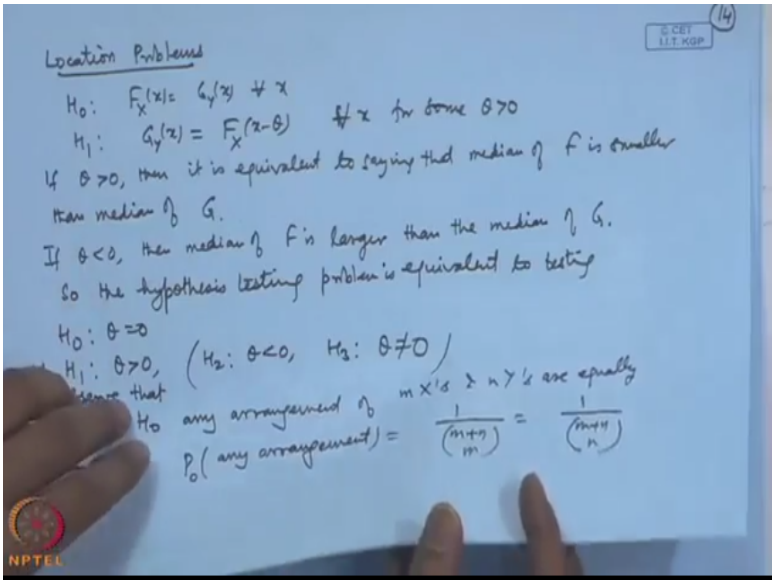
Statistical Methods for Scientists and Engineers
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology - Kharagpur

Lecture - 37
Non parametric Methods - X

Friends in the last class, I had introduced various tests for the single sample location problems and then I had also introduced a 2 sample location problem. Let me recapitulate this thing.

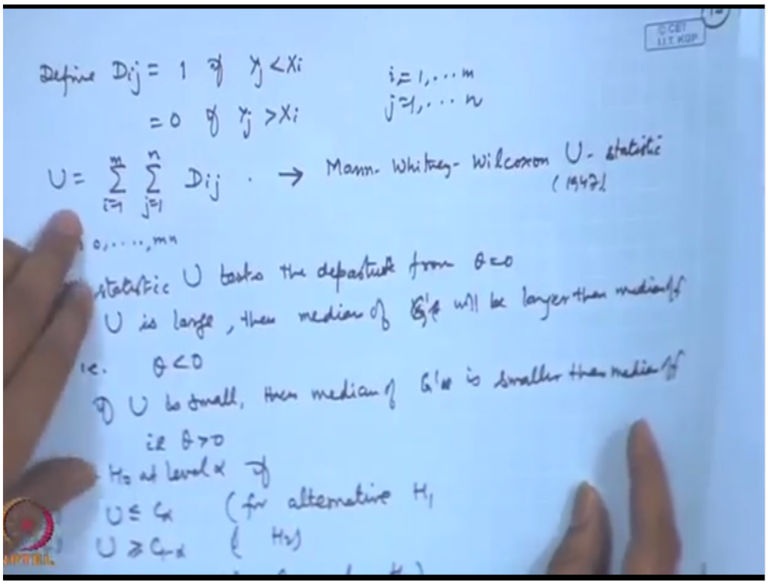
(Refer Slide Time: 00:35)



We have 2 distributions F and G and so we want to basically check whether one of the distributions is a location shift from the other one. So if we consider say theta>0 theta<0 or theta != 0, it is meaning that the median of the distribution of F is either smaller than the median of G or it is larger than the median of G or it is simply != the median of G.

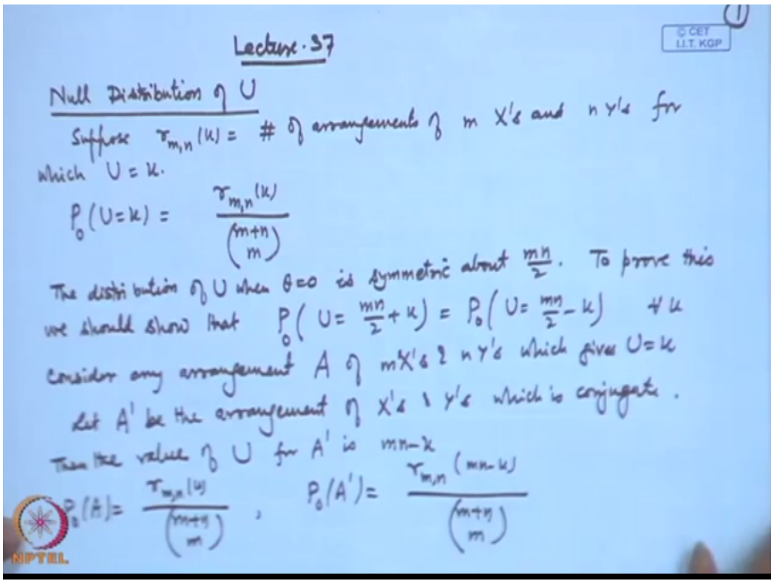
Now for this one we had proposed a 2 sample test based on the observations X1, X2, Xm from F distribution and Y1, Y2, Yn based on the G distribution.

(Refer Slide Time: 01:19)



So we had defined a Mann Whitney Wilcoxon U statistic, which is given by double summation Dij where Dij is 1 if Yj<Xi, it is = 0 if Yj>Xi. So if U is large then naturally it means that the median of G will be larger than the median of F. If U is small, so like that we propose the test here. Now we discuss the null distribution of U etc. So let us start the discussion on that.

(Refer Slide Time: 02:00)

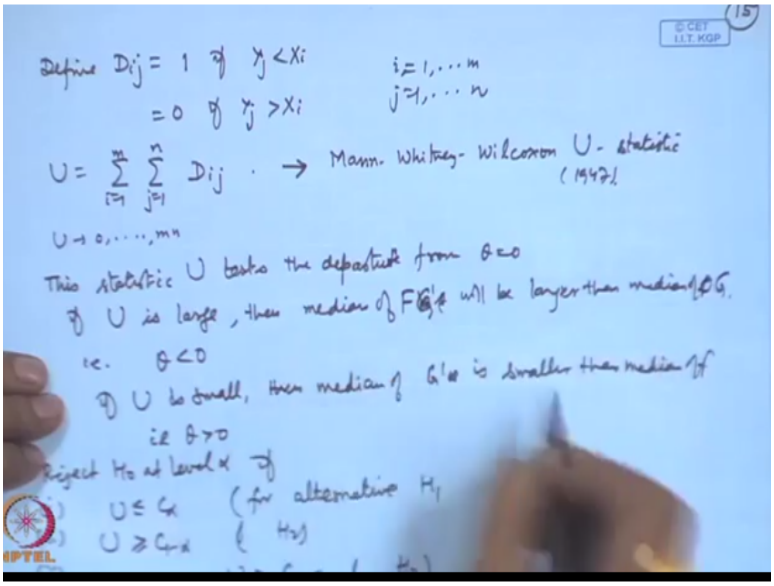


So we start with the null distribution of U. So suppose r m, n u= to the number of arrangements of m X's and n Y's for which U=u. So U consider then it is = r m, n u/the total number of choices m+n C m as we know that the values of u can be from 0 to mn. The first fact that we observe is that the distribution of U when theta=0 is symmetric about the mean value that is mn/2.

