Data Analysis and Exploration
Homework on Bayes Theorem
Due Friday May 9th

---

1. Bob receives three kinds of emails. Important emails, mundane emails and spam. Ten percent of emails are important, seventy percent are mundane and twenty percent are spam. Important emails are marked as spam only one percent of the time, mundane emails are marked as spam ten percent of the time and actual spam emails are marked as spam eighty percent of the time.

   a Compute the probability of an email being important if it was marked as spam. [4]

   b Compute the probability of an email being spam if it was marked as spam. [4]

   c Compute the probability of an email being important if it was marked as not spam. [4]

   d The probability of an email being mundane if it was not marked as spam. [4]

2. Three fair six sides dice are rolled and their values are denoted by the random variables X_1, X_2 and X_3.We define $Y_1 = X_1 + X_2$ and $Y_2 = X_2 + X_3$. Notice that Y_1 and Y_2 are dependent random variables as both have X_2 as a contributor.

   Compute:

   a $\mathbb{P}(X_1 > 4 \cap X_2 > 3)$ [2]

   b $\mathbb{P}(Y_2 > 6 | X_1 = 1)$ [3]

   c $\mathbb{P}(Y_1 > 7 | X_1 = 5)$ [3]

   d $\mathbb{P}(Y_1 > 8 | Y_2 = 11)$ [6]

3. We have a possibly weighted coin which has been flipped 12 times and produced 8 heads.

   a Set up a one-sided hypothesis test and compute the p-value for the null hypothesis that $p = 0.5$. Do not use the normal approximation. [10]

   b Set up a one-sided hypothesis test and compute the p-value for the null hypothesis that $p = 0.7$. Do not use the normal approximation. [10]

4. Data scientists at the Willy Wonka's factory are allowed as much ice cream as they want!!!

   (a) 175 data scientists like chocolate ice cream and 130 like vanillia. If 260 like at least one of chocolate or vanillia, how many like both chocolate and vanillia [3].

   (b) If 150 students like Strawberry and 110 of those like at least one of chocolate or vanillia how many students like at least one of chocolate, strawberry and vanillia? [3]

   (c) If 25 students like all three flavours how many enjoy exactly two? [4]

   [10]

5. The DIVOC-91 virus has infected the much of the world's rodent population. In fact a quarter of the worlds rodents have been infected. You're worried about your pet mouse Ulric so you buy a test which is ninety-nine percent accurate weather the mouse is infected or not. Compute:

   a The probability that Ulric is infected if the test declares him to be infected. [5]

   b The probability that Ulric is infected if the test declares him to be uninfected. [5]

6. The disease of Greyscale is feared by all in the seven kingdoms. Fortunately there is an imperfect test for the early detection of this disease. If someone has greyscale it'll detect the disease ninety-five percent of the time. If someone does not have the disease it'll falsely detect it only three percent of the time. A soldier from Dorne has just tested positive for the disease! Determine the probability that he is actually infected if:

   (a) The fraction of the population infected is 0.01 (one person in a hundred is infected). [3]

   (b) The fraction of the population infected is 0.1 (one person in ten is infected). [3]

   (c) The fraction of the population infected is t. [4]