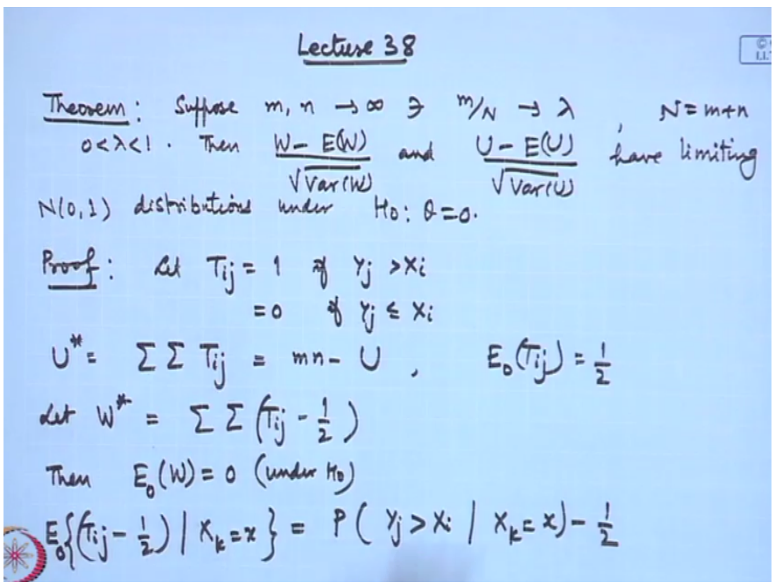
Statistical Methods for Scientists and Engineers
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology – Kharagpur

Lecture - 38
Nonparametric Methods - XI

We were discussing the theory of general linear rank statistics in the previous class and we have discussed the distributions, then we also talked about how to find out the asymptotic distribution of the general linear rank statistics. Now using that theory, we will derive the asymptotic distribution of Mann–Whitney U statistic and the Wilcoxon rank sum statistics for two sample problems. We will show that they are actually asymptotic normal.

So this is proved in the following theorem.

(Refer Slide Time: 00:53)



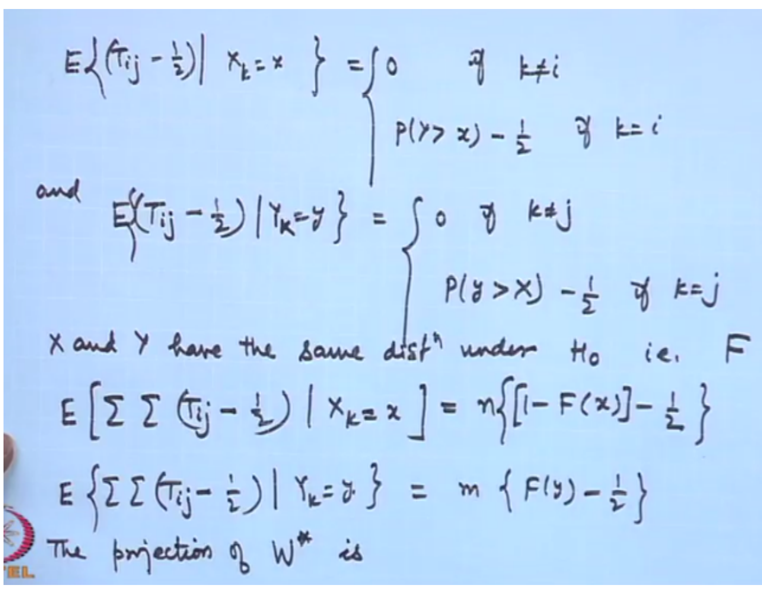
So we consider that m and n tend to infinity such that m/n goes to lambda. That means where N=m+n, that means basically what we are saying is that it is not that abnormally one of the sample sizes becomes very large. It will be that both of them will have a fixed ratio. So here, of course, your lambda will be between 0 and 1. Then the standardized Wilcoxon statistic and similarly the standardized Mann-Whitney statistic.

That is the W-expectation W/square root variance of W and U-expectation U/square root variance of U. These 2 have limiting normal 0, 1 distributions under the null hypothesis that is theta=0. To prove this, let us consider here the definitions of this. Let me rewrite this thing. Suppose I consider Tij=1 if Yj is >Xi and it = 0 if Yj is </=Xi. So if I consider U*=double summation Tij, then actually it is =mn-U.

Because U was defined as the sum of sigma Tij where Tij was one when Yj is < Xi, that means it is reverse of this. We are assuming that the ties are not occurring, then we will have expectation of Tij=1 under the null hypothesis. That is when theta = 0. Let us consider say W*, which is based on Tij. If we define this thing, then we will have expectation of W also = 0 under H not that is when theta = 0.

Now let us consider the conditional expectation of Tij-1/2 given that Xk = X under the null hypothesis, then it is nothing but the probability of Yj>Xi given that Xk=X-1/2. So we can then write.

(Refer Slide Time: 04:05)

