
Kolmogorov-Smirnov test

Kolmogorov test¹ measures the maximum absolute deviation between the empirical distribution function of a sample and the empirical distribution function of a theoretical function:

$$D_n(x) = \sup_x |F(x) - F_n(x)|$$

Smirnov test (better known as Kolmogorov-Smirnov test) is based on Kolmogorov statistics and concerns the comparison of two samples². The aim of this test is to compare two distributions, f and g with respectively n and m observations, in order to test whether the two independent samples involved have been drawn from the same population.

Deriving from Kolmogorov test statistics, the test statistics $D_{n,m}$ computation involves the two empirical distribution functions F and G . If the two samples have been drawn from the same population distribution, the cumulative distributions of both samples may be expected to be close to each other; if the two sample cumulative distributions differ, this suggests that the samples come from different populations.

$$D_{n,m}(x) = \sup_x |F_n(x) - G_m(x)|$$

The asymptotic probability is the following:

$$\lim_{n \rightarrow \infty} P(\sqrt{n}D_n > z) = 2 \sum_{k=1}^{\infty} (-1)^{k-1} \cdot e^{-2k^2 z^2}$$

where the quantity λ is defined:

$$\lambda_{obs} = \left[\sqrt{\frac{n \cdot m}{n + m}} + 0.12 + \frac{0.12}{\sqrt{\frac{n \cdot m}{n + m}}} \right] \cdot D_{obs}$$

¹A.Kolmogorov, *On the empirical determination of a distribution function*, Giornale dell'Istituto Italiano degli Attuari, 1933.

²N.V.Smirnov, *On the estimation of the discrepancy between empirical curves of distributions for two independent samples*, Bull. Math. Univ. Moscou, 1939.