CS 4964: Foundations of Data Analysis **Due: Wednesday 9.13 at 2:45pm**

Homework 1: Probability and Bayes' Rule

Instructions: Your answers are due at 2:45, before the beginning of class on the due date. You must turn in a pdf through canvas. I recommend using latex (`http://www.cs.utah.edu/~jeffp/teaching/latex/`) for producing the assignment answers. If the answers are too hard to read you will lose points, entire questions may be given a 0 (e.g. **sloppy pictures with your phone's camera are not ok, but very careful ones are**)

Please make sure your name appears at the top of the page.

You may discuss the concepts with your classmates, but write up the answers entirely on your own. **Be sure to show all the work involved in deriving your answers! If you just give a final answer without explanation, you may not receive credit for that question.**

---

1. **[20 points]** Using the probability table below (one entry is missing, but you should be able to figure it out) for the random variables X and Y, derive the following values

   (a) $\mathbf{Pr}(Y = 1)$
   (b) $\mathbf{Pr}(X = 2 \cap Y = 1)$
   (c) $\mathbf{Pr}(X = 2 \mid Y = 1)$
   (d) $\mathbf{Pr}(X = 3 \cap Y = 2)$

	$X = 1$	$X = 2$	$X = 3$
$Y = 1$	0.25	0.1	0.15
$Y = 2$	0.1	0.2	??

2. **[25 points]** An "adventurous" athlete has the following running routine every morning: He takes a bus to a random stop, then hitches a ride, and then runs all the way home. The bus, described by a random variable B, has four stops where the stops are at a distance of 1, 3, 4, and 7 miles from his house – he chooses each with probability 1/4. Then the random hitchhiking takes him further from his house with a uniform distribution between -1 and 4 miles; that is it is represented as a random variable H with pdf described

$$f(H = x) = \begin{cases} 1/5 & \text{if } x \in [-1, 4] \\ 0 & \text{if } x \notin [-1, 4]. \end{cases}$$

   What is the expected distance he jogs each morning?

3. **[30 points]** Consider a data set D with 10 data points $\{-1, 6, 0, 2, -1, 7, 7, 8, 4, -2\}$. We want to find a model for M from a restricted sample space $\Omega = \{0, 2, 4\}$. Assume the data has ~~Gaussian~~ Laplace noise defined, so from a model M a data point's probability distribution is described $f(x) = \frac{1}{4}\exp(-|M - x|/2)$. Also assume we have a prior assumption on the models so that ~~[old: $\mathbf{Pr}(M = 0) = 0.4$]~~ $\mathbf{Pr}(M = 0) = 0.3$, $\mathbf{Pr}(M = 2) = 0.3$, and $\mathbf{Pr}(M = 4) = 0.4$. Assuming all data points in D are independent, which model is most likely.

4. **[25 points]** Use python to plot the pdf and cdf of the Laplace distribution ($f(x) = \frac{1}{2}\exp(-|x|)$) for values of x in the range $[-3, 3]$. The function `scipy.stats.laplace` may be useful.