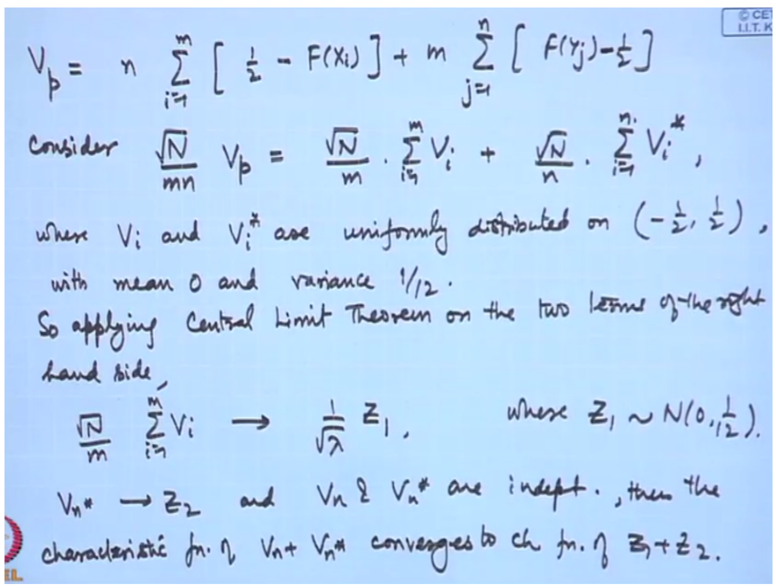


Expectation of Tij-1/2 given Xk=X. This is = 0 if k is != i and it is = probability of Y > X-1/2 if k is = i. Similarly, if I consider expectation of Tij -1/2 given Yk is =Y, then this is =0 if k is != j and it is equal to probability of Y>x-1/2 if k=j. Now we are assuming that under the null hypothesis that is theta = 0, X and Y will have the same distribution under H not, that is F. So if we consider expectation of double summation Tij – 1/2 given Xk=X.

Then that will become simply =, now you see here that I will get this value when k=i for all other values it will be =0. So how many times that will occur when k=i, that is how many Xi are there. There are n, so this will become n * 1-FX-1/2. So this I can write like this. Similarly, if I consider expectation of double summation Tij-1/2 given Yk=Y, then this is = m*FY-1/2. You can see here, this is X<Y, so that is FY and here it is Y>X, so it is a CDF of Y that is F.

So it is 1-probability Y</=X that is 1-FX here. Now the projection of W*.

(Refer Slide Time: 06:54)



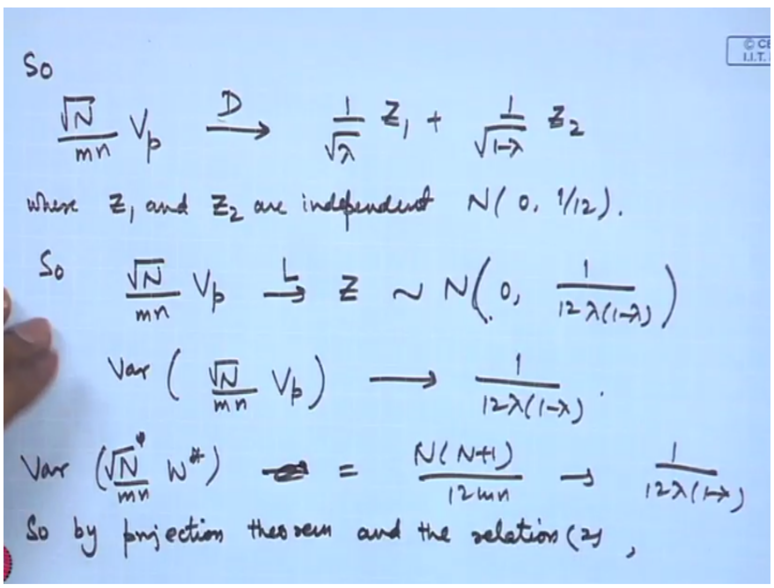
We are writing it as Vp=n*sigma 1/2, so this is i=1 to n, - F(Xi) + m* sigma j=1 to n F(Yj-1/2). So let us consider here square root n/mn Vp that is = square root n/n * sigma Vi, i=1 to n+ square root n/n sigma Vi* i=1 to n. What are these Vi and Vi*, we are considering Vi and Vi*, they are uniformly distributed on interval -1/2 to 1/2. So they will have basically 0 mean and variance will become 1/12. Now these are in the form of the summation.

So we can apply the central limit theorem on the 2 terms on the right hand side. Basically what I have done, I adjusted these terms here. See this particular term because of the probability, integral transform, this becomes uniform distribution on the interval 0 to 1, F(Yj) becomes uniform distribution on the interval 0 to 1. So 1/2-F(Xi) becomes uniform distribution on the interval – 1/2 to 1/2.

Similarly, F(Yj) – 1/2 becomes uniform distribution on the interval -1/2 to 1/2. So both of these are summations now and we apply the central limit theorem. So applying central limit theorem on the 2 terms of the right hand side, what will happen, we will get square root n/n sigma Vi, I=1 to n, this will converge to 1/root lambda Z1 where Z1 is following normal 0 (1/12) and in a similar way, if we consider Vn* converging to Z2 and Vn and Vn* are independent.

Then the characteristic function of Vn+Vn* converges to characteristic function of Z1+Z2. Basically what we get then here.

(Refer Slide Time: 10:43)



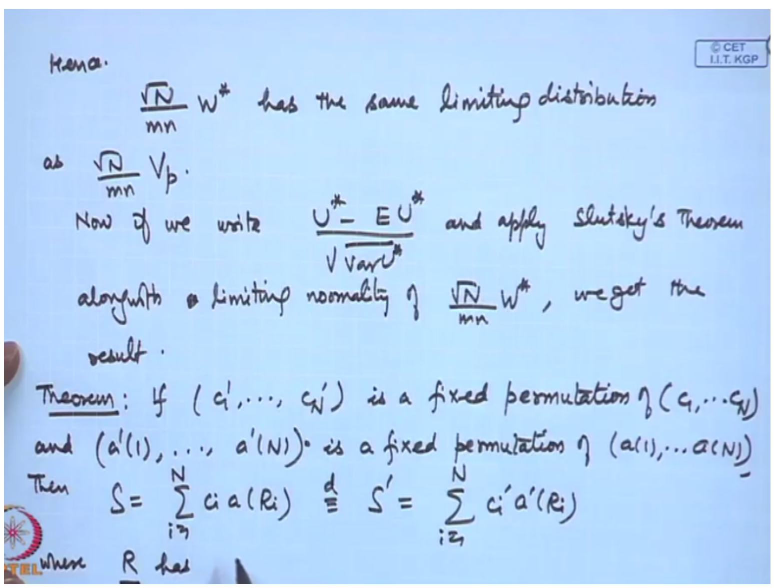
So root n/mn Vp, this will converge to, that means I am considering the sum hereafter adjustment here. So this is converging to 1/root lambda Z1 and this is converging to 1/square root 1-lambda Z2. So this coming to 1/root lambda A1 + this means converges and distribution 1/root 1-lambda Z2 where Z1 and Z2 are independent normal 0, 1/12. So if we apply the linearity property of the normal distribution, we get that root n/mn Vp converges and distribution to Z, which follows normal 0, 1/12 lambda *1-lambda.

