider a1 <= a2 <= to some an and we define s = sigma of ai u(xi) I = 1 to n. So as examples you can see if we are considering ai = 1 then it is signed test. So these are called signed scores. Second example is if ai = I then s is Wilcoxon where these are called then Wilcoxon scores. So actually this ai are called score function there are some others also like.

For example you may choose ai = in terms of the cumulative distribution function of standard normal that is half + i/twice n + 1. These are called normal scores and then there is another one called Fraser's normal score where this zi is actually the ith order statistics among modulus of z1, modulus z2, modulus zn where zis are iid normal 0, 1.

So if I consider a random sample from a standard normal distribution and I consider the absolute values. Then the relative position of modulus zi that is ith order statistics. Then based on that if I define this then it is called Fraser normal scores. since the null distribution of uxi is known therefore we can look at the mean variance of etc of s in the general sense here.

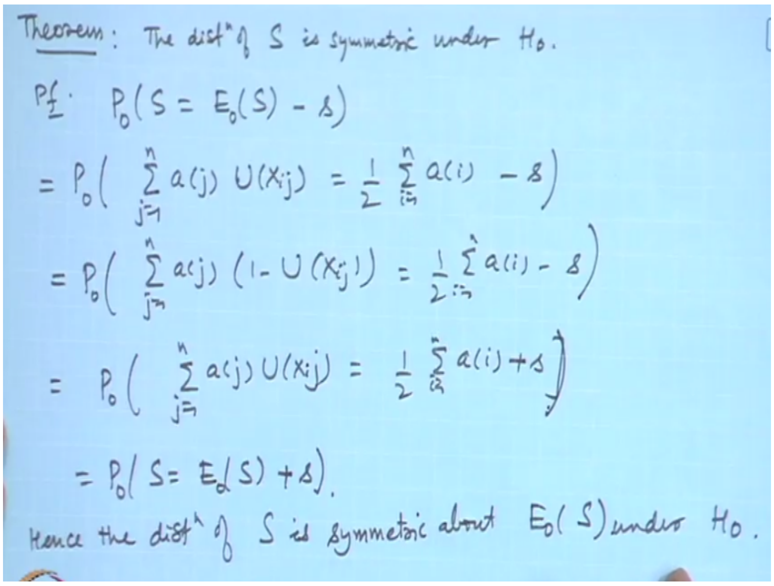
(Refer Slide Time: 19:36)



So if I look at say expectation of s that = sigma of ai that is half so it is sigma of ai I = 1 to n. So in all the cases when these are only permutation of number 1 to n then this will become simply n * n + 1/2. For example in the Wilcoxon score it was like this. For sign this was I so it was n/2. So like that there can be various choices here. If I look at variance of this then it is equal to sigma a square I variance of u(xi) I = 1 to n = so half - half square that is 1/4.

So it is 1/4 sigma of a square i, I = 1 to n. So this s = sigma a of ri plus u(xi) can be written as sigma aj u(xij). Based on this if I consider the expectation of set sector then what I will get? U(xi). We have already proved that the distributions of Ri + and xi are independent so this becomes expectation of a(Ri +) * expectation of U (xi), but this is half here. So it is simply becoming half times expectation of a (Ri +) but if I summing over all of them then it is simply all the values are coming here that is all and variance of s will be simply 1/4 sigma a square i, I = 1 to n.

(Refer Slide Time: 22:23)