To prove this, we should show that P0 U=mn/2+u=P0 U=mn/2-u, for all u it should be true. Now consider any arrangement and I name it as A, arrangement A of m X's and n Y's which gives U=u. Now we consider the conjugate arrangement of X's and Y's that means in which the positions of X's are replaced by the position of the Y's that means the roles of X and Y's are interchanged.

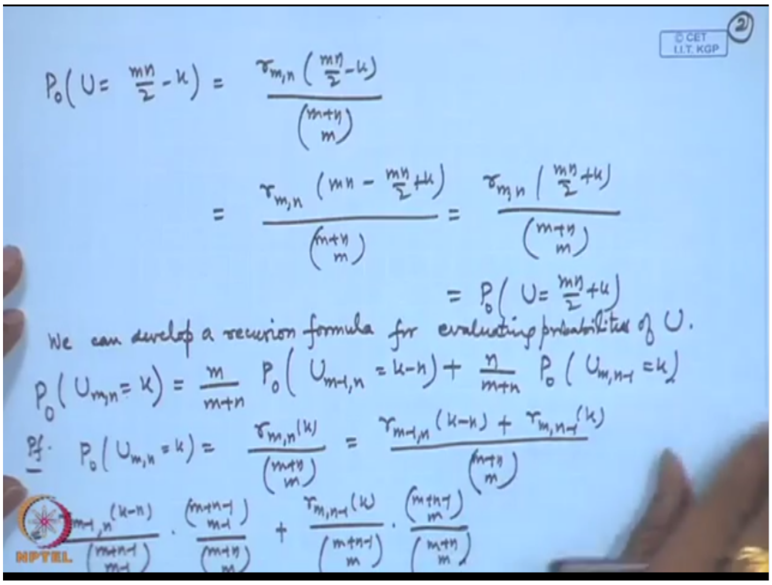
I call that as the arrangement A prime. Let A prime be the arrangement of X's and Y's, which is conjugate. Then the value of U for A prime that will be mn-u because all the X's have become Y's and Y's have become X's, so if we look at this definition here.

(Refer Slide Time: 05:11)



In this definition it will become reverse of that. So if you do that then this will become mn-u. Another point which we have seen that if U is large that means there are more number of Xi's which are larger than the Yj's. Then the distribution of X will be larger than the distribution of the F, so I think I have written it the reverse here, median of F will be larger than the median of G. So we can make a correction in that way. Now if I consider P0 A=r m, n u/m+n C m, P0 A prime= r mn mn-u/m+n C m.

(Refer Slide Time: 06:23)



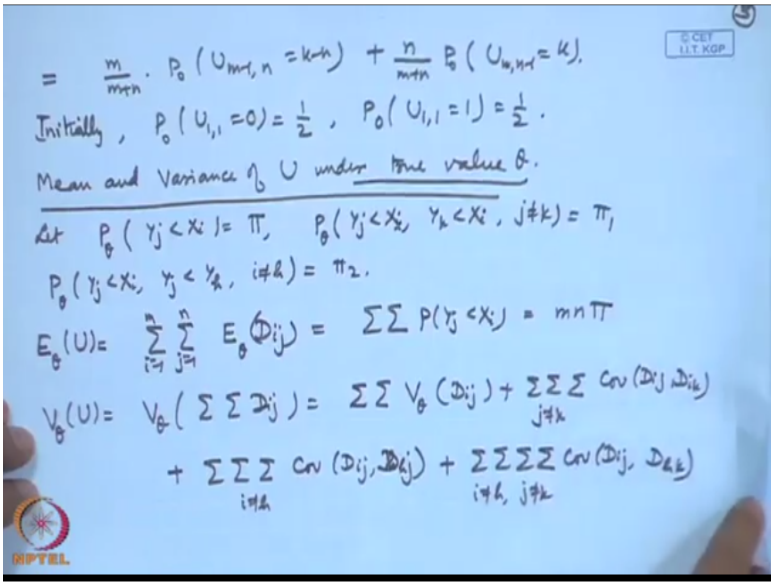
So P0 U=mn/2-u=r m, n mn/2-u/m+n C m. Now that is = r m, n mn-mn/2+u because here we have just interchanged them divided by m+n C m=r m, n mn/2+u/m+n C m=probability of U=mn/2+u. So we have proved that the distribution of U is symmetric about mn/2. To derive the distribution of U actually in general if I have m and n values, then what is the probability of U=u as I have written it is r m, n/m+n C m.

So if I take any values of mn then it is quite complicated because (()) (07:44) the number of permutations will be very large. So we can develop a recursion formula for this. We can develop a recursion formula for evaluating probabilities for distribution of U that means if I consider P0 U m, n that means based on mn observations m observations from X and n observations from Y.

Then it is m/m+n P0 U m-1, n=u-n+n/m+n P0 U m, n-1=u. Let us look at the proof. If I consider P0 U mn=u that is r m, n u/m+n C m that is r m-1, n u-n+r m, n-1 u/m+n C m because if the last observation is X and it is the largest then the value will increase by n otherwise it will remain the same. So either it is u-n or it is u in the previous step.

So then that is = now this you can adjust as r m-1, n u-n/m+n-1 C m-1 m+n-1 C m-1/m+n C m+r m, n-1 u m+n-1 C m m+n-1 C m/m+n C m.

(Refer Slide Time: 10:30)

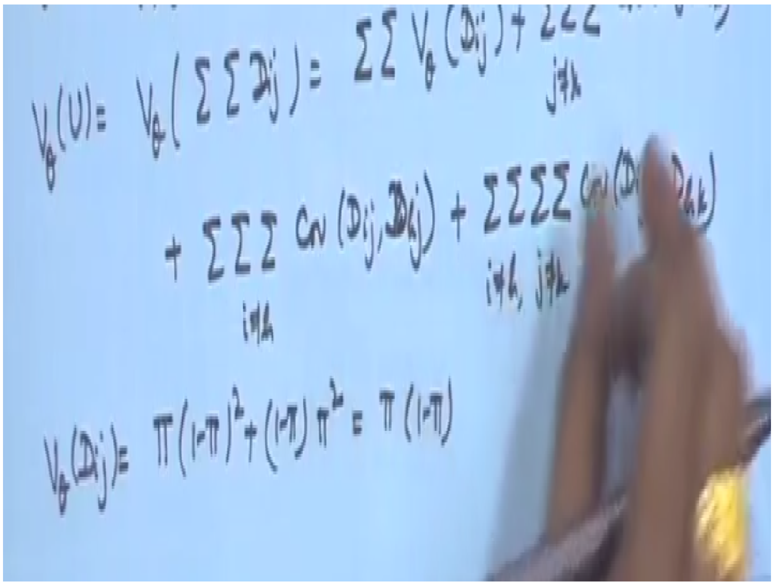


So these 2 terms can be simplified and we get this = m/m+n P0 U m-1, n=u-n+n/m+n P0 U m, n-1=u and for evaluation for higher order thing, we look at what is U 1, 1? U 1, 1 can take 2 values 0 and 1, so it will be 0 with probability 1/2 and P0 U 1, 1=1 that will be with probability 1/2. Now let us look at the mean and variance of the u statistic under general hypothesis that means when the true parameter value is theta.

