

HUDM6026 Homework_07

Chenguang Pan

Mar 17, 2023

Q1:

Do some research on the two-sample Kolmogorov-Smirnov (K-S) test for equality of distributions. Describe the null and alternative hypotheses and discuss how the test statistic is computed

MY SOLUTION:

1.1 Two sample K-S test's concept

The two-sample Kolmogorov-Smirnov(K-S) test is a nonparametric test used to test whether two datasets are from the same population. K-S test works by comparing the cumulative distribution functions (CDF) of the two datasets.

The null hypothesis is that two datasets are drawn from the same population or that they have the same underlying probability distributions. The alternative hypothesis is that they are not from the same population or that they do not have the same underlying probability distributions. That is,

$$H_0 : F_1(x) = F_2(x) \quad \text{v.s.} \quad H_1 : F_1(x) \neq F_2(x)$$

, where $F_1(x)$ and $F_2(x)$ are the distributions of two sets of random variables X_1 and X_2 .

1.2 How to compute the two-sample K-S test

The K-S test compares the largest vertical distance between the two empirical cumulative distribution functions(ECDF). This is also called the Kolmogorov-Smirnov statistic (D). The test statistic is defined as:

$$D = \max|F_1(x) - F_2(x)|$$

, where $F_1(x)$ and $F_2(x)$ are the ECDF of two datasets being compared.

The maximum distance (i.e., the D) between the two CDFs is then compared to a critical value

$$c(\alpha) \cdot \sqrt{\frac{m+n}{m \cdot n}}$$

, where $c(\alpha) = \sqrt{-0.5 \ln(0.5\alpha)}$, depending on the sample size and significance level chosen.

If the test statistic is greater than the critical value, the null hypothesis is rejected, indicating that the two sets of data are not drawn from the same population.