We can also talk about the asymptotic variance. So variance of square root n/mn Vp, this will converge to 1/12 lambda *1-lambda and variance of square root n/mn W* that will also converge to that is = n*n+1/12 mn that is also converging to same value. Because this was just a linear combination of this thing. So if we now use the projection theorem, which I gave in the last class, let me just repeat it here.

That expectation V-W square is minimized by choosing Pi*x as expectation of V given xi and this is the projection and expectation of V-Vp square = variance of the –variance of Vp. So we use this result here now. So by projection theorem and the relation 2 which I just now showed you, we get that expectation of root n/mn W*- root n/mn Vp. This goes to 0. So if we use the theorem, which I gave for the limit part.

That is if Wn is asymptotic distribution and expectation of Un-Wn square goes to 0, then Un also has an asymptotic normal distribution. So if we use this.

(Refer Slide Time: 14:17)

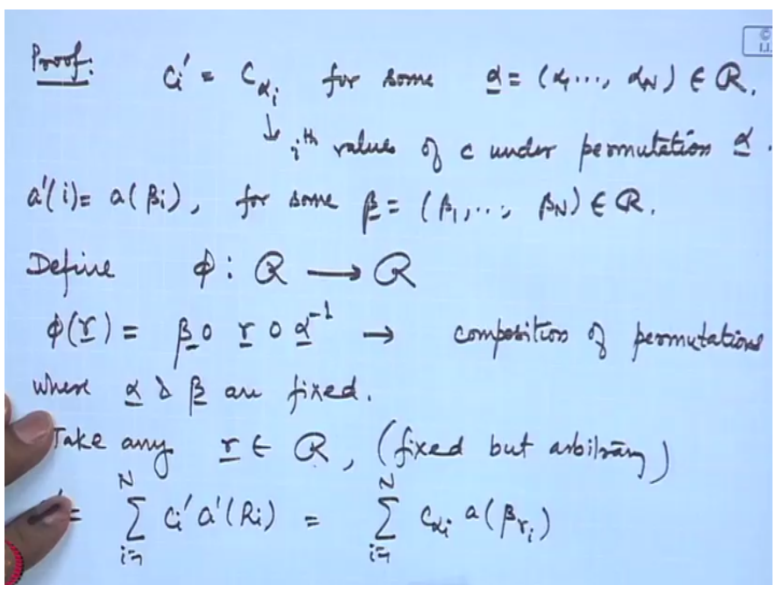


Hence square root n/mn W* has the same limiting distribution as root n/mn Vp. So now if we write U*-expectation U*/square root variance of U* and apply Slutsky's theorem along with limiting normality of square root n/mn W*. We get the result. So thus we have obtained the asymptotic distribution of the Wilcoxon rank sum statistic and the Mann-Whitney U statistic and both are found to be asymptotically normal.

Now in the general linear rank statistics, we are writing the statistic of the form sigma of Ci A of Ri. Now in this one, we may consider some sort of permutation of the ranks or you can say permutation of the indices. Then what happens to the distribution. Our next result is regarding the distribution of the permuted form of this. So if C1 prime, C2 prime, Cn prime is fixed permutation of C1, C2, Cn.

A prime 1 and so on A prime n, this is a fixed permutation of A1, A2, An, then S=sigma Ci A (Ri) i=1 to n is having the same distribution as S prime = where the vector of ranks has uniform distribution over the set of all permutations of the numbers 1 to n. So this is on the testing result and it allows us to use the ranks in it. That means basically the way the data has been obtained, it will not matter when we consider the distribution of the linear rank statistics, which is based on that. Let me give the proof of this here.

(Refer Slide Time: 17:53)

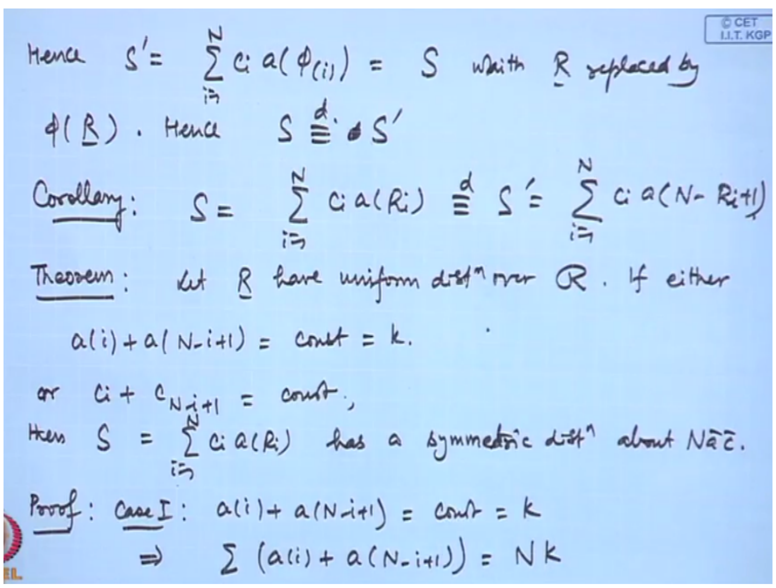


Since C1 prime, C2 prime, this is a permutation, so we can write C1 prime as some C of alpha i for some alpha = alpha 1, alpha 2, alpha n belonging to R. Basically this means that it is the i-th value of C under permutation alpha. Similarly, we can consider A prime i as = A of beta i. This is for some permutation beta of the numbers 1 to n. Let us define a function from R to R as phi R = beta composition with R alpha inverse.