So since it is dependent upon the probability of Yj<Xi or when we have 2 then Yj<Xi, Yk<Xi or Yj<Xi Yj<Xh so we have to give some notation to that. Let us consider say P theta Yj<Xi=pi P theta Yj<Xi Yk<Xi for j != k=pi 1 P theta Yj<Xi Yj<Xh for i != h=pi 2. In this one j is same and here i is the same. So let us consider the expectation of U=double summation expectation of Dij i=1 to m, j=1 to n.

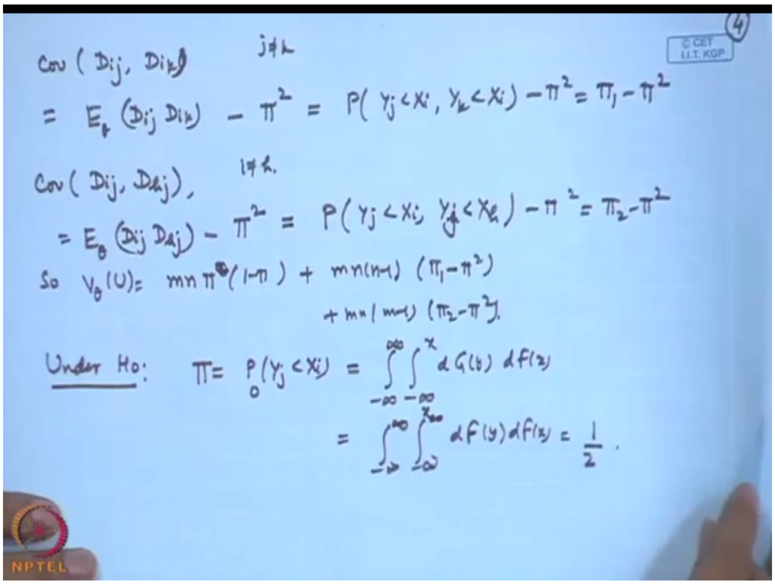
So Dij will be 1 when Yj<Xi this probability=pi=simply mn pi. Similarly, if I look at variance of U=double summation variance of Dij+covariance between Dij Dik. j is != k and then there will be other terms also like there will be terms covariance between Dij Dhj+i != h, j != k, covariance between Dij and Dhk.

(Refer Slide Time: 14:06)



So variance of Dij that is becoming pi*1-pi square+1-pi*pi square=pi*1-pi. Let us look at the various covariance terms here, this covariance, this covariance and so on.

(Refer Slide Time: 14:38)



Covariance between Dij and Dik where j is != k. Then this is = expectation of Dij Dik-individual expectations that is expectation of Dij*expectation of Dik that is pi square. Now this term is going to be 1 when you have Yj<Xi and Yk<Xi for j != k but this value we have assumed to be pi 1 so this is becoming pi 1-pi square. Similarly, if I look at covariance between Dij and Dhj term where i is != h.

Then this is equal to expectation of Dij Dhj-expectation Dij*expectation Dhj both are pi so this is becoming pi square. Now this value will be = 1 only if you have Yj<Xi and Yj<Xh. So

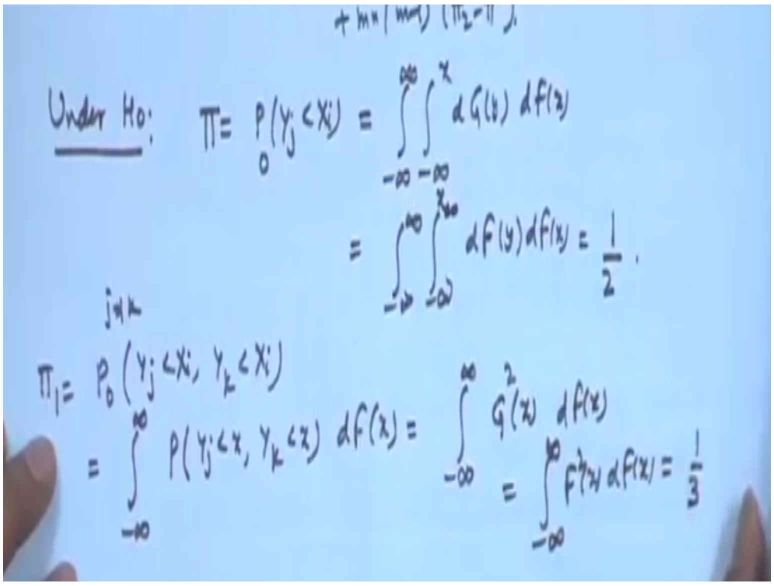
this value we have assumed to be pi 2. So that is = pi 2-pi square. So now if we look at variance of U after substitution of all the terms here and of course this last one will be 0 why?

Because this is involving Yj and Xi and this is involving Yk and Xh since X1, X2, Xm, Y1, Y2, Yn they are independent random variables. Therefore, Dij and Dhk will be independent and therefore the covariance between them then will become 0. So then we are left with these terms and let us count how many terms will be coming. So the first one is the sum over all the values.

So that is = mn pi*1-pi+how many terms are here that will = mn in this one mn*n-1 terms, pi-pi square+mn*m-1 pi 2-pi square. Now let us also see what are these values under H0? Under H0 what happens to pi? That is probability of $Y_j<X_i$ that will be simply = 1/2 because this is becoming dGy dFx -infinity to x -infinity to infinity. Under H0 they are same dFy dFx.

So that is equal to simply in the first integral this will give me Fx and then Fx dFx that will give F square x/2 so that is = 1/2.

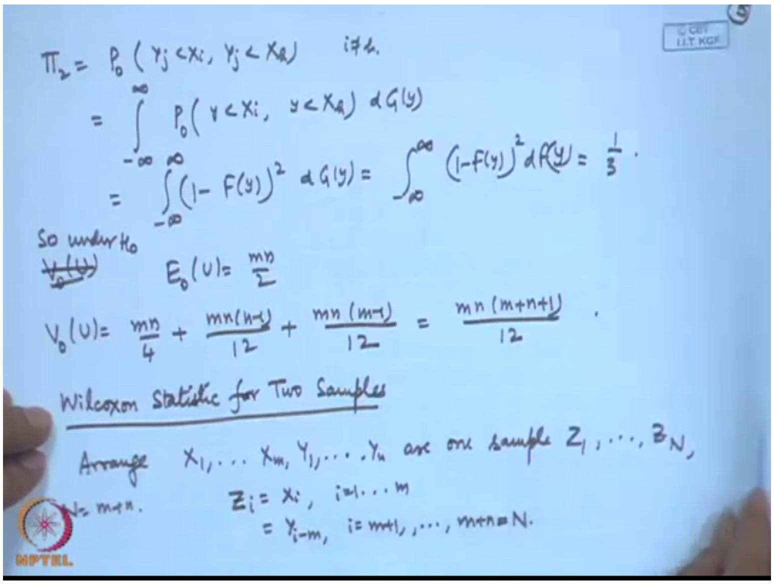
(Refer Slide Time: 18:35)



Similarly, we can evaluate pi 1 and pi 2 under H0, pi 1=probability of $Y_j<X_i$ $Y_k<X_i$ where j is != k so that is = integral probability of $Y_j<x$ $Y_k<x$ dFx. When x is fixed, then Yj and Yk become independent. So this can become equal to the product of these values and that is simply becoming G x square dFx. Now under H0, G=F so it is becoming F square x dFx and this is nothing but 1/3.

