

Lecture 59: Two-sample Kolmogorov–Smirnov test

Statistical Inference–II ■ Units IV–V: Nonparametric Inference

Kapil Kumar

Department of Statistics
Chaudhary Charan Singh University, Meerut

M.Sc. Statistics, Semester III
Academic Session 2026–2027

Previous lecture

Lecture 58: One-sample Kolmogorov–Smirnov test.

Today

- Kolmogorov–Smirnov Tests
 - Two-Sample K–S Test
 - Ties and Discrete Distributions

Two-sample K-S test

For independent samples of sizes m and n with e.d.f.s F_m and G_n , the statistic is

$$D_{m,n} = \sup_x |F_m(x) - G_n(x)|.$$

Under $H_0 : F = G$ and continuity, its null distribution is distribution-free: $D_{m,n}$ depends only on the pooled ordering of the labels, and all $\binom{m+n}{m}$ label allocations are equally likely. A common large-sample rejection rule is

$$D_{m,n} > c_\alpha \sqrt{\frac{m+n}{mn}}.$$

The two-sample test is sensitive to differences in location, spread and shape.

Exact critical values for small samples (equal sample sizes; unequal sizes up to 10): Tables in Appendix B of the lecture notes.

Example: two-sample K-S test

Example 59.1

Let $X = (1, 2, 3, 4)$ and $Y = (3.5, 4.5, 5, 6)$. Test $H_0 : F = G$ at the 5% level by the two-sample K-S test.

Solution

Comparing the two e.d.f.s at the pooled ordered values gives

z label	1	2	3	3.5	4	4.5	5	6
	X	X	X	Y	X	Y	Y	Y
$F_4(z)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	$\frac{3}{4}$	1	1	1	1
$G_4(z)$	0	0	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	1
$ F_4(z) - G_4(z) $	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	0

so $D_{4,4} = 3/4 = 0.75$, attained at $z = 3$ and $z = 4$.

Example: two-sample K-S test (cont.)

Solution (cont.): exact p -value

Under H_0 each of the $\binom{8}{4} = 70$ allocations of the labels is equally likely. Let S_t be the number of X 's minus the number of Y 's among the t smallest pooled observations; then $D_{4,4} = \max_t |S_t|/4$, and $D_{4,4} \geq 3/4$ exactly when the path $S_0 = 0, S_1, \dots, S_8 = 0$ reaches $+3$ or -3 .

By the reflection principle (reflect the path after its first visit to 3), the paths with four steps up and four down that reach $+3$ correspond one-to-one to the paths ending at $+6$, i.e. with seven steps up and one down; there are $\binom{8}{1} = 8$ of them. Likewise 8 paths reach -3 , and no path of eight steps reaches both ($3 + 6 = 9 > 8$ steps would be needed). Hence

$$\mathbb{P}(D_{4,4} \geq 0.75) = \frac{16}{70} = 0.2286.$$

Example: two-sample K-S test (cont.)

Solution (cont.)

This agrees with complete enumeration of the 70 allocations: 2 give $D_{4,4} = 1$, 14 give 0.75, 38 give 0.50 and 16 give 0.25.

H_0 is not rejected. Indeed $\mathbb{P}(D_{4,4} = 1) = 2/70 = 0.0286$, so the exact 5% test rejects only for complete separation of the samples.

The asymptotic critical value $1.36\sqrt{8/16} \approx 0.962$ leads to the same decision, but for samples this small the exact calculation should be used.

The classical K–S null distribution assumes continuity.

With discrete data, the statistic remains meaningful but the usual critical values are not exact and are generally conservative.

Exact discrete calculations, permutation methods or simulation under the null should be used.

Practice (Lecture Notes, Chapter 3)

- Exercises (Section 3.15): 3.26.

Next lecture

Lecture 60: Comparison and selection of procedures; summary and revision.