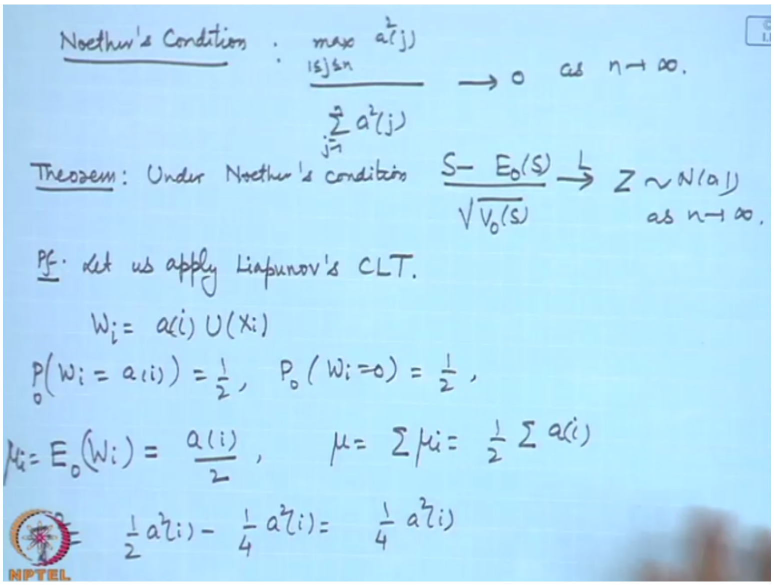
In general we can prove the following theorem that the distribution of s is symmetric under H0. So for a proof let us look at what is the probability that s = - some s. So we will prove actually it is symmetric about its mean so it is probability of sigma aj U (xij) expectation is we have already calculate so it is simply - s here. Since the distribution of u(xij) and I - u(xij) is the same because what is u(xij)?

U(xij) takes value one with probability half and 0 with probability half when the null hypothesis is true that is when the median is assumed to be zero. If that is so then if I look at 1 - u(xij) that is also having the same distribution because that is also taking value 0 and 1 only each with probability half. So in this statement I can replace 1 - u(xij) that is half sigma ai - s. Now this term we take to the other side so you are getting it is equal to P0 sigma aj u(xij) j = 1 to n = half times sigma ai + s.

That is probability of s = expectation s + s. So the distribution of s is it is symmetric about expectation of s. if it is not under H0 then the distribution will not be symmetric because then the distribution of u(xij) and 1 - u(xij) will not be the same. Here since it is under the null hypothesis so both the probability of xij being positive or negative is half. Now this part you can see.

This is a general theory. Because I am as you may get particular form of s here I am writing general scores ai is there. We have seen that the asymptotic distribution of Wilcoxon signed ranked statistics, the asymptotic distribution of the signed test statistics they are all normal asymptotically. There we were able to do the exact calculation so here it is in the terms of ai if we impose certain condition on these score that ai etc then here also we can obtain the asymptotic distribution to be normal. So we impose some condition. These are called Noether's condition named after Emy Noether.

(Refer Slide Time: 26:13)

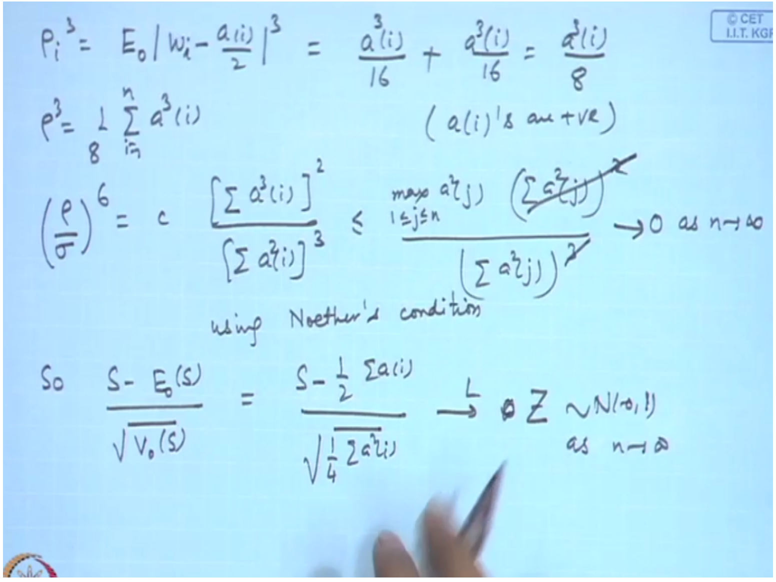


So let us define this Noether's condition. What is the Noether's condition? The Noether's condition is maximum of aj square for 1 <= j <= n/sigma a square j j = 1 to n. This goes to 0 as n tends to infinity. So this is called Noether's condition. Then we have the following result that is under Noether's conditions. The distribution of s - expectation is under H0/square root variance of S. This converges to standard normal as n tends to infinity.

Now you see here the expression for the general linear rank statistics I have written in terms of summation here. So if we use this Liapunov's central limit theorem we can do that thing. So let us apply Liapunov's central limit theorem. What the expressions here? The value of once again let us go back to this expression here. Let us call this expression is some wi. Then what is the value of wi, it is either + ai with probability half and it is zero with probability half. So let us write that.