Because this is becoming F cube/3 so from –infinity to infinity this will be evaluated to be 1/3.

(Refer Slide Time: 19:42)



Similarly, if we look at pi 2 that is probability of Yj<Xi Yj<Xh where i is != h. Then that is = P0 y<Xi y<Xh dGy=1-F of y whole square because when y is fixed Xi and Xh are independent. So this becomes probability of Xi>y that is 1-Fy and this becomes probability of Xh>y that is also 1-Fy so this becomes square dGy. So when G=F under the null hypothesis, then this is becoming 1-Fy square dFy that is = -1-Fy cube/3.

So at +infinity this will become 0 and at – infinity it will become 1 so this is also = 1/3. So under the null hypothesis when F=G, the value of pi is 1/2, the value of pi 1 is 1/3, the value of pi 2 is also = 1/3 and we can look at the expressions here, 1/3-1/4 if I substitute the values here pi=1/2 then this becomes mn/4. This value will become 1/3-1/4=1/12.

And here also it will become 1/3-1/4 this will become 1/12, so we can simplify so under H0 expectation of U=mn/2, variance of U=mn/4+mn*n-1/12+mn*n-1/12. These can be simplified. This actually becomes = mn/m+n+1/12. So for various purposes this distribution of U can be utilized here. The general use of this 2 sample Mann Whitney Wilcoxon U statistic is to test the location.

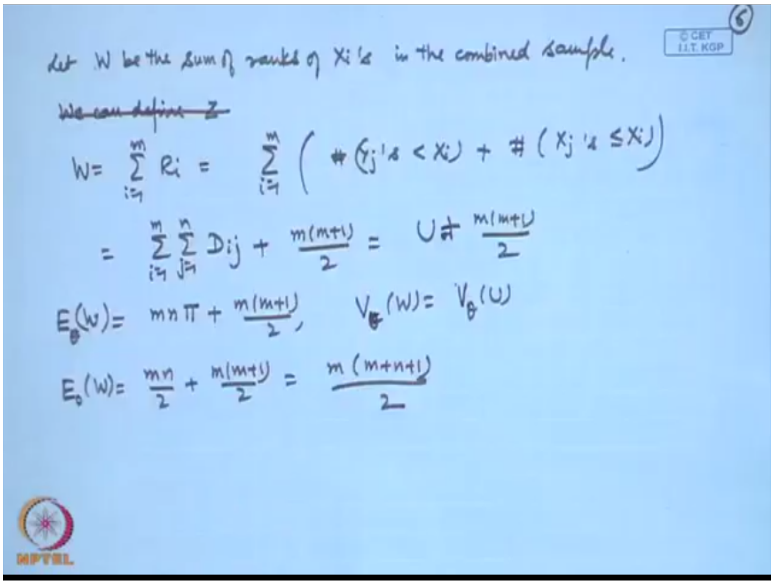
That means whether the median of one of the distributions is larger than the median of the other or less or it is simply !=. We have been able to derive the null distributions so it can be used for several purposes. Now let us consider a variation of this that is called simply the

Wilcoxon statistic for 2 samples. So first we do that we combine all the observations X1, X2, Xm and Y1, Y2, Yn and we treat it as 1 sample.

Let us call it Z1, Z2, Zn. Arrange X1, X2, Xm, Y1, Y2, Yn as one sample say call it Z1, Z2, Zn where N=m+n that means we are saying Zi=Xi for i=1 to m and it is = Yi-m for i=m+1 to m+n=N okay.

Now if the null hypothesis is true that means if the 2 distributions are the same then basically it becomes simply one random sample from the entire population F. Otherwise, there will be some discrepancy that means we are mixing some different kind of things.

(Refer Slide Time: 24:38)

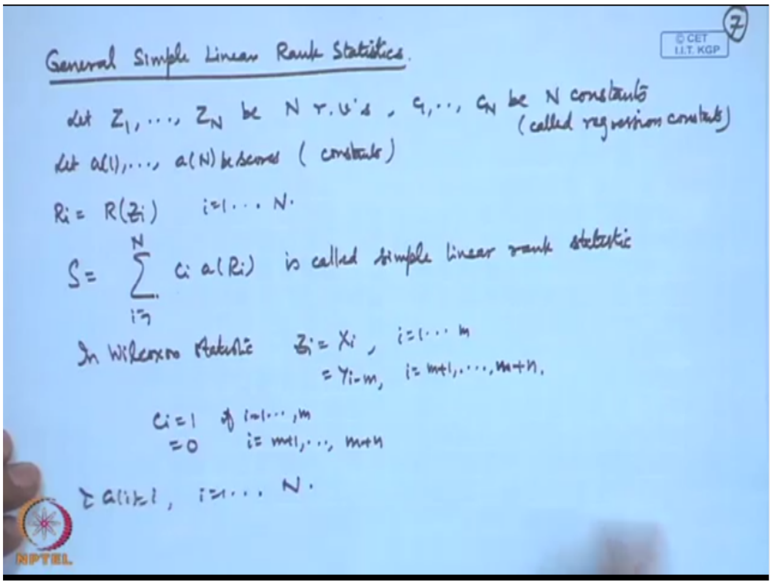


Let W be the sum of ranks of Xi's in the combined sample. So if we consider say W=summation of Ri i=1 to m okay. That is = sigma number of Yj's which are < Xi+number of Xj's which are <= Xi and this we are doing for all i=1 to m. So if I sum this this is nothing but the Dij's i=1 to m, j=1 to n and the second term if you look at when I sum this, this is simply m*m+1/2.

Because what we are doing, how many Xj's are <= 1 particular Xi and this we are doing for every i, then this is nothing but the sum of all the ranks so it is becoming m*m+1/2. So basically you are saying this Wilcoxon W statistic is U+m*m+1/2 so it is simply a shift from u. Therefore, this can also be used for testing the hypothesis here. So we can have in general expectation of this will become = mn pi+m*m+1/2.

