CS 3190: Foundations of Data Analysis **Due: Tuesday 9.14 at 11:50pm**

Homework 1: Probability and Bayes' Rule

Instructions: Your answers are due at 11:50pm submitted on canvas. You must turn in a pdf through canvas. I recommend using latex (`http://www.cs.utah.edu/~jeffp/teaching/latex/`, see also `http://overleaf.com`) 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 for the random variables X and Y, derive the following values

   (a) $\mathbf{Pr}(X \neq 1)$

   (b) $\mathbf{Pr}(X = 0 \cap Y = 0)$

   (c) $\mathbf{Pr}(X = 0 \mid Y = 1)$

   (d) Are X and Y independent? and explain why.

	$X = 0$	$X = 1$
$Y = 0$	3/30	7/30
$Y = 1$	6/30	14/30

2. **[25 points]** An "adventurous" track 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 2, 3, 4, and 10 miles from his house – the first three stops have probability 1/5 of occurring. The 10 mile stop has probability 2/5 of occurring. Then the random hitchhiking takes him further from his house a uniformily distributed number of miles on the distances -2 to 8; that is it is represented as a random variable H with pdf described

$$f(H = x) = \begin{cases} 1/10 & \text{if } x \in [-2, 8] \\ 0 & \text{if } x \notin [-2, 8] \end{cases}$$

   Note that a negative distance means that the runner is taken closer to his house. For example, if $H = -1$, then the runner is taken 1 mile *closer* to his home.

   What is the expected distance he jogs each morning?

3. **[30 points]** Consider a data set D with three data points $\{-1, 2, 10\}$. We want to find a model for M from a restricted sample space $\Omega = \{1, 3, 6\}$. Assume the data has Laplacian

noise defined, so from a model M a data point's probability distribution is described $f_M(x) = \frac{1}{2}\exp(-|M - x|)$. Also assume we have a prior knowledge assumption on the model that $\textbf{Pr}(M = 1) = 0.2$, $\textbf{Pr}(M = 3) = 0.5$, and $\textbf{Pr}(M = 6) = 0.3$. Assuming all data points in D are independent, which model is most likely?

4. **[25 points]** The Rayleigh Distribution, indexed by scale parameter σ, has probability density function given by $f(x) = \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}$ for $x \geq 0$ where $\sigma > 0$. Plot the pdf and cdf of a Rayleigh Random Variable with $\sigma = 3.5$ for values of x is range $[0, 11]$. The function `scipy.stats.rayleigh` may be useful.