EP 4130/PH 6130 Assignment 5

Deadline **1** March 2024 before **23:59 hrs**
Please show the source code for each of the problems.

1. Download the asteroid dataset from `http://astrostatistics.psu.edu/datasets/asteroid_dens.dat`. Apply the Shapiro-Wilk test to both the asteroid density values and the natural logarithm of the density values. From the p values, which of these is closer to a Gaussian distribution? Verify this by plotting histograms of both density and its logarithm and overlaying the best-fit normal distribution (Look up `stats.norm.fit`) (25 points)

2. Download the Hipparcos star catalog from `http://iith.ac.in/~shantanud/HIP_star.dat`. Detailed explanation of the columns in this dataset can be found in `http://astrostatistics.psu.edu/datasets/HIP_star.html` under "Dataset". Calculate using two-sample t-test whether the color (B-V) of the Hyades stars differs from the non-Hyades ones. The Hyades stars have Right Ascension between 50° and 100°, declinations between 0 and 25°, proper motion in RA between 90 and 130 mas/year, proper motion in DEC between -60 and -10 mas/year. Any other star which does not satisfy any of the above conditions is considered a non-Hyades star. (25 points)

3. The T90 distribution for Beppo-Sax T90 data can be found at `http://www.iith.ac.in/~shantanud/beppoSax.txt`. Apply GMM to $\log_{10}$ of T90 data and find the optimum number of components using AIC and BIC by plotting BIC as a function of number of componts (20 points)
(Hint: Look at the source code for astroML figure 6.6)