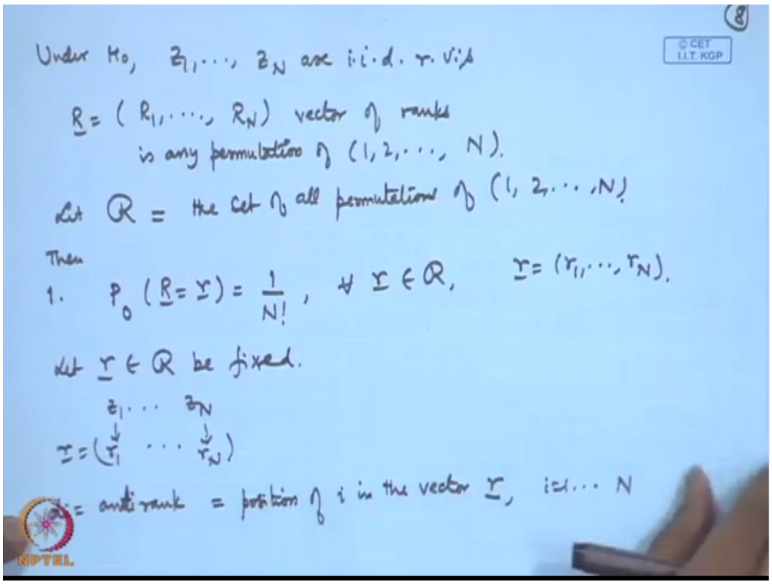
The variance of W will be same as the variance of U because it is simply a location shift. Also the null expectation of this will become mn/2+m*m+1/2=m*m+n+1/2. So the use of Wilcoxon W is same as the use of Mann Whitney U. Both can be used interchangeably. In certain problems it is easier to calculate W rather than the U. Now I consider general simple linear rank statistic for the 2 sample problems.

(Refer Slide Time: 27:28)



Let Z1, Z2, ZN be N random variables. c1, c2, cN be N constants. Here we call them regression constants and let us call a1, a2, aN scores. So these are also some constants, but I call them scores so these have to be chosen. So now let us consider say Ri=rank of Zi, i =1 to N. So then S=sigma of ci a of Ri i=1 to n. This is called simple linear rank statistic. See in Wilcoxon case, we have chosen Zi to be Xi for i=1 to m and it is = Yi-m for i=m+1 to m+n and ci is 1 if i =1 to m and it is = 0 for i=m+1 to m+n and ai=i for i=1 to N.

(Refer Slide Time: 29:47)

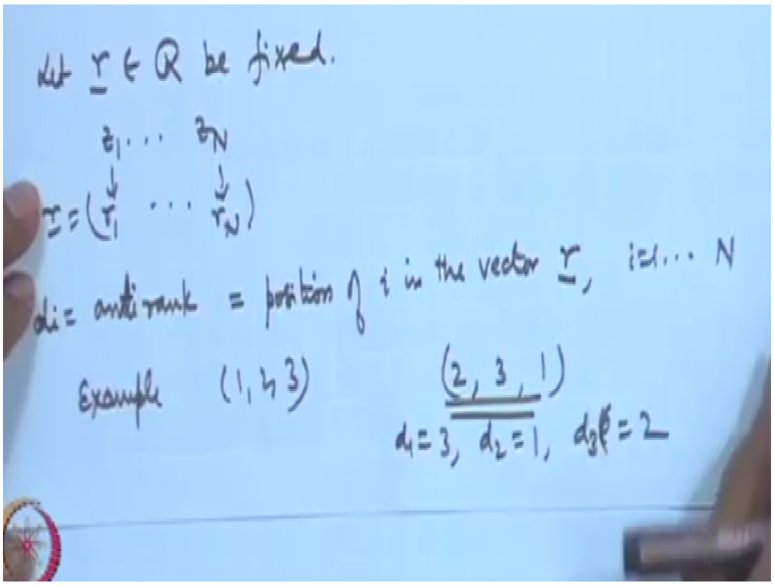


Let us also consider what happens under H0. Under H0, this Z1, Z2, ZN they become independent and identically distributed random variables because they are coming from the same distribution if F=G. So if I consider R=R1, R2, RN that is the vector of the ranks, so this is any permutation of numbers 1 to N. Let us consider say script R the set of all permutations okay of the numbers 1 to N.

Then first result is that under the null hypothesis each of the permutation will be equally likely. Let us look at an elementary proof of this. If we consider say let us fix say some value R as a fixed value in the set of permutations. Now if I am considering say Z1, Z2, ZN then corresponding to this we are having r1, r2, rN these are the ranks here okay. Let us consider di to be the anti-rank so this is the new terminology that I am introducing here.

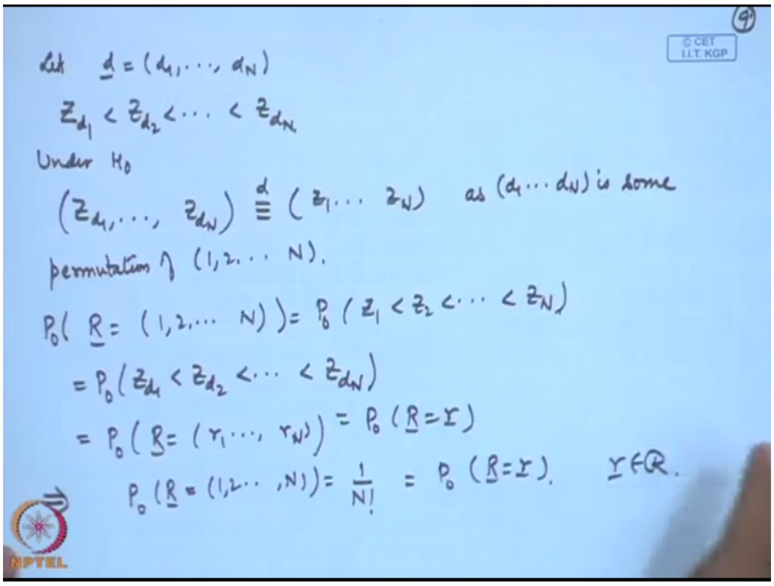
This is nothing but the position of i in the vector r, i=1 to N. See this is like this suppose I consider 3 numbers.

(Refer Slide Time: 32:01)



Suppose I have say 1, 2, 3 okay and I consider say the arrangement of the ranks as say 2, 3, 1 suppose this is an arrangement here. Then what is the anti-rank here? d1 is 3, d2=1, and d3=2. These are the anti-ranks here.

