Statistics for Astronomers
Homework #4 (Due before 5:00 PM on Wednesday, 2020.10.28)

Prof. Sundar Srinivasan

November 5, 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. **The distribution of the sample variance of a normally-distributed variable.**
The sum of squares of N independent, normally-distributed variables has a χ^2 distribution with N degrees of freedom, with expectation value N and variance $2N$. Answer the following questions using only this information and the properties of the expectation value and variance.

$$\text{Define } Y = \sum_{i=1}^{N}\left(\frac{X_i-\mu}{\sigma}\right)^2,\ U = \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{N}}\right)^2 \text{ and } W = \sum_{i=1}^{N}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2.$$

   (a) (**3 points**) What are the distribution, expectation value, and variance of Y? What is its interpretation?

   (b) (**3 points**) What are the distribution, expectation value, and variance of U?

   (c) (**4 points**) What are the expectation value and variance of W? What is its interpretation? Can you guess its distribution from its relationship to the above quantities?
   Hint: $\sum_{i=1}^{N}\left(X_i-\mu\right)^2 = \sum_{i=1}^{N}\left(X_i-\overline{X}\right)^2 + N\left(\overline{X}-\mu\right)^2.$

   (d) (**3 points**) Using the solutions for the above, what are the distribution, expectation value, and variance of the sample variance of a normally-distributed variable?

2. (**3 points**) The total mass of 100 passengers on Venture Airways Flight VEN 155 is 8450 kg. Assuming that the passenger masses are distributed according to $\mathscr{N}(\mu,\sigma^2)$, with $\sigma = 15$ kg, what is the 95% CI on the population mean of passenger masses? Note that you have 100 data points for the passenger mass!

3. (**5 points**) Write a code that produces 10 standard normal deviates. Use these 10 data points to produce a 95% CI for the population mean. Repeat this procedure 100 times. Print out the fraction of CIs that contain the true mean of zero. Does the result agree with your expectation (explain why/why not)?

4. Sarah computes metallicities for four previously-unstudied Solar Neighbourhood stars, obtaining [Fe/H] values of 2.878, 1.318, 1.545, 0.888 dex.

   (a) (**3 points**) Assuming that the Solar Neighbourhood metallicities are normally distributed, based on her data, what is the 95% confidence interval for the population mean? Note that you have 4 data points!

   (b) (**1 point**) Sarah's colleague then tells her that Casagrande et al. A&A 530, A138 used data for almost 1500 late-type Solar Neighbourhood stars to constrain the population mean tightly around [Fe/H] ≈ 0 dex (cf. Fig. 5 in the paper). Sarah decides that she will publish a paper about her stars being anomalous if the Casagrande et al. mean is different from her sample mean at less than 5% significance. Should she start working on an article?