CS 6190: Probabilistic Modeling **Due Tu 1/23**

Homework 0: Getting Started

Instructions: Be sure to electronically submit your answers in either R Markdown (*.Rmd) or Jupyter (*.ipynb) format. Include all of the output of your code, plots, and discussion of the results in your write up. You may work together and discuss the problems with your classmates, but write up your final answers entirely on your own.

Written Part

1. Let X and Y be continuous, real-valued random variables. Prove the following:
   (a) $\mathrm{E}[\mathrm{E}[X \mid Y]] = \mathrm{E}[X]$
   Hint: The inside expectation is over X, the outside is over Y.
   (b) $\mathrm{Var}(X) = \mathrm{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathrm{E}[X \mid Y])$
   Hint: Expand the two terms on the right, and use your answer to part (a).

   NOTE: These are basic identities, whose proofs can easily be found on the web, but try to prove them on your own before you look up the answers! If you do use an online source for help, remember to cite it.

2. Consider random variables X and Y with joint pdf

$$p(x, y) = \begin{cases} x + y & \text{for } 0 \leq x \leq 1, \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

   (a) What is the marginal pdf $p(x)$?
   (b) What is the conditinal pdf $p(y \mid x)$?
   (c) What is the conditional expectation $\mathrm{E}[Y \mid X]$?
   (d) What is the covariance $\mathrm{Cov}(X, Y)$?

3. Let $X \sim \mathrm{Exp}(\lambda)$, i.e., the exponential distribution with pdf $p(x) = \lambda \exp(-\lambda x)$. Let $Y = \sqrt{X}$.
   (a) What is the density function $p(y)$?
   (b) What is the cdf, $F(y) = P(Y \leq y)$? Verify that $F(0) = 0$ and $F(\infty) = 1$.
   (c) What is the quantile function F^{-1}?
   (d) Compute the mean, $\mathrm{E}[Y]$, and variance, $\mathrm{Var}(Y)$. **Hint:** Use integration by parts.

4. Given a realization $y_1, y_2, \ldots, y_n$ from your random variable Y in the previous problem, what is the maximum likelihood estimate for λ?

R Coding Part

4. Recall that if F is a cdf and $U \sim \text{Unif}(0, 1)$, then $Y = F^{-1}(U)$ will be a random variable with cdf F. Write a function that uses this method to generate n random numbers from the distribution of Y in Problem 3 above (n and λ should be parameters to the function).
   Hint: Look at the `runif` function in R for simulating uniform random varibles.

   (a) Generate 10,000 realizations of the random variable Y with $\lambda = 2$.

   (b) Plot a histogram of these numbers, using the `hist` function with option `freq = FALSE`.

   (c) Use the `lines` command to plot the pdf of Y on top.

   (d) Compute the sample mean and variance of your 10,000 realizations. Do they roughly match the values for $\text{E}[Y]$ and $\text{Var}(Y)$ you calculated above?

5. Now generate 20 realizations from Y with $\lambda = 2$. Plot the log likelihood function for this data. Plot a vertical line at the maximum likelihood estimate for lambda (using the equation you got in Problem 3).