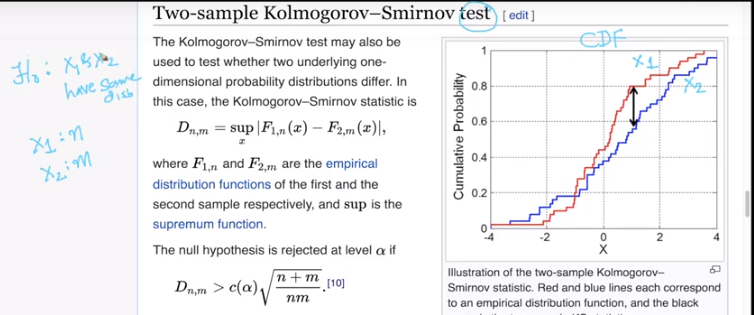
K-S test is used for to test whether two distribution are similar or not.

To understand K-S test for similarity of two distribution we will take an example and understand steps:

Step – 1:

Suppose we have two R.V with n and m no. of observations and we draw CDF for the same

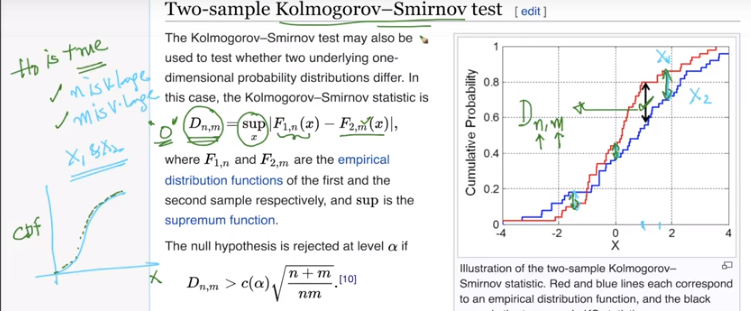


Step-2:

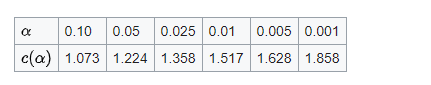
First assume that the values of n and m are too large may be 5000 and 1000 than the CDF for both the R.V will be almost same(IF they are having same distribution) and so the value of D(n,m) will be 0.

NOTE:- D(n,m) is the maximum distance between the CDF of two R.V’s.

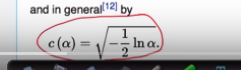
In below graph in image the black line is the D(n,m)



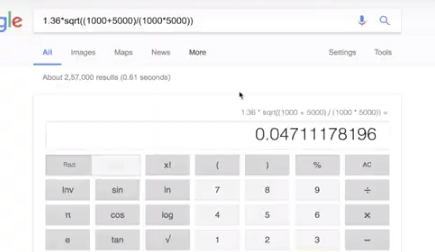
The closed range formula which actually determines whether to accept or reject the hypothesis has variable c(alpha) and we can get its value by looking up into table which is already available.



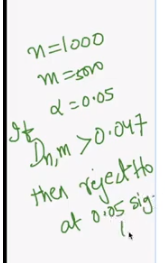
**And c(Alpha) is basically**



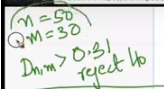
**Now putting the values in formula and computing the Range value for D(n,m)**



**We got D(n,m) > 0.047 which means if this condition is satisfied then we will reject our null hypothesis with 0.05 significance.**



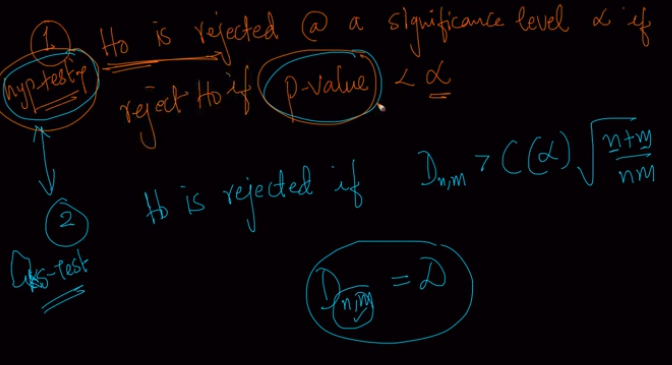
**If we change the value of n and m for understanding the concept better than**

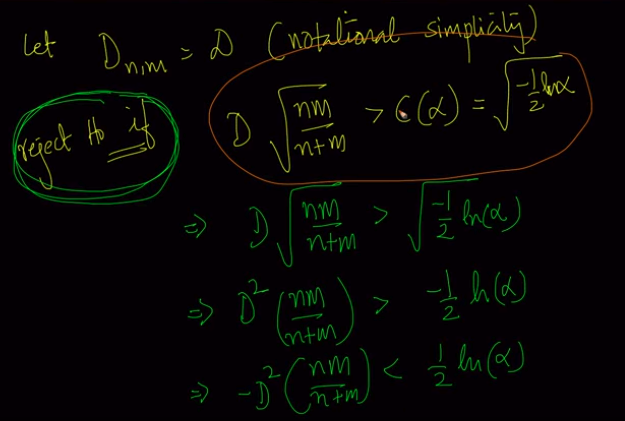


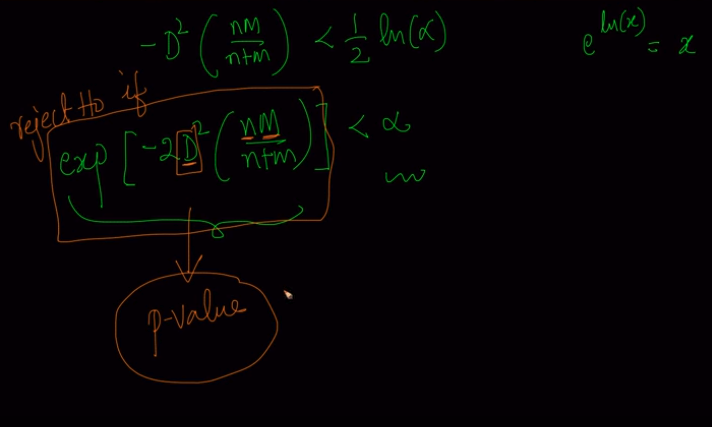
**we get D(n,m) > 0.31 which is almost 30% so this is very big difference we have got which means for 0 till 0.31 value of D(n,m) we will accept H0 and beyond that we will reject H0.**

For more understanding:

Below steps shows how to calculate p-value so that k-s test can be converted to hypothesis test.

3





Here we get our p-value and significance(alpha) can be anything depending upon domain of problem, usaually we keep it as 0.05 or 5%, but for medical purpose we need confirmation of 99%, and that’s why we keep significance as 0.01 or 1%.

**Some comments:**

1. Does QQ plot and K-S test affected by sample size. And what is difference between QQ plot and K-S test.

1. Yes, just like many other statistical methods. More data implies better results.  
2. Both QQ-Plot and KSTest would be impacted by smaller sample sizes. KSTest is a proper statistical test with p-values and a proper test statistic(D) and we use it when we want to make a decision based on numerical values of the p-value. QQ-Plot is a more visual technique which is a great way to communicate your results with others.