Let us write say wi = ai u(xi). We can also write it is aj u(xij) as we have done in the previous one it does not matter. So wi = ai under the null distribution = half and probability that wi = 0 that is also half. So if I look at the expectation of wi that is ai half and therefore if I look at that I call it mu I then it = half sigma of ai. Now let us look at the second one sigma square. So that will be equal to half of a square I - 1/4 a square I = 1/4 a square i. So sigma square then that is becoming equal to 1/4 sigma of a square i, I = 1 to n.

(Refer Slide Time: 29:30)



We also need the third central moment here. So the third central moment here will become equal to wi - ai/2 cube. So when it is equal to ai it is simply becoming ai cube/8 and then you are dividing by two so it is becoming by 16. Then when wi is 0 then it is becoming again ai cube/8 then half. So ai cube/16 that is = ai cube/8. So rho cube = sigma of a cube I 1/8 I = 1 to n. Of course there is one comment here. I did not mention about ai.

What are the values of ai. ai are either 1s or 0s in this for example in the case of in the signed rank it is 1 to 0 otherwise it is I etc. So in general actually we are taking ai are positive. So this term I did not mention earlier but it is requiring otherwise you have to again put modulus here. So we have to consider rho/2. It is more convenient if I take the power 6 here so it will become that is 2 to power 3 so it is becoming 2 to power 18 and here you are having 2 square so that will become 2 to the power 12.

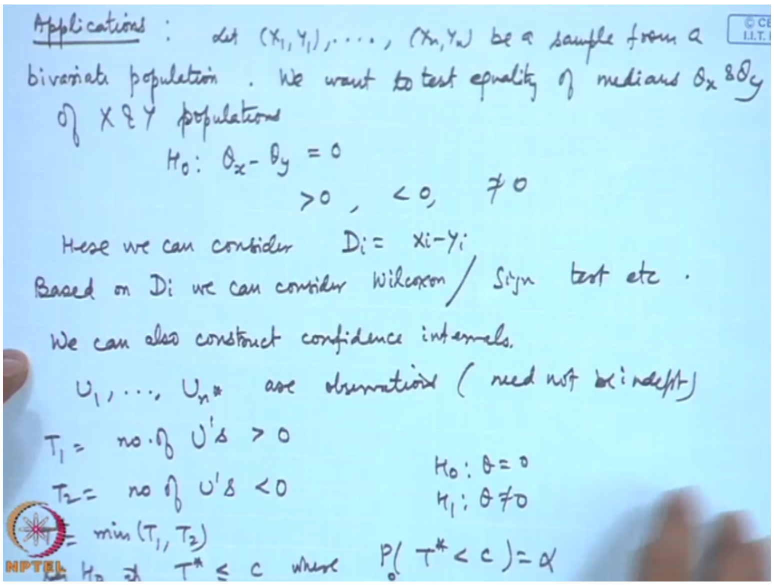
So some coefficient times, so some constant times you will get sigma of a cube I whole square divided by sigma a square I whole I cube. Now this term we separate out. this is <= in one of them we put maximum here. So this is <= maximum of a square j for 1 <= j <= n. So basically what I am doing I am splitting it and in the second term I am writing it as simply sigma a square j whole square and in the denominator.

I am having sigma a square j whole cube so this term gets cancelled out so by Noether's condition this goes to 0 as n tends to infinity. So s - expectation s/square root of variance of s that is actually s - expectation s. We have already calculated here that is half sigma ai/square root 1/4 sigma of a square i. This goes to Z as n tends to infinity. So the asymptotic distribution of the general linear rank statistics are satisfied and actually for the signed test statistics for the Wilcoxon singed rank statistics it is already satisfied.

I wrote two more scores that is the normal score and the Fraser normal score. for that also one can actually check that this will be satisfied. Now the procedure that I have developed for single sample problem in some cases they can be also extended to two sample problem for example if we consider bivariate and we still want to compare the locations of both of them then we can take the differences.

Now based on the differences. Now based on the difference if you look at the distribution of that and we define this ranks of that difference then this test statistics will again work here. So let us consider these extensions to various other cases. We can also look at some confidence interval procedures etc.