Here the alpha and beta are this. This is actually the composition of the permutations. So you can look at it like this, that R is a vector in R, that means it is permutation of the numbers 1 to n. On that we apply beta from the left end and alpha inverse from the right end. So here alpha and beta are fixed. As we have mentioned here that these are fixed permutations. So for fixed permutations, this result is being proved.

We have already fixed alpha and beta here. So now let us consider. Take any R belonging to R which is arbitrarily fixed. So S prime that is based on i=1 to n=sigma C alpha i A beta Ri, i=1 to n. Why this is so, because A prime i=A of beta i. So if I am writing Ri here, then this will become beta Ri here. So this I can now write as sigma of Ci A(beta R) since I have changed alpha to i that means I have taken the inverse transformation for i, then this will become alpha i inverse, i=1 to n. So this has then become = sigma Ci A (phi i) i=1 to n.

(Refer Slide Time: 21:35)



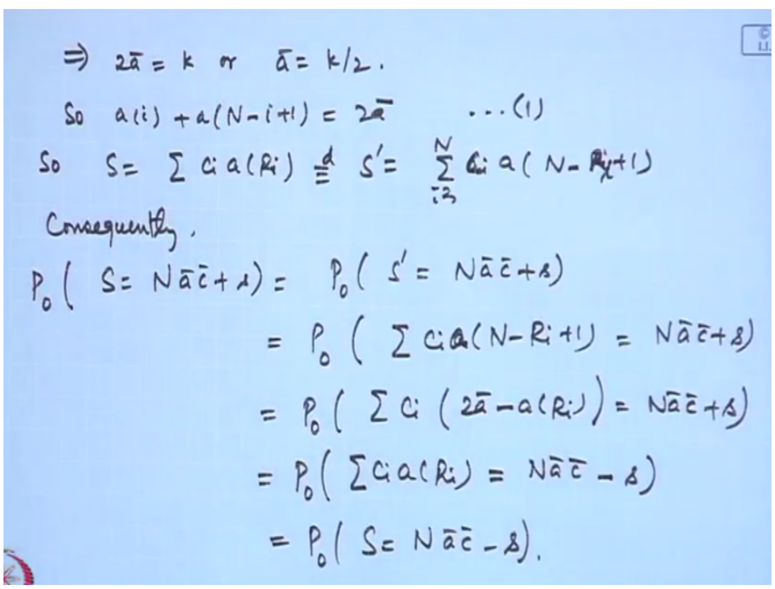
Hence S prime = Ci A (phi i) i=1 to n = S with R replaced by phi R. Hence S will have the same distribution as S prime. So we denote by this, S and S prime have the same distribution. Let us repeat the argument here. I am expressing S prime, which is sigma Ci prime, A prime Ri as here Ci prime has become Ci again and here A prime Ri becomes A of phi i here. So you can say that phi is a 1 to 1 function. Because what is happening there is R.

R is transferred using beta and alpha inverse. So for a given R phi R is uniquely defined. If that is so, then basically the original combination will be preserved here for the distribution. That means whatever probability we are saying for that particular thing, it will remain the same. As a corollary if we consider like we are going 1 to n and then we take from the reverse side, so if we consider the permutations, which are counted from the left hand side, if you count from the right hand side, then the distribution must be the same.

So as a corollary we have the following result, that is the sigma Ci A (Ri). This will have the same distribution as S prime = sigma i=1 to n Ci A (n-Ri+1). As a consequence, we can prove another important theorem. Let R have uniform distribution over R. That means we are considering each permutation is equally likely. If either Ai + A n-i+1 = a constant say Kor Ci+C (n-i+1) is a constant, then S=sigma Ci A (Ri) has a symmetric distribution about n A bar C bar.

We will take both the cases. Firstly, when Ai+A (n-i+1) is a constant and secondly the case when Ci+C (n-i+1) is a constant. So Ai+A (n-i+1) that is a constant = K. This implies sigma Ai+A (n-i+1) = Nk.

(Refer Slide Time: 25:51)

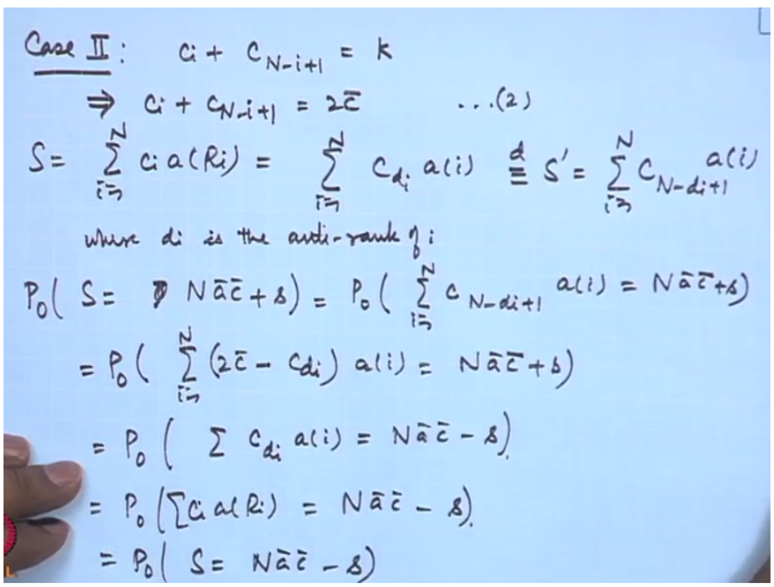


This implies 2 A bar = Kor A bar = K/2. So Ai + A (n-i+1) = twice A bar. Let me call this relation #1. So S = sigma Ci A Ri, this is having the same distribution as S prime, that is = sigma Ai Ci A (n-Ri+1). So as a consequence, let us consider the probability of S = N A bar C bar + S. That is probability of S prime = N A bar C bar + S. That is probability of sigma Ci A (n-Ri+1) = N A bar C bar + S. This we can write as probability of sigma Ci.