(Refer Slide Time: 32:38)

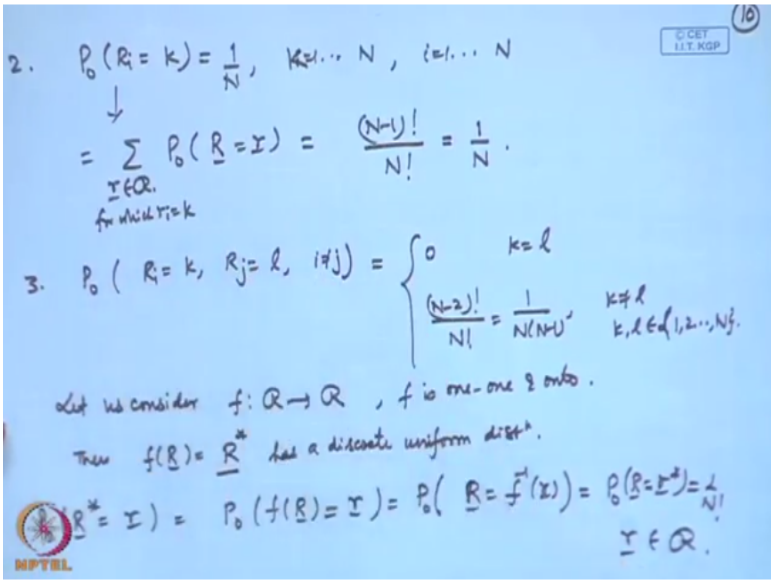


So let us write say d=d1 d2 dN that is the vector of the anti-ranks here. Then what we are saying is that Zd1 is < Zd2 < ZdN. Now under H0, Zd1, Zd2, ZdN will have the same distribution as Z1, Z2, Zn because this is simply one permutation of numbers 1 to N. It is some permutation of numbers 1 to N here. So if I consider say probability of say R=1 to N then that is = probability that Z1<Z2<ZN.

That is = probability of Zd1<Zd2<ZdN because the distributions are the same. But this is nothing but the probability that R=r1, r2, rN that means what I am saying for any permutation

it is equal to the same probability that means each of them will have the equal probability 1/N factorial. So this proves that the distribution of the ranks is discrete uniform distribution over all permutations.

(Refer Slide Time: 34:57)



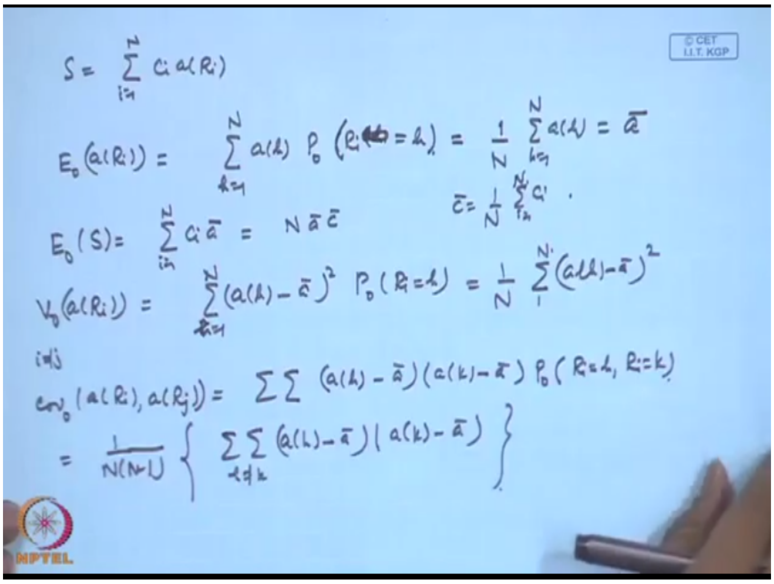
Now we consider individual ranks also. That means I consider the ith rank, then of course this can be from 1 to N then we will prove that it is actually for k=1 to N for i=1 to N. For each of them it will take the same number of values with same probabilities. See how do I derive this? This is equal to the sum over r for which ri=k. So how many such things will be there? It will be N-1 factorial/because 1 rank I am fixing for the ith one and other N-1 positions will interchange.

They can be permutated in N-1 factorial ways so it is becoming simply = 1/N. Similarly, we can consider say probability of say Ri=k or j=l where i is != j. Then of course this is 0 if k=l I am dealing with the continuous distributions so I will not assume that the 2 values can be same because the 2 values will be same with probability 0.

Now if I am fixing 2 values then N-2 factorial/N factorial=1/N*N-1 where k is != l and both k and l can vary from 1 to N. So the joint distribution of 2 ranks can also be obtained and that is also bivariate, you can say discrete uniform distribution. Let us consider say a function which is on 2 function from R to R and also one-one and onto. Then fR let me call it R star.

So this is actually a vector here, so this has a discrete uniform distribution. That is if I consider then that is = P0 fR=r that is probability of R=f inverse r=probability of R=r star=1/N factorial for r belonging to R. So we are able to talk about this basic distribution of the ranks of the observations when I consider the combined samples. So under the null hypothesis it is from the same distribution.

(Refer Slide Time: 38:25)

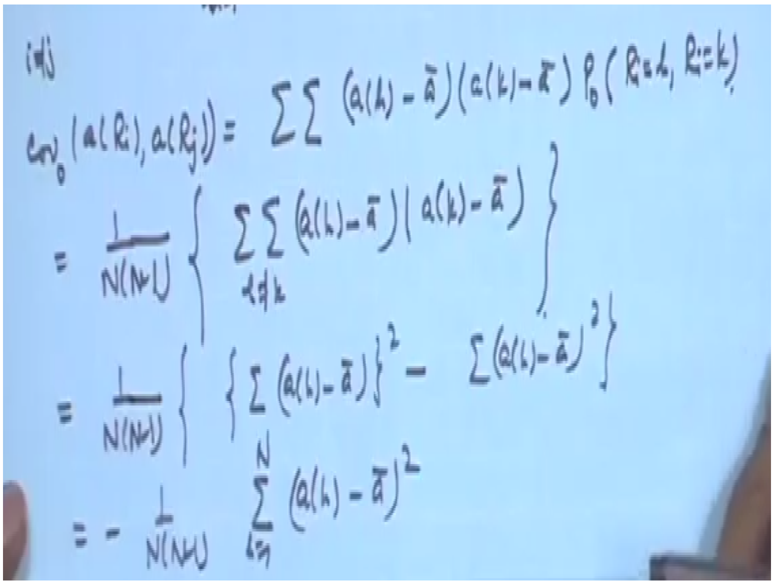


Therefore, these statements are valid. Let us consider now S=sigma ci a of Ri that was our expression for the simple linear rank statistic. So if I consider expectation of a Ri=sigma ah probability that Ri=h. This will be Ri=h, for h=1 to N because Ri can take values 1 to N here. This probability is 1/N so it is simply becoming 1/N sigma ah, h=1 to N. Let us denote this quantity by a bar here.

So expectation of S that is becoming sigma ci a bar, which you can also write as N a bar c bar where c bar is nothing but the mean of ci's i=1 to N here. We can also consider the variance here, so firstly let us consider the variance of a Ri=sigma ah-a bar square probability of Ri=h, h=1 to N. So this is 1/N so it is becoming 1/N sigma ah-a bar whole square, h=1 to N. For i != j, let us consider the covariance between a Ri and a Rj.

