Statistics for Astronomers
Homework #1 (Due before 5:00 PM on Monday, 2020.09.28)

Prof. Sundar Srinivasan

September 21, 2020

1. **(2 points)**
   A multiple-choice quiz contains two questions with 5 possible answers for each.
   (a) For Question 1, I select an answer at random, assuming that it has a 20% chance of being the correct answer. What rule did I use to arrive at this conclusion?
   (b) For Question 2, I notice that one of the answers contains terms that the professor repeatedly used in class. I conclude that this answer has a $> 20\%$ chance of being correct. Which interpretation of probability did I use to justify this estimate?

2. (**4 points**)
   An algorithm is used to automatically classify mid-infrared spectra into oxygen-rich ("O") and carbon-rich ("C") chemistries. When tested on a training sample containing 802 O-rich and 82 C-rich spectra, it is found that 20 O-rich objects are misclassified as "C" and 3 C-rich objects are misclassified as "O".
   What is the probability that an object randomly selected from the sample is either classified as "O" **or** misclassified?

3. (**3 points**)
   Table 1, adapted from Boyer et al. (2011 AJ 142 103), shows the numbers of so-called "far-infrared (FIR) objects"[1] in the Small and Large Magellanic Clouds (SMC and LMC, respectively). The FIR objects are extracted from two populations: a fainter population consisting of red giant branch (RGB) stars and a more luminous population made up asymptotic giant branch (AGB) and red supergiant (RSG) stars.
   Given that a randomly selected source from this sample is classified as "FIR(RGB)", what is the probability that it is associated with the LMC?

Population	N_{SMC}	N_{LMC}
FIR(RGB)	303	1262
FIR(RSG+AGB)	57	224

Table 1: Table for Question 3.

[1] Defined in Boyer et al. (2011) as sources with a higher flux density in the *Spitzer* MIPS 24 μm band than in the *Spitzer* IRAC 8 μm band.

4. (**6 points**)
Suppose that HIV is known to infect 0.25% of the population of a country. The ELISA test can be used to check for the presence of HIV antibodies. The test is very accurate: 99.5% of infected subjects test positive, and only 7.2% of healthy subjects test positive.

    (a) Given that a person tests positive for HIV, what is the probability that they are actually infected?

    (b) The ELISA test is repeated on this person, and they test positive again. What is the probability that they are actually infected?

5. (**5 points**)
(Adapted from Chapter 2 of "All of Statistics: A Concise Course in Statistical Inference" by L. Wasserman)
Five coins have probabilities $p_1 = 0, p_2 = \frac{1}{4}, p_3 = \frac{1}{2}, p_4 = \frac{3}{4}$, and $p_5 = 1$ of landing heads if tossed. A coin is selected at random and tossed twice. Let C_i denote the event that coin i is selected, H_1 the event that the first toss results in heads, and H_2 the event that the second toss results in heads.

    (a) Given that the first toss results in a head, compute the probabilities that coin number i ($i = 1, 2, 3, 4, 5$) was selected (*i.e.*, compute the probabilities $P(C_i|H_1)$).

    (b) Given that the first toss results in a head, compute the probability that the second toss also results in a head (*i.e.*, compute the probability $P(H_2|H_1)$).