Statistics for Astronomers
Homework #5 (Due before 5:00 PM on Thursday, 2020.11.26)

Prof. Sundar Srinivasan

November 19, 2020

Notes: (1) You are welcome to use `Python` functions to evaluate probabilities for various distributions, and `Mathematica/Wolfram Alpha` to compute integrals if necessary. **Just mention your source in each case!** (2) Email me your `Python` scripts and any/all resulting output plots/images.

1. (**5 points**)
   The file simple_bootstrap.txt contains a sample of 10 points drawn from an unknown distribution. Write a Python script to compute 100 bootstrap resamples of this dataset. Plot the bootstrap resampled distribution of the sample median and print out the mean value of its bootstrap distribution.

2. (**12 points**) **Histogram with uncertainties.**
   The file M31_GCs.txt contains 2MASS K_s-band magnitudes for globular clusters associated with the Andromeda Galaxy (M31). In this problem, you will generate a histogram of these magnitudes and use the bootstrap procedure to add standard errors to the count in each histogram bin.

   Write a Python script to generate 100 bootstrap resamples of the original dataset. Compute a histogram for each bootstrap resample, making sure that the bin edges of the histograms remain the same each time (you can control this using the `bins` keyword in `numpy.histogram`, for example). Finally, compute the mean and standard deviation of the counts in each bin of the 100 bootstrapped histograms. Plot your result as a histogram, showing the bootstrapped mean count for each bin, with the standard deviations as error bars (here's an example of such a plot).

3. (**8 points**) **Paired bootstrap for a linear regression problem.**
   It is suspected that two quantities X and Y are linearly correlated. Only measurements of Y are affected by random error. In such a case, the least-squares estimators for the slope β and intercept α are given by

$$\widehat{\beta} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N}(x_i - \overline{x})^2} \qquad \text{and} \qquad \widehat{\alpha} = \overline{y} - \widehat{\beta}\,\overline{x},$$

   where $\overline{x}$ and $\overline{y}$ are the sample means of the observations for X and Y.

   The file data_for_paired_bootstrap.csv contains 75 pairs of observations for X and Y. A paired bootstrap is a procedure in which one resamples pairs of (X, Y) values from the dataset such that each

bootstrap resample contains as many pairs as exist in the original dataset.

Write a Python script to obtain 100 paired bootstrap resamples of the dataset. Compute the least-squares estimates of the slope and intercept for each bootstrap resample, thus obtaining the bootstrap distributions for the slope and intercept. Use these distributions to compute and print out the 68% central confidence intervals for the slope and intercept.