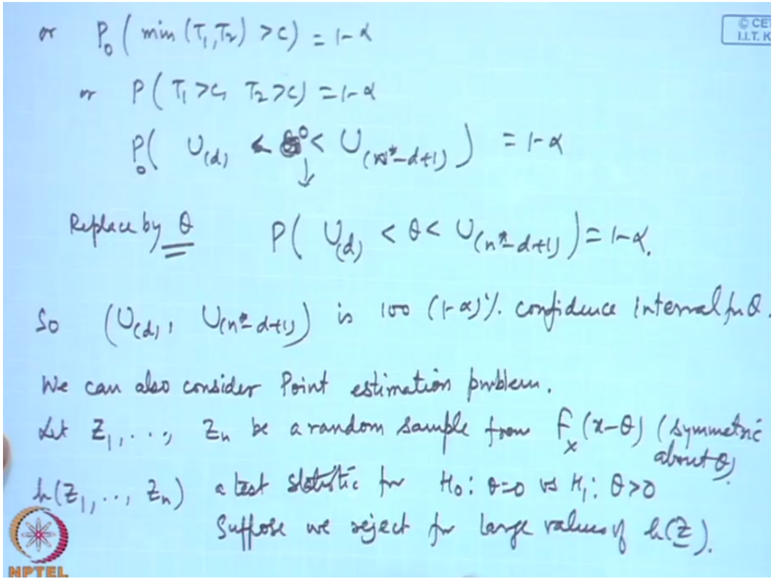
(Refer Slide Time: 35:00)



Let x1, y1, x2, y2, xn, yn be a sample from a bivariate population and we want to test say equality of medians theta x and theta y of x and y populations that is the separate population. That means my hypothesis testing problem is something like this theta x - theta y = 0 > 0, < 0 not = 0 etc. So here we can consider say Di = xi - yi. Now based on Di we can consider say Wilcoxon test or sign test etc.

Another application is to look at the confidence interval. We can also construct confidence intervals. Suppose we are considering some other number say u1, u2, un *. Suppose they are observations and they need not be independent also. We can consider say T1 = number of u which are > 0. T2 is the number of us which are < 0. Then we can take minimum of T1, T2. So based on this we can consider let us call it say T*. We can reject H0 if so hypothesis say theta = 0 against theta not = 0. We can reject this if T* is <= some C where this should be equal to alpha.

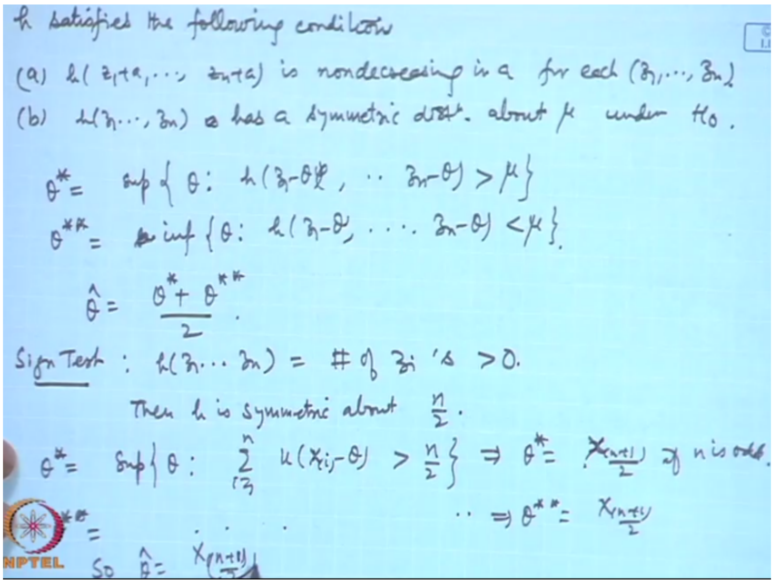
(Refer Slide Time: 38:25)



Now this can be determined as we are saying minimum of T1, T2 > c is = 1 - alpha or we can say T1 > c, T2 > c = 1 - alpha and this you can consider as say u of d < 0 < u n* - d + 1 = 1 - alpha. Now in place of 0 we replace by theta. Then this is becoming ud < theta < u n* - d + 1 = 1 - alpha. So we are getting that ud to un* - d + 1.

