

# 1) KS Test

Given  $p(x) = 2x$

Let this target distribution be Y.

Given threshold  $C = 0.25$

Given Samples  $X = \{0.592, 0.774, 0.245, 0.424, 0.685, 0.436, 0.648, 0.959, 0.842, 0.995\}$ .

Null Hypothesis  $H_0$ : The given samples X follow the distribution Y.

Alternate Hypothesis  $H_1$ : The given samples X do not follow the distribution Y.

CDF of Y =  $F_Y(X) = \Pr(X \leq x)$

$$= \int_0^x p(x) dx$$

$$= \int_0^x 2x dx$$

$$= 2(x^2 / 2)$$

$$= x^2$$

X	$F_Y(X) = X^2$	$F^-(X)$	$F^+(X)$	$ F(X) - F^-(X) $	$ F(X) - F^+(X) $
0.245	0.06	0	0.1	0.06	0.04
0.424	0.179	0.1	0.2	0.079	0.02
0.436	0.19	0.2	0.3	0.01	0.11
0.592	0.35	0.3	0.4	0.05	0.05
0.648	0.42	0.4	0.5	0.02	0.08
0.685	0.47	0.5	0.6	0.03	0.13
0.774	0.59	0.6	0.7	0.01	0.11
0.842	0.709	0.7	0.8	0.0089	0.091
0.959	0.92	0.8	0.9	0.12	0.02
0.995	0.99	0.9	1	0.1	0.01

$$\text{Max} (F_Y(X) - \hat{F}(X)) = 0.13$$

$$\text{Max} (F_Y(X) - \hat{F}(X)) < C$$

Therefore, the null hypothesis is accepted.

## 2) Permutation Test

Should we consider/not consider the repeated permutations while performing the permutations test?

Actually, it does not matter if we consider/did not consider the repeated permutations, as the obtained p\_value will be same in both scenarios.

How?

Consider a permutation [a,b,a,c]. This permutation will be repeated two times. Let m and n be the lengths of the two samples we consider in this scenario. We have to note that if [a,b,a,c] is repeated two times, symmetrically [c,a,b,a] is also repeated two times. If T is greater than  $T_{\text{obs}}$  in first case, then T will be less than or equal to  $T_{\text{obs}}$  in the second case. This is true in all the possible cases of m and n i.e  $m > n$ ,  $m < n$ ,  $m = n$ .

So, if we remove all the repeated permutations, if one 1 is removed, one 0 will also be removed. So, the final  $\Pr(T > T_{\text{obs}})$  will not change.

Therefore we can conclude that the P\_Value is irrespective of the decision that we are including/not including the duplicate permutations. So to reduce the calculational complexity, it is advised not to include the duplicate permutations.

## Given Problem

Given  $X = \{3, 2\}$  and  $Y = \{2, 5\}$

Let us use the difference of the means of the distributions as the measure for this experiment.

$$X_{\text{mean}} = 2.5$$

$$Y_{\text{mean}} = 3.5$$

Null Hypothesis  $H_0$ : The two samples follow similar distributions

Alternate Hypothesis  $H_1$ : The two samples do not follow similar distributions.

$$T_{\text{obs}} = |2.5 - 3.5| = 1$$

Let us generate all the permutations of these distributions(excluding the repeated permutations. These repetitions are caused by the element 2, which is repeated twice).

Index	X	Y	X_mean	Y_mean	$T =  X_{\text{mean}} - Y_{\text{mean}} $	$I(T > T_{\text{obs}})$
0	[3, 2]	[2, 5]	2.5	3.5	1	0
1	[3, 2]	[5, 2]	2.5	3.5	1	0
2	[3, 5]	[2, 2]	4	2	2	1
3	[2, 3]	[2, 5]	2.5	3.5	1	0
4	[2, 3]	[5, 2]	2.5	3.5	1	0
5	[2, 2]	[3, 5]	2	4	2	1
6	[2, 2]	[5, 3]	2	4	2	1
7	[2, 5]	[3, 2]	3.5	2.5	1	0
8	[2, 5]	[2, 3]	3.5	2.5	1	0
9	[5, 3]	[2, 2]	4	2	2	1
10	[5, 2]	[3, 2]	3.5	2.5	1	0
11	[5, 2]	[2, 3]	3.5	2.5	1	0
Total						4

$$P_{\text{Value}} = 4/12 = 0.33$$

$$P_{\text{Value}} > 0.05$$

Therefore the Null Hypothesis is Accepted.