And here we change A (n-Ri+1) as 2 A bar – A (Ri) using this relation here. Because Ai+A (n-i+1) = 2A bar = N A bar C bar + S, so that is = probability of sigma Ci A (Ri) and this becomes N C bar, so twice N A bar C bar, this you bring to the left hand side, and you take this term to the right hand side. So you get N A bar C bar – S, which is same as probability of S = N A bar C bar – S. So this will prove that the distribution of S is symmetric about N A bar C bar.

So we have proved this theorem for the case when Ai + A (n-i+1) is a constant. Now let us take the second case.

(Refer Slide Time: 28:27)

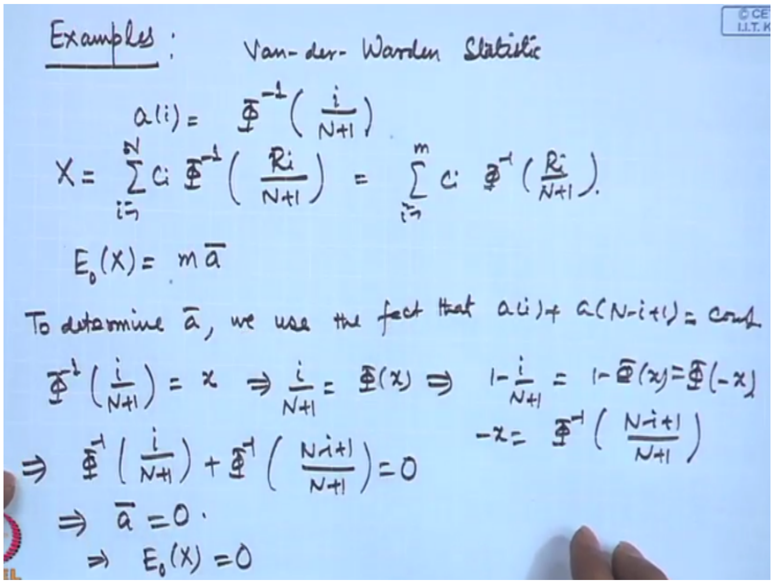


Ci + C (n-i+1) = K, which implies that Ci + C (n-i+1) = 2C bar. This proof will be same because we can sum over all the values, we will get 2K and then since both the sums will be the same, therefore this is = 2C bar therefore K=2C bar. So now let us consider S=sigma Ci A (Ri) i-1 to n. Now this we write as Cdi, A (i) i=1 to n where di is the anti-rank of. Basically what we are doing is that if the i-th observation has rank Ri, so i will have the reverse di.

That is have changed Ri/i so what is the corresponding reverse value here, so that is called Cdi. So this will then have the same distribution as sigma C (n-di+1) A (i), i=1 to n. So if I consider the probability of S=N A bar C bar + S, then it is = probability of sigma C (n-di+1) A (i) = N A bar C bar + S = probability of 2C bar – Cdi A (i), i=1 to n that is = N A bar C bar + S = probability of sigma Cdi Ai, again this term you take to the other side.

So this becomes N A bar C bar –S that is same as saying sigma Ci A(Ri) = N A bar C bar – S. So once again, you are proving that the distribution of S is symmetric about N A bar C bar. We can actually apply this result to various statistics and therefore they can be used for the testing problems in the 2 sample testing problems. Let me give some examples.

(Refer Slide Time: 32:00)

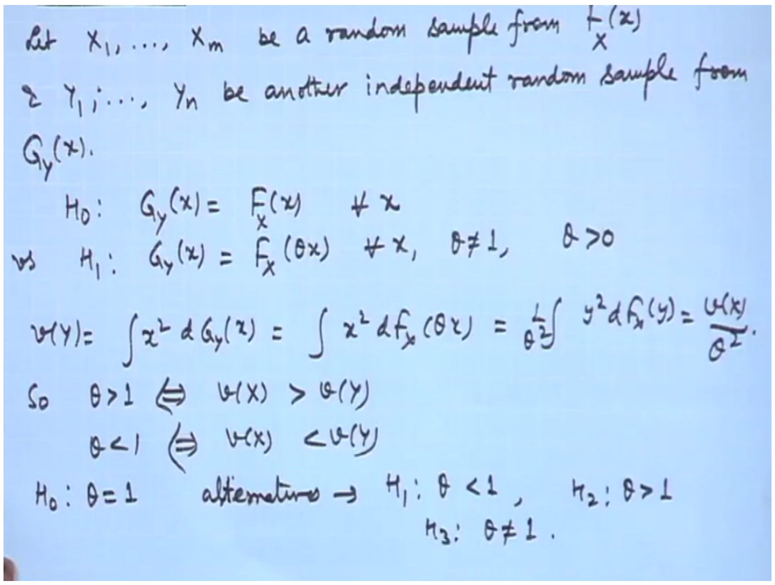


1 is called Van der Waerden statistic. Here the scores are taken as based on the CDF of the standard normal distribution, that is phi inverse i/n+1. So if we consider the statistic as sigma Ci phi inverse i (Ri/n+1), i=1 to n. So this will be = sigma Ci because for n+1 up to n, this will be 0. This is i=1 to m Ci phi inverse Ri/n+1. So expectation of x under the null hypothesis is = mA bar. Actually you can determine A bar here.