This is 100 1- alpha percent confidence interval for theta. We can also consider point estimation problem so let us consider say z1, z2, zn be a random sample from a location parameter distribution fx - theta and symmetric about theta and let us consider say h of z1, z2, zn a test statistic for H0 theta = 0 against say h1 theta > 0 and suppose we reject for large values of h(z).

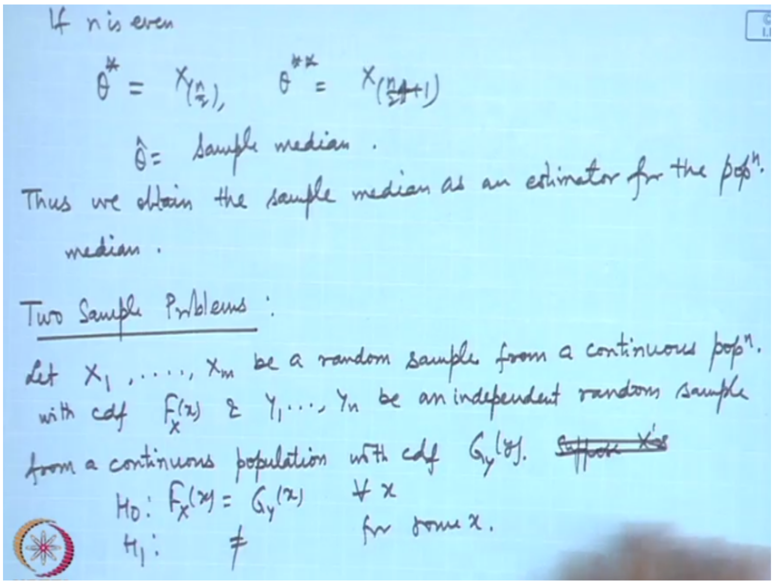
(Refer Slide Time: 41:08)



Then you look at this. H satisfies the following conditions. One is that h of z1 + a and so on, zn + a is non-decreasing in a for each of z1, z2, zn. Second is that h of z1, z2, zn has a symmetric distribution about mu under H0. If we consider say theta star = supremum of those values for which h of z1 - theta and so on zn - theta is > mu and theta double star say = infimum of those values theta for which h of z1 - theta and so on zn - theta is < mu. If I consider say theta head = theta* + theta** /2.

For scientist for example, here h of z1, z2 zn is the number of zi which are positive then h is symmetric about n/2. So here theta* will then become equal to supremum of theta, sigma of u(xi) - theta > n/2 I = 1 to n that will give me theta* = xn +1/2 if n is odd and similarly theta** that will become equal to xn + 1/2. So what we are getting that theta head = xn + 1/2 as the estimator of the median theta. So basically we are getting a sample median as the estimator the population median.

(Refer Slide Time: 44:20)



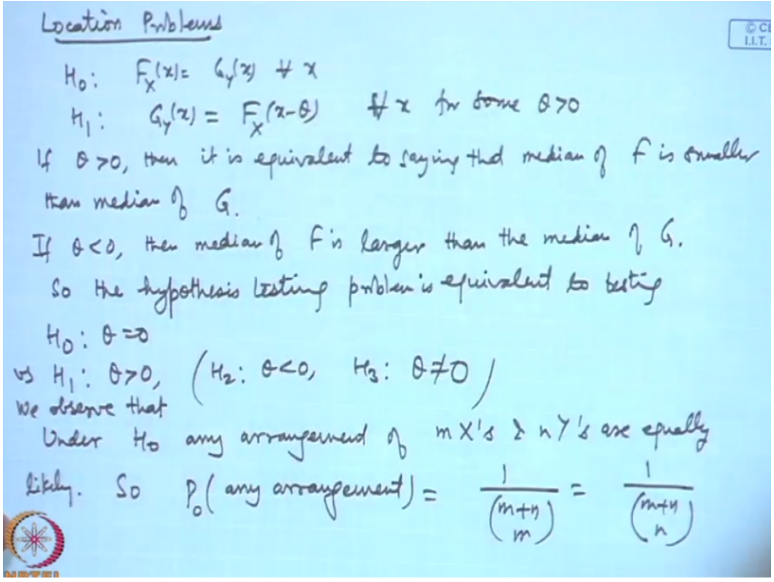
If n is even then if you calculate these quantities theta lower, upper star that will be = xn/2 and theta** this will become = xn/2 + 1. So theta head is then again sample median. Thus we obtain the sample median as an estimator for the population median. So these linear rank statistics or the score function that is actually useful in deriving confidence interval for deriving the point estimates.