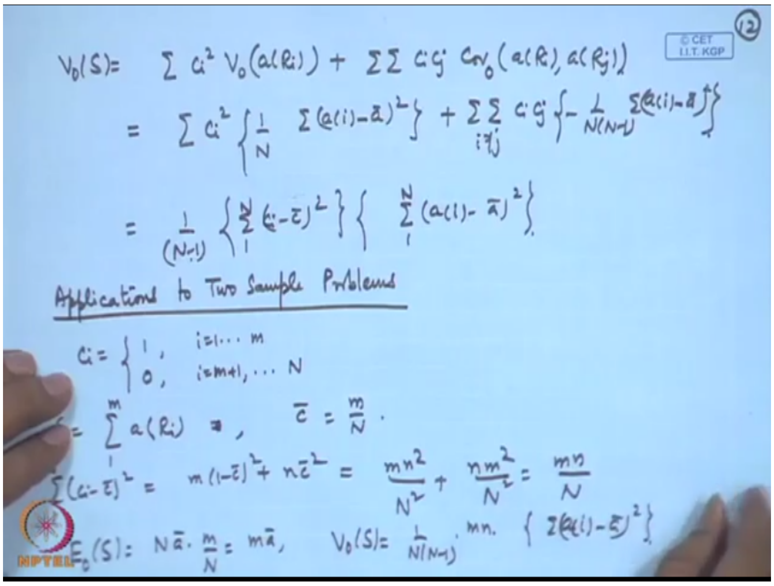
That is equal to double summation ah-a bar ak-a bar probability of Ri=h Ri=k=1/N*N-1 ah-a bar ak-a bar, h != k.

(Refer Slide Time: 41:05)



This we can write as 1/N*N-1. This term we write as the square of the sum-sum of the squares that means it is = sigma ah-a bar whole square-sigma ah-a bar square. So this term becomes 0, so we are left with -1/N*N-1 sigma of ah-a bar square, h=1 to N. So these 2 terms we can use in the variance of S and we will get here.

(Refer Slide Time: 41:53)



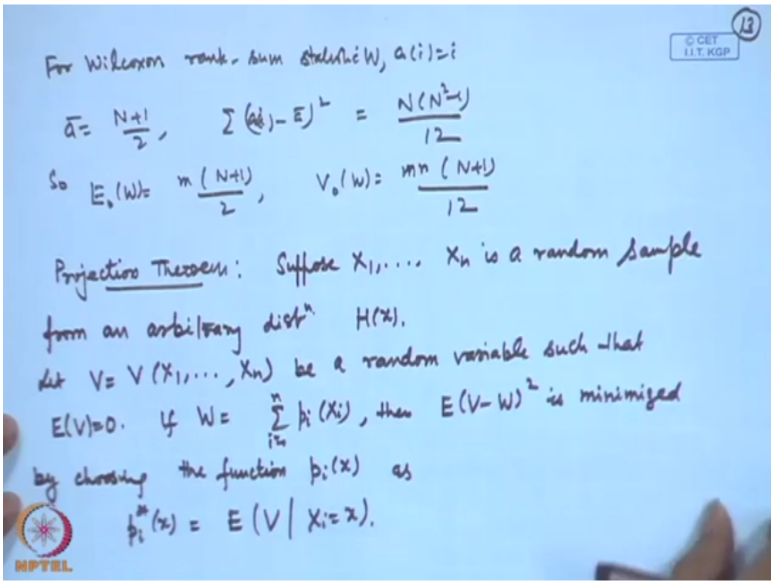
Variance of S=sigma ci square variance of a Ri+double summation ci cj covariance a Ri a Rj. The values of variance a Ri and covariance of a Ri and a Rj have just been calculated. So we substitute here so we get sigma ci square and this term is nothing but 1/N sigma ai-a bar square+double summation i != j Ci Cj that is 1/N*N-1 sigma ai-a bar square. So this is becoming 1/N-1.

See this term I can take outside, so this will become simply sigma of ci-c bar square 1 to N*sigma of ai- a bar square 1 to N. So in the general function that means if I consider general constants and that means regression constants and general score function we can derive the null mean and the null variance of the distribution of the linear rank statistics. As an application you can see to some of the 2 sample problems.

Let us consider some applications to 2 sample problems. Let us consider say ci=1 if i=1 to m and it is = 0 for i=m+1 to N. When S=sigma a of Ri 1 to m and c bar=m/N, so sigma of ci-c bar square 1 to N=m*1-c bar square+n c bar square=mn square/N square+nm square/N square=mn/N. I can take out mn then this will become m+n that is N so N square cancels out to get U mn/N.

So under this if I consider expectation of S=N a bar m/N=m a bar and variance of S=1/N*N-1 mn sigma ai-a bar whole square. For Wilcoxon rank-sum statistic ai=i so if I put that value here what I will get?

(Refer Slide Time: 45:32)

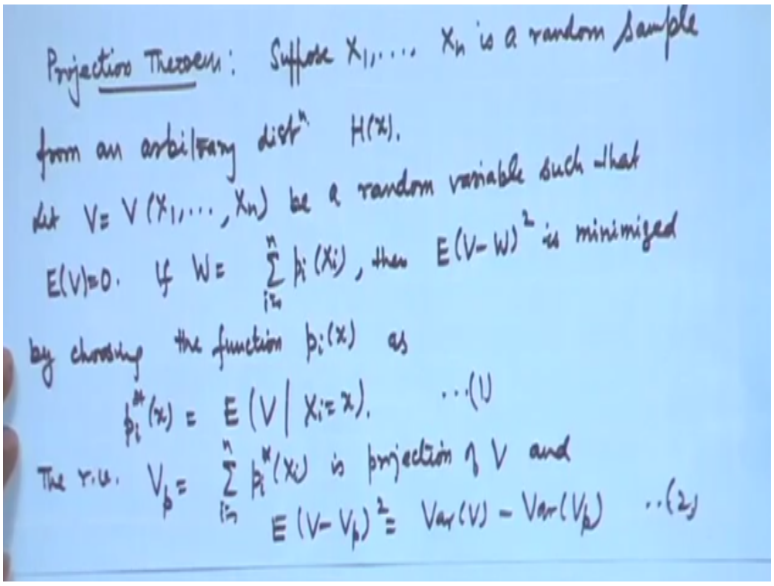


For Wilcoxon rank-sum statistics ai=i so a bar becomes N+1/2 and sigma ai-a bar that will become = N*N square-1/12 that is the mean of the discrete uniform distribution and the variance of the discrete uniform distribution. So if I consider expectation of W=m*N=1/2 and the variance=mn*N+1/12. So if we compare with the values that I derived earlier, you can match here whether it is the same or not.

So if you look at here, this was = m*m+n+1/2 m+n=N so it is the same value. Variance of this was same as the variance of U, which was actually mn*N+1/12 so here also you get mn*N+1/12. So you can see that this general structure helps us to perceive of various other new test statistics that can be utilized for various purposes in the testing problems. Next, we consider the concept of projection.

So we state a theorem here, we call it projection theorem. Suppose say X1, X2, Xn is a random sample from an arbitrary distribution Hx. Let V=V of X1, X2, Xn be a random variable such that expectation of V=0. Now if W=sigma pi Xi i=1 to n then expectation of V-W square is minimized by choosing the function pi x as pi star x=expectation of V given Xi=x.