Here if we consider the property of the standard normal CDF here that is to determine A bar, we use the fact that A (i) + A (n-i+1) that is constant. If I write say phi inverse i/n+1 = some x, then this will mean that i/n+1 = phi of x, this will mean that 1-i/n+1=1-phi of x = phi of (-x). So you will get –x=phi inverse of n-i+1/n+1. So what do you get then. Phi inverse i/n+1+phi inverse n-i+1/n+1 = 0. That means this constant is actually becoming = 0.

This means that your A bar is 0 and therefore you will have expectation of x under the null hypothesis that is also 0. That is mA bar. We can also write the expression for the variance of x that is mn/N*N+1 sigma i=1 to n, phi inverse i/n+1. This kind of statistics are quite useful for Gupta 2 sample testing problems. Let me also introduce the scale problem here and the 2-sample scale problem.

(Refer Slide Time: 35:28)

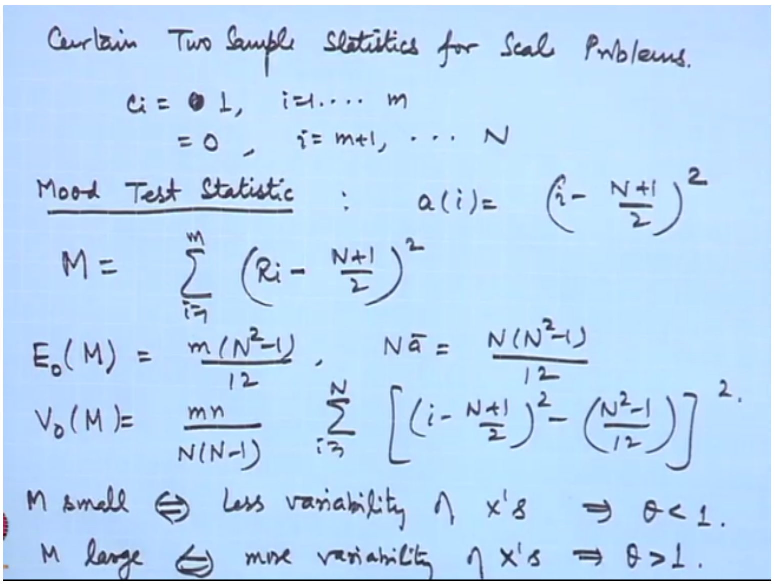


So we have a random sample, let X1, X2, Xm be a random sample from the CDF, F(x) and Y1, Y2, Yn be another independent random sample. This is from Gyx. So our null hypothesis is whether the 2 distibutions are identical and alternatives that this is theta x for all x where theta is != 1. So this is basically the scale model because I have introduced a scale parameter here. So when you have theta = 1, then the 2 will be same, so that is the null hypothesis.

We may also consider it in terms of the variability. So if we consider say X square dGy=X square dFx theta x = 1/theta square y square dFxy, so that is = Vx/theta square. So theta>1 will imply that Vx>Vy and theta<1 will imply that Vx<Vy. In some sense, we can say that this testing problem is equivalent to testing, which distribution has more variability. That is the distribution of x or the distribution of y.

So basically we can consider this null hypothesis, and alternatives will be theta<1 that means whether the variability of X is less than the variability of the distribution of Y or theta>1 that means whether the variability of X is more than the variability of Y or simply say that the variability of X is different from the variability of Y. So all the 3 alternatives can be considered here. So some of the 2 sample statistics that are introduced for the scale problem, they are as follows. Let me give few of them.

(Refer Slide Time: 38:37)

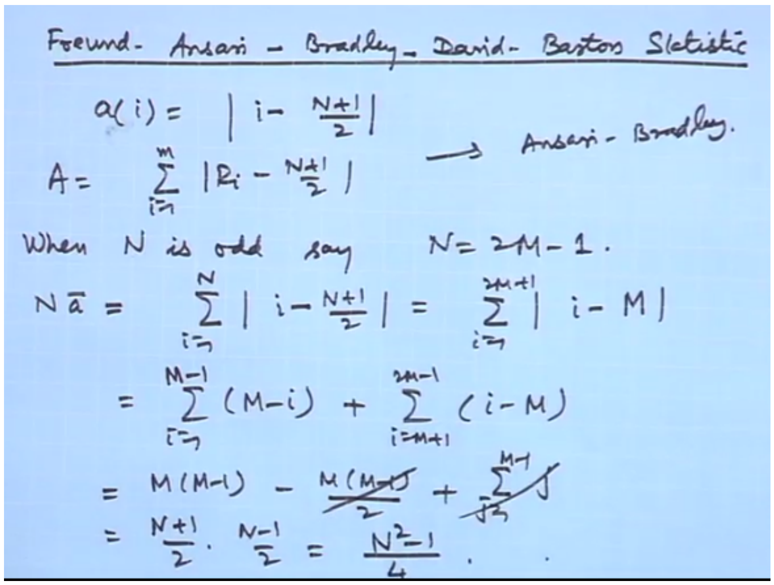


Certain 2 sample statistics for scale problems. So here I am taking Ci=1 for i=1 to n and =0 for i=n+1 to n. So when we are mixing the 2 samples, I am assigning the value 1 and in the second 1, I am assigning the value 0. So 1 is Mood test statistic. In Mood test statistic, we take the scores as i-N+1/2 square. You can easily understand what does it represent here. It will represent that how much difference each value basically when we put ARi.

So N+1/2 is the mean rank here. So how much each rank is different from this 1. So this is a major of variability of the ranks and therefore if I consider by statistics based on that. For example, if I write the Mood statistic as sigma Ri - N+1/2 square, i=1 to m. So basically this is the rank of Xi in the sample, so if the ranks are closer to the mean value of all the, that means the sample is well mixed up that is x's and y's are well mixed up and therefore it will mean that theta is closer to 1.