It can help us in testing about the equality of the median in a bivariate problem also. So these are various applications of general linear rank statistics. Then the other problems that come is comparing the medians of two independent sample. So in that case this directly cannot be used. So let us consider separately the two sample problems. So our next topic in this non parametric methods is two sample problems are two sample location problems so let us consider say. In fact, I have already given the form of ui and u bracket I etc so we will see how these terms are used.

So x1, x2, xn be a random sample from a continuous population with cdf say fx and y1, y2, yn be an independent random sample from a continuous population with cdf gy. Suppose this x. So we are already assuming that they are independent and in general we want to test say fx = g(x) for all x against not equal to some x. So this is the general problem of equality of the two distributions, but in particular we can consider location equality problem.

In equality problems scale equality and inequality problems and general alternatives. Let us consider say location problems here.

(Refer Slide Time: 47:37)



In the location problem we consider H0 as that fx is equal to gx for all x and in h1 gy is a location shift of f for all x for some theta. So this is interesting here. I have actually if you consider the standard problems in the normal distribution etc then these problems are directly related to the equality of the location that is the mean etc. Here it will become in terms of median you can say.

So we can consider if theta is > 0 then we are basically saying it is equivalent to saying that median of f is smaller than median of g. If theta is < 0 then median of f is larger than the median of g. so the hypothesis testing problem is this is equivalent to testing H0 theta = 0 against H1 either theta > 0 or theta < 0 or theta not = 0. These are the alternative. So as you can see that it has come down to the original type of problem here.

Now we will introduce a test statistic it is called Mann-Whitney-Wilcoxon test. When the null hypothesis is true then basically we are saying that two distributions are same and then this x1, x2, xn and y1, y2, yn. This can be actually considered as one sample and if it is one sample then all the arrangements of this x1, x2, xn, y1, y2, yn among the numbers 1, 2, m + n. They will be equally likely.

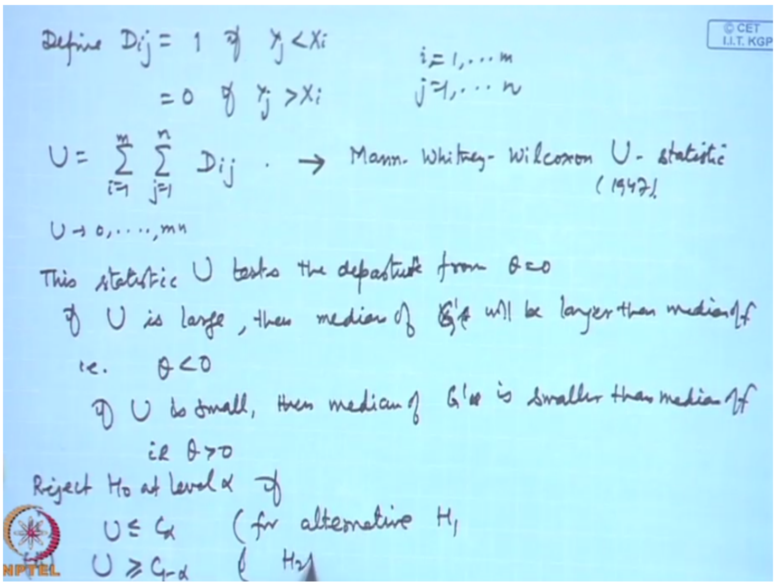
If I consider the probability of any arrangement of mx is among m + n observations, then it will be 1/ m + n cm or 1/m + n cn. So we utilize this concept here. Basically we consider counting of how many yj are < xi, how many yj are > xi etc. Actually that will directly give us a hint of this testing problem. You can see that in the case of parametric frames we look at the means of the observations.

Since here if I look at the means of the observations etc, the distribution will be extremely complicated because we do not know actually what is the form of f and what is the form of g. Therefore, we have to do or we have to actually work with the numbers only that is the ranks or the how of many of them are positive, negative, how many of them are > the other on because the probabilities can be calculated in terms of F and G.

