

In Python, the Scipy.stats package has hypothesis testing functions of KS Test and AD Test.

These functions can be called with the inputs as the fitted distribution and the data set. The test function will then return the test statistics along with the p-value. If p-value is large (taken as 0.05) we cannot reject null hypothesis that the data is from the fitted distribution.

Task 1

- Make a copy of the ipynb notebook “Goodness-of-Fit.ipynb” from Moodle to your GDrive
- Download the MTTF.txt file from Moodle.
- Open your copy of the “Goodness-of-Fit.ipynb” and upload the MTTF.txt file in the Files Tab on the left side panel.
- The distribution fitting steps using KS test & AD Test is written as a function, MyFit().
 - Read it and understand the code.
 - This is a template code. DO NOT MODIFY THAT CELL.
- Follow the instructions in the ipython notebook to run the examples. The case of fitting Exponential distribution and Normal distribution is provide in the notebook.
- Identify which “best fit” distribution the MTTF data set using KSTest/ AD Test.

Show your work to TA

Task 2: The arrival time of 500 customers has been collected (in minutes) and provided in Moodle (arrivaltimes.txt). Use the data to find the best fit distribution of the arrival process (Hint: Think IAT)

Task 3: BEST operates a mini Bus from KanjurMarg to Hiranandani Gardens Powai. The number of passengers carried by the bus over 1000 trips is recorded (dataset1.txt). Find the best fit distribution of the data.

Show your work to TA

Task 4: There are several GUI based distribution fitting software such as ExpertFit, EasyFit, Stat:Fit, InputAnalyzer, etc. Unfortunately, many of them may only work in Windows, and may not be free for use. EasyFit has an evaluation version for 30 days. Anyway, search and install/use any ONE of such software, and carry out tasks 1-3 in that. Carefully read and interpret the outputs. Identify drawbacks, if any, on using such software results.