Whereas the more variability will imply that m is large. So we can consider here, expectation of M=MN square-1/12, Na bar will be equal to N*N square-1/12, and variance of N=mn/N*N, I am not giving the derivations here. But this can be done in a easy way, M small is equivalent to less variability of x's that is theta is < 1, and M large will imply that more variability, that is theta is > 1. So this can be used for testing, the test of hypothesis here.

(Refer Slide Time: 42:00)

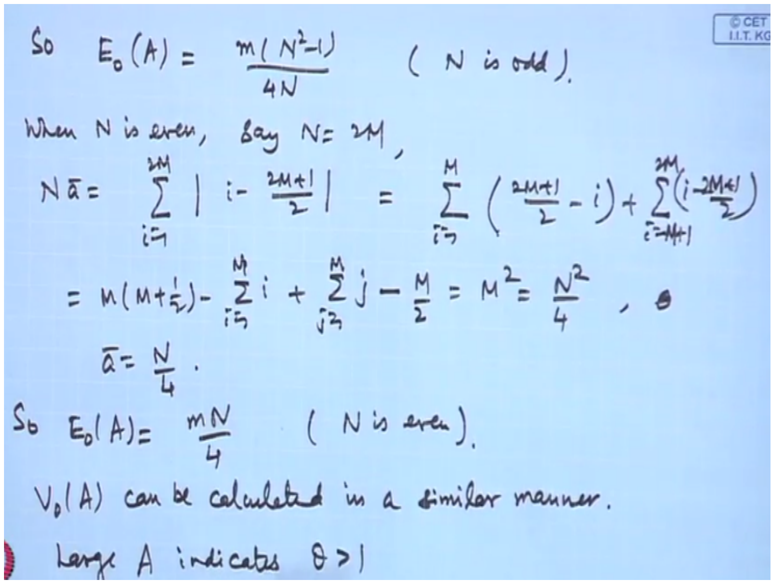


We can also consider another statistic which is named by several authors actually Freund, Ansari, Bradley, David, Barton. So if you look at this Mood statistic, here it is taking the square deviation here, so from the square deviation we take the absolute deviation, then we get this statistics. So these are all very natural choices for their core functions here. So Ansari, Bradley is given by sigma Ri-N+1/2, i = 1 to m.

Actually there are several variations of this that means whether you take directly like this or so. So this one is actually Ansari, Bradley choice. Let us take the case when N is odd, that means N=2M-1 kind of thing. If that is so, then Na bar, that is=sigma modulus i-N+1/2, that is becoming sigma i-M, that is 2M-1+1/2, so i=that is=sigma M-i + sigma i-M. So i=1 to m-1, and this is from M+1 to 2M-1, corresponding to i=m this term becomes 0.

So if you look at this, this is becoming M*M-1/2 and this will also give the similar thing -M*M-1/2+sigma j, j= 1 to m-1. So this actually gets cancelled out, so you simply get M*M-1, which is N+1/2 N-1/2, that is N square-1/4 here. So if Na bar is this one, then we are able to get the value.

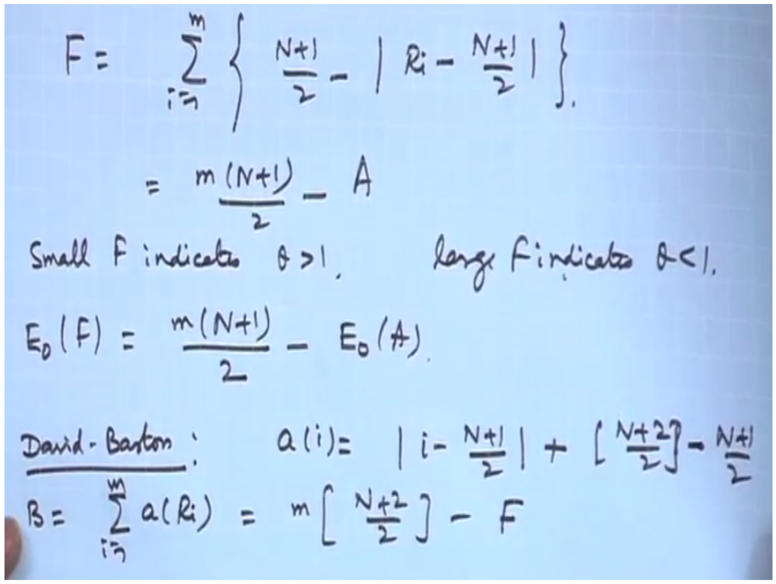
(Refer Slide Time: 44:52)



Expectation A that is = mN square-1/4N, of course this I have done for N odd. If N is even, say N=2M, in that case you consider Na bar that is = sigma i-2M+1/2, that is N+1/2, i=1 to 2M. So again we split in to 2 parts i=1 to N, then this is = 2M+1/2-i+sigma i-2M+1/2 for i = M+1 to 2M. Once again we can easily simplify these terms, it becomes M*M+1/2-sigma i, i=1 to M, +sigma j, j=1 to M-M/2, that is M square, that is N square/4.

So a bar is then = N/4, so expectation of A in this case becomes = mN/4. So variance of A can be calculated, once again if actually the interpretation of this is same as the Mood test statistics, because if A is large it will mean that there is more variability in the X data, that is theta is > 1, so basically we say that large A indicates theta>1, small A indicates theta is <1. Now some variation of this form is given here, see you are considering the absolute derivation here. So and we are taking direct sum here.