But we cannot calculate expectation and another quantity is if we do not have the basically we cannot find out the distribution of the sums of the observations or the means of the observations. So that is the difference between actually the methods of the parametric inference and the nonparametric inference. In the parametric inference we directly go down to the sufficient statistics we check whether it is complete or not and then we base our inferences on that.

In the case of nonparametric that is not possible and therefore we work with the order statistics we work with the signs of the things we work with the ranks. So under H0 we observe that under H0 any arrangement of mx and nys are equally likely. so probability of any arrangement is 1/ m + ncm which is also equal to 1/ m + cn.

(Refer Slide Time: 53:09)



Let us consider say define Dij = 1 if yj is < xi and = 0 if yj > xi. So this is defined for all I = 1 to m and j = 1 to n and then we define u = double summation Dij for I = 1 to n and j = 1 to n. This u it is actually known as Mann-Whitney-Wilcoxon U statistics. It was given in 1947.

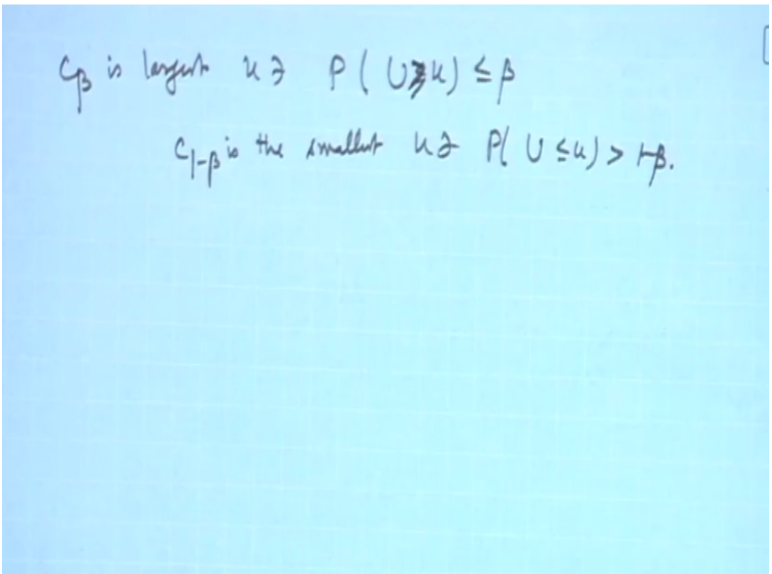
Now what are the possibilities here. It could happen that all the yj are greater than all xi. In that case value of u will be 0.

If you have all the yj < all of xi then all the Dij will be 1 and therefore this value will become equal to mn. So the values of u this will vary from 0 to mn and therefore it will test the departure from the equality of the median. So for example if more of the yj < xi then that means Dij is the higher value that means of median of Y is < the median of x and that is equivalent to saying that median of G is smaller than the median of this one then it is actually equivalent to theta < 0.

Similarly, if this is smaller then we are getting theta < 0 that means if u is smaller than more of ij are larger than the xi that means median of y may be tending to become higher than the median of this. If that is so then you will get theta > 0 so this hypothesis will be true and similarly for either very large or very small you will have theta noT+ 0 so all the three cases will be actually satisfied here.

So this statistic u tests the departure from theta = 0 if u is large then median of y will be larger than median of g will be larger than median of f that is theta is < 0. If u is small then median of G is smaller than median of f that is theta is > 0. So we can consider the following test that is reject H0 at level alpha is U is <= c alpha. This is for alternative H1. U is >= c1 - alpha. This is for alternative H2 and for the third one u <= c alpha/2 or u >= c1 - alpha/2. This is for alternative H3 and,

(Refer Slide Time: 57:22)



C beta is the largest u such that probability of u <= u is <= beta or c1 - beta is the smallest u such that probability of u < = u is > 1 - beta. I think this should be >= u. In the next lecture we will discuss the null distribution of us how it is obtained, the mean and variance under the general hypothesis, there is a related one which is called Wilcoxon statistics for the two. We will define the general rank statistics for the two sample problem. We will look at the asymptotic distributions of that so these are various things that I will be taking up in the next lecture.