CS 3190: Foundations of Data Analysis   Due: Wednesday 9.11 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 for the random variables X and Y, derive the following values

    (a) $\mathbf{Pr}(Y = 1)$

    (b) $\mathbf{Pr}(X = 0 \cap Y = 1)$

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

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

	$X = 0$	$X = 1$
$Y = 0$	0.05	0.25
$Y = 1$	0.2	0.5

2. **[25 points]** A student at Data University realizes she needs to file some registration paperwork 5 minutes from the deadline. She will be fined \$1 for each minute that it is late. So if it takes her 7 minutes, she will be 2 minutes late, and be fined \$2. The time is rounded down to the nearest minute late (so 2.6 minutes late has a \$2 fine).

    Her friend tells her that the usual time to finish is a uniform distribution between 4 minutes and 8 minutes. That is the time to finish represents a random variable T with pdf

    $$f(T = x) = \begin{cases} 1/4 & \text{if } x \in [4, 8] \\ 0 & \text{if } x \notin [4, 8]. \end{cases}$$

    (a) What is the expected amount of money she will be fined if she starts immediately?

    (b) What is the expected amount of money she will be fined if she starts in 2 minutes?

3. **[35 points]** A glassblower creates an intricate artistic glass bottle, and attempts to measure its volume in Liters; the 10 measures are: $\{1.82, 1.71, 2.34, 2.21, 2.01, 1.95, 1.76, 1.94, 2.02, 1.89\}$. To sell the bottle, by regulation, she must label its volume up to 0.1 Liters (it could be 1.7L or 1.8L or 1.9L and so on). Her prior estimate is that it is 2.0L, but with a normal distribution

with standard deviation of 0.1; that is, her prior for the volume V being x is described by the pdf

$$f_V(x) = C \cdot \exp(-1 \cdot (2.0 - x)^2/(2 \cdot 0.1^2)),$$

for some unknown constant C (since it is only valid at increments of 0.1). Assuming the 10 empirical estimates of the volume are unbiased, but have normal error with standard deviation of 0.2, what is the most likely model for the volume V?

4. **[20 points]** Use python to plot the pdf and cdf of the Cauchy distribution $(f(x) = \frac{1}{\pi}\frac{1}{x^2+1})$ for values of x in the range $[-3, 3]$. The function `scipy.stats.cauchy` may be useful.