Statistics for Astronomers
Homework #2 (Due before 5:00 PM on Tuesday, 2019.03.05)

Prof. Sundar Srinivasan

March 11, 2019

Note: for questions 4 and 5, email me your `Python` scripts and any/all resulting output plots/images **before the deadline**.

1. **3 points**
   Given that about 60% of the sources in the Hubble Guide Star Catalogue are binary stars,

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

2. **3 points**
   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) 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) what area of sky could one expect to survey before the probability of finding *no* quasars was less than 1%?

3. **8 points**
   (Adapted from Chapter 2 of "Practical Statistics for Astronomers" by J. V. Wall & C. R. Jenkins)
   Anna tosses a fair coin $N = 100$ times. She gets a point for each head.

   (a) What is the probability distribution for Anna's final score? What is the mean and variance?

   (b) Suppose, in addition to gaining a point for each head, Anna loses a point for each tail. How do the answers to (1) change?

   (c) Anna gains a point on the 100$^{\text{th}}$ (final) turn. When do you expect was the last time Anna lost a point (*e.g.*, "the 45$^{\text{th}}$ turn")?

   (d) Using a random number generator in `Python` to simulate the coin toss, "play" the above game $M = 1000$ times (100 turns each time) and generate the distributions for (a), (b), and (c). Compute the means and variances. Do these values agree with your answers for (a), (b), and (c)?

4. **4 points**
Two random variables, X and Y, have a joint pdf given by

$$p_{XY}(x, y) = e^{-y}, \quad 0 < x < y < \infty, \quad \text{zero elsewhere.}$$

(a) Determine the marginal pdf of x and the marginal pdf of y.

(b) Determine the conditional pdf of x given y and of y given x.

(c) Are x and y independent random variables? Why (not)?

5. **7 points**
Generate $N = 1000$ random points that are uniformly distributed inside a circle of radius $R = 4$.

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

(b) Write a `Python` script that, for a given N and R pair, generates the sample and outputs a scatter plot showing their distribution.