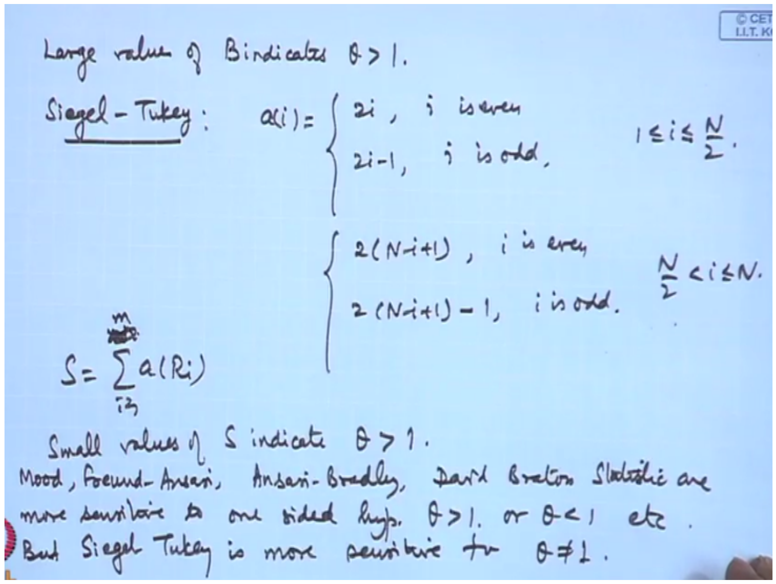
(Refer Slide Time: 47:30)



We may also consider the reverse form ai=N+1/2- as you can see the logic behind it, that this is just a little variation from this. This form is actually called Freund-Ansari form. So this function let me write i=1 to m, N+1/2-Ri-N+1/2, so that is = m*N+1/2- basically Ansari-Bradley. So it is just the reverse 1. So small F indicates theta>1, and large F indicates theta is<1. Expectation of F will be m*N+1/2-expectation of A, and the expectation term, I already calculated here when N is odd then it is m*N square-1/4N, and when N is even it is = mN/4.

There is yet another variation of this which is David-Barton variation. You can consider ai as this is the Ansari Bradley choice, but you shift it little bit, that means basically it is just adjustment of the even and odd values. So basically, B = sigma a of Ri, i=1 to m, that is = m*N+2/2 - the Freund Ansari choice, that is this choice here, because if I take this - this - in the bracket, then this will become the Freund Ansari choice. Therefore, once again we can obtain, this is also the N+2/2-m*N+1/2+Ansari Bradley choice.

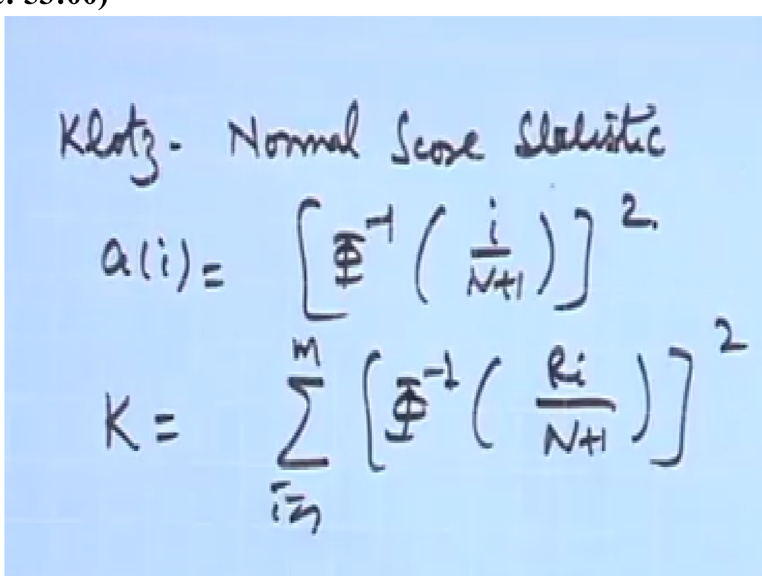
(Refer Slide Time: 50:24)



So here large value of B indicates that theta is>1, and of course small B will indicate theta is<1. Then there is Siegel Tukey choice. In Siegel Tukey choice, we take ai to be = 2i if i is even, it is = twice i-1, if i is odd, for 1</=i</=N/2. And it is = 2*N-i+1, if i is even, and it is = twice N-i+1-1 if i is odd, N/2<i<=N. And S = sigma a (Ri), this is the small m here. Small values of S indicate that theta is>1.

We have one final comment here, that this mood, Freund- Ansari, Ansari-Bradley and David-Barton, these statistics they are more sensitive to 1 sided hypothesis, that is theta>1, or theta<1. But the Siegel- Tukey, this is more sensitive for theta !=.

(Refer Slide Time: 53:00)



Let me briefly mention one more here, that is called the Klotz normal score. In this 1 we define phi inverse i/N+1 whole square. You can compare it with the 1 which I gave earlier, that was phi inverse of i/N+1, Van Der Waerden statistics. In this 1 we had phi inverse i/N+1 and here you can see this is = phi inverse i/N+1 whole square here. So the Klotz statistics is given by phi inverse Ri/N+1 whole square, i=1 to N.

I am not getting to much into detail of the working out of this of course it is slightly more complicated than the Van Der Waerden statistics, because if I take this one I am getting the sum X square+1-X whole square, so that is not a constant here. So this will require some working out to get the outcome of this. In the next class, I will discuss about the (()) (54:21) 2-sample test, I will discuss the null distribution of this and we will also introduce the concept of the consistency of the statistical test.

You might have seen that when we considered the parametric tests, so in the parametric test we discuss about the power of the test and we considered the type 1 error and type 2 error. But when we consider the 2-sample test since we are not having the form of the distribution so we are not using you can say most powerful, that the usual Neyman-Pearson theory is not being applied here.

And therefore the test functions are based on these linear ranking statistics and the exact distribution are quite complicated. So we consider the asymptotic properties of these things. So in the next lecture, I will be discussing about the asymptotic properties of the test here. Firstly, I will discuss about the (()) (55:25) and then we will discuss about other asymptotic properties.