Statistics for Astronomers
Homework #4 (Due before 5:00 PM on Tuesday, 2019.04.09)

Prof. Sundar Srinivasan

April 2, 2019

1. **(13 points)**
   The K-band luminosity function of carbon stars in the Large Magellanic Cloud.
   This problem will use K-band photometry from the 2MASS survey for stars in the Large Magellanic Cloud that were classified as carbon-rich by Boyer et al. (2011). The data is available here in the form of a two-column comma-separated file, with the first column containing the K-band magnitude and the second column containing the uncertainties in these magnitudes. In what follows, the histogram of K-band magnitudes will be referred to as the K-band luminosity function (KLF).

   (a) Plot the KLF. Generate $N_{\rm iter} = 1000$ realisations of this KLF and compute the 95% CI for each bin. Combine these CIs to generate a 95% point-wise confidence band for the KLF. Overlay this confidence band onto the original KLF.

   (b) The location of the peak of the carbon-star LF can place strong constraints on the efficiency of the third dredge-up process (see, *e.g.*, Marigo et al. 1999). Generate $N_{\rm iter} = 1000$ realisations of the KLF using the magnitude uncertainties. For each realisation, find the magnitude at which the KLF peaks. Use these values to compute a 95% CI for the magnitude of the KLF peak. Plot the computed range onto the figure generated in 1a.

   Warning: make sure that the bin edges/locations don't change during the multiple realisations of the KLF!

2. **(12 points)**
   Comparing globular cluster luminosity functions.
   Download K-band data for globular clusters in the Milky Way and M31. The Milky Way data is corrected for distance (*i.e.*, they are absolute K magnitudes), but the M31 data aren't. Correct the latter assuming a distance modulus of 24.9 mag. Use these two datasets to answer the following questions:

   (a) Perform the 2-sample KS test on the datasets to determine if they are drawn from the same distribution to within 95% significance.

   (b) What are the sample mean and standard deviation for each sample? Use these values to "studentise" the data. Perform the KS test to see if the studentised data in each case are consistent with a standard normal distribution to within a 95% significance. Which dataset is less likely to be drawn from a standard normal?

   (c) For this dataset, plot the histogram of studentised data and compare it to a normal distribution. Which end (bright or faint) deviates more from the normal? Observationally, what could this discrepancy be due to?