Statistics for Astronomers
Homework #2 (Due before 5:00 PM on Monday, 2020.10.05)

Prof. Sundar Srinivasan

October 6, 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) For Question 6, email me your `Python` scripts and any/all resulting output plots/images.

1. Sarah writes a code to print out the sum of 10 independent random deviates drawn from a probability distribution $p_X(x)$ that has mean 3.5 and variance 0.025. She runs the code twice, generating sums S_1 and S_2.

   (a) **2 points** Explain why S_1 and S_2 are (i) independent and (ii) identically distributed random numbers.

   (b) **1 point** What are the mean and variance of the distribution from which the sums are drawn? Why?

2. **(3 points)**
   A 100-seater plane has a passenger load limit of 8450 kg. Assuming that the passenger masses are independent and identically distributed according to $\mathcal{N}(\mu, \sigma^2)$, with $\mu = 80$ kg and $\sigma = 15$ kg, what is the probability that the load limit is exceeded?

3. *(Adapted from Chapter 2 of "Practical Statistics for Astronomers" by J. V. Wall & C. R. Jenkins)*
   Given that about 60% of the sources in the Hubble Guide Star Catalogue are binary stars,

   (a) **(1 point)** what is the probability that a random sample of $N = 10$ stars contains 3 non-binaries (corresponding to a 30% contamination from non-binaries)?

   (b) **(3 points)** what is the smallest value of N for a >99% probability that there are at least 2 non-binaries in a randomly selected sample?

4. *(Adapted from Problem 2 in Example Set 2 of the Astronomical Data Analysis lectures by Martin Hendry of Glasgow University.)*
   The Baryon Oscillation Spectroscopic Survey (BOSS) program identified 87 822 quasars over a survey area of 3 275 deg^2 (Pâris et al. 2012 A&A 548, A66). Assuming that the projected distribution of quasars can be modelled as a Poisson distribution,

   (a) **(1 point)** what is the probability of observing less than 4 quasars in a given square degree of the sky? Justify your choice of probability distribution.

   (b) **(2 points)** what area of sky could one expect to survey before the probability of finding *no* quasars was less than 1%?

5. Assume that stars of the same spectral type have the same radial speed $v_{\rm rad}$, but that their directions are oriented randomly. Thus, the projected radial velocities are $v_{\rm rad}\cos\phi$, where ϕ (the angle between the line-of-sight and the radial velocity vector) is drawn from Uniform$[0,\pi)$.

   (a) **(2 points)** What is the probability distribution of the projected velocities?

   (b) **(1 point)** What is the population mean?

   (c) **(2 points)** What is the population variance?

6. Generate $N = 1000$ random points inside a circle of radius R such that each unit area contains approximately the same number of points.

   (a) **(5 points)** Justify the probability distribution(s) you use to generate the locations of these points.

   (b) **(2 points)** Write a `Python` script that, for a given N and R pair, generates the sample and outputs a scatter plot showing their distribution. (You can use this plot to verify that the distribution of points is uniform)