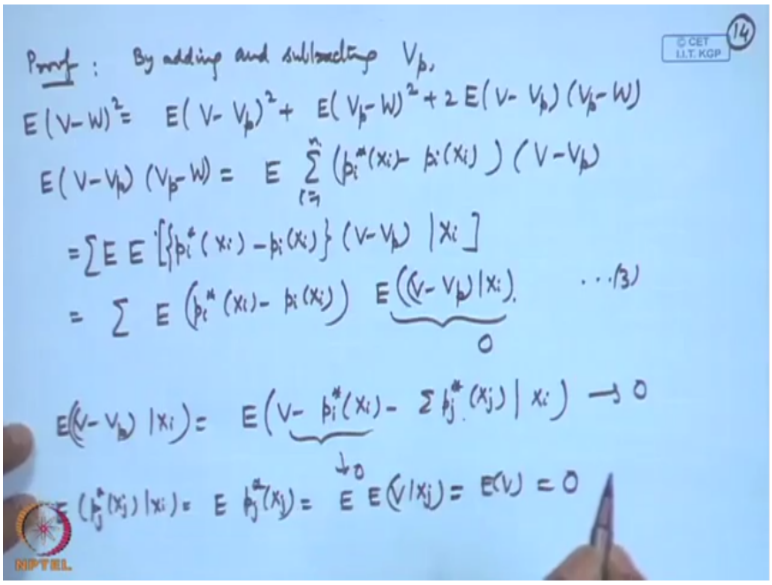
(Refer Slide Time: 49:04)



So this random variable which is obtained as Vp, the random variable Vp that is defined as sigma pi star Xi, this is called projection of V and expectation of V-Vp square=variance of V-variance of Vp. Let me name these relations as 1 and 2 here okay. Let us look at this. What we are saying is that we have the function WS sigma pi Xi, so this is minimized when we consider the conditional expectation of V with respect to Xi.

And then when we do this for every Xi then if we sum it then it is called the projection of V okay.

(Refer Slide Time: 50:14)

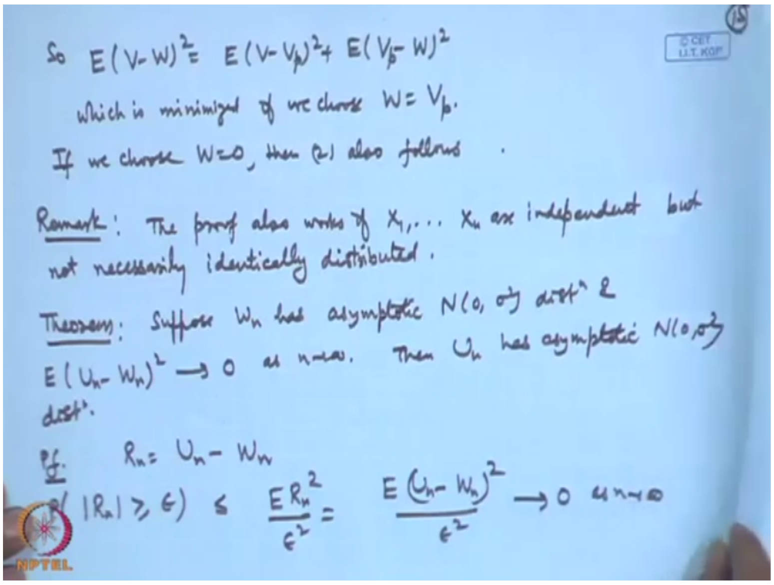


Let us look at the proof of this here. By adding and subtracting Vp, so expectation of V-W square that is = expectation of V-Vp square+expectation of Vp-W square+twice expectation V-Vp Vp-W. So if I look at the expectation of V-Vp*Vp-W=expectation of sigma pi star Xi-pi Xi*V-Vp for i=1 to n. Now this we can write as the summation expectation of expectation pi star Xi-pi Xi*V-Vp given Xi.

So this becomes = expectation of pi star Xi-pi Xi, this term can be separated out, expectation of V-Vp given Xi. Now if we consider expectation of V-Vp Xi, so V-Vp given Xi=expectation of V-pi star Xi-sigma pj star Xj given Xi. Now if we look at the definition 1 here, then this is actually = 0 so this part is 0 and so this part becomes 0 and if I look at this term, expectation of pj star Xj given Xi then what it is equal to?

Expectation of pj star Xj because Xi and Xj are independent so it is equal to expectation of expectation V given Xj that is expectation of V that is = 0 because I am assuming V to be random variable such that expectation V is 0, so this term si also 0. So basically this entire term is becoming actually = 0, this term is becoming 0 so this term is entirely becoming = 0.

(Refer Slide Time: 53:23)

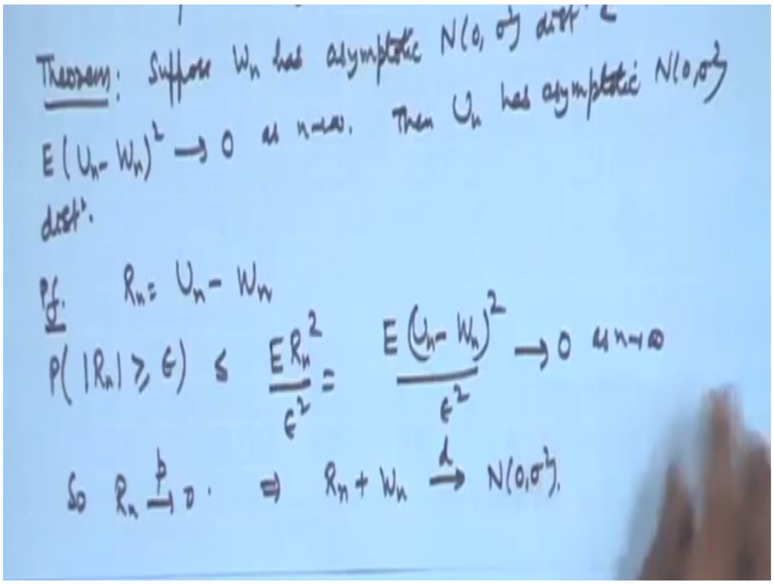


So what we are getting is that expectation of V-W square is nothing but expectation of V-Vp square+expectation of Vp-W square. That means it is the expectations of the 2 positive terms not negative terms. So this is minimized if we choose W=Vp here. If we choose W=0 then the expression 2 also follows. So this completes the proof of this projection theorem. As a remark, let me mention here.

The proof also works if X1, X2, Xn are independent, but not necessarily identically distributed. So in some applications this theorem can be used because when Xi's are coming independently but they are not having the same distribution then also this concept of projection can be used here. So we have the following theorem, which is following from here.

Suppose Wn has asymptotic normal 0, sigma square distribution and expectation of Un-Wn square this goes to 0 as n tends to infinity. Then, Un has asymptotic normal 0, sigma square distribution. For proving let us define Rn=Un-Wn, so probability of Rn >= epsilon that is <= expectation of Rn square/epsilon square=expectation of Un-Wn square/epsilon square. This goes to 0 as n tends to infinity.

(Refer Slide Time: 56:17)



So this proves that Rn goes to 0 in probability. So now you add it here, this implies Rn+Wn that will converge in distribution to normal 0, sigma square. Using these properties, I will be deriving the asymptotic distributions of the Mann Whitney U statistic and the Wilcoxon rank-sum statistics in the next lecture, so that I will be covering in the next lecture.