Assignment #8
Due: 6:00pm April 9, 2021

Upload at:

Assignments in COS 302 should be done individually. See the course syllabus for the collaboration policy.

Remember to append your Colab PDF as explained in the first homework, with all outputs visible.
When you print to PDF it may be helpful to scale at 95% or so to get everything on the page.

Problem 1 (20pts)

Consider the following cumulative distribution function for a random variable X that takes values in $\mathbb{R}$:

$$F(x) = P(X \leq x) = \frac{1}{1 + e^{-x}}$$

(A) What is the probability density function for this random variable?

(B) Find the inverse distribution (quantile) function $F^{-1}(u)$ that maps from $(0, 1)$ to $\mathbb{R}$.

(C) In a Colab notebook, implement inversion sampling and draw 1000 samples from this distribution. Make a histogram of your results.

Problem 2 (20pts)

Consider two random variables X and Y that are exponential and Poisson, respectively. The exponential random variable X has parameter λ:

$$X \sim \text{Exponential}(\lambda)\,.$$

The notation $\sim$ is like saying "drawn according to the distribution ...". The Poisson distribution, however, has as its parameter the random variable X:

$$Y \mid X = x \sim \text{Poisson}(x)$$

This notation is saying "the random variable Y, conditioned on $X = x$ is drawn according to a Poisson distribution where x is the parameter".

(A) Write the joint probability density function $p(x, y)$ for these random variables.

(B) Compute the marginal distribution for Y. Here are the rough steps you should follow to do this, which are reflective of the kinds of computations one often does in probabilistic machine learning:

1. Write down the definite integral that you solve in order to get $p(y)$ from $p(x, y)$.

2. Pull terms that don't depend on x outside the integral.

3. Inside the integral, collect exponents.

4. See if the pieces inside the integral have a form that you recognize. If they look like an unnormalized probability density with known form, then you can sometimes solve the integral by inspection using that PDF's normalization constant. This kind of thing comes up a lot. (Hint: for this problem you want to have a look at the gamma distribution.)

5. Simplify. In this problem you'll probably want to take advantage of the relationship between the factorial and the gamma function.

Problem 3 (18pts)

Recall that two random variables X and Y are independent if and only if $P(X,Y) = P(X)P(Y)$. Two random variables X and Y are **conditionally** independent given a third event Z if $P(X,Y|Z) = P(X|Z)P(Y|Z)$. The following problem explores the relationship between independence and conditional independence, specifically whether independence implies conditional independence or vice versa.

(A) Imagine that you have two coins: one regular fair coin ($P(\text{heads}) = 0.5$) and one fake two-headed coin ($P(\text{heads}) = 1$). Consider the following experiment: choose a coin at random and toss it twice. Define the following events:

- A is the event that the first coin toss results in a heads.
- B is the event that the second coin toss results in a heads.
- C is the event that the regular fair coin has been selected.

Are the events A and B independent? Give a qualitative answer, i.e., either yes because... or no because... Are they conditionally independent given C? Show your work, i.e., explicitly show the definition holds as given in the problem statement.

(B) Roll a single six-sided die and consider the following events:

- X is that you roll 1 or 2, i.e., $X = \{1,2\}$.
- Y is that you roll an even number, i.e., $Y = \{2,4,6\}$.
- Z is that you roll 1 or 4, i.e., $Z = \{1,4\}$.

Are the events X and Y independent? Are they conditionally independent given Z? For both of these questions show the definitions hold as in the problem statement.

Problem 4 (40pts)

In 1777, Georges-Louis Leclerc, Comte de Buffon introduced a simple Monte Carlo method for estimation, now called the Buffon's needle problem. It's a fun way to estimate the constant π via physical simulation. In this problem, we'll look at the following variant: you have a piece of paper on a tabletop and drawn on it is a square with sides of length w. Centered and entirely contained in the square is a circle with radius r. (So $2r \leq w$.) You drop n points (say, grains of sand) onto the piece of paper in a spatially uniform way. You count the grains and find that the ratio of grains inside the circle (n_c) to the total number grains inside the square is α:

$$\alpha = \frac{\text{\# of grains inside circle}}{\text{\# of grains inside square}} = \frac{n_c}{n} \; .$$

(A) Write an estimate for π in terms of w, r, n and n_c.

(B) What kind of probability distribution will n_c follow? What is the standard deviation of n_c as a function of n and α?

(C) Assume that the radius of the circle is 1. What box width w minimizes the standard deviation of your estimate of π when you form this Monte Carlo estimate? (Hint: use the standard deviation from your answer in in (B), and reason about the relationship between w and α. Do you want w to be as small as possible, as big as possible, or somewhere in between?

(D) Perform a real-life version of this experiment and estimate π with physical simulation. You could do it with paper and sand or something, as described above. Or you can be more creative! Some ideas:

- Grate a known amount of cheese onto a square pan that has a pizza in the middle. Weigh the cheese that doesn't fall onto the pizza.
- Put a circular cup into a box and drop coins or peas or something into the box and count the fraction that fall into the cup.
- Make a batch of cookies that are all perfect circles. Spread sprinkles over the entire baking pan and count (or weigh) how many don't make it onto cookies. You'll need to work out the compensation for having more than one cookie.
- Put a round bucket inside a square bucket and stick it out in the rain. Measure the difference in water collected.

You'll need to measure the dimensions of the rectangular region and the radius of the circle. Draw a couple of dozen samples and report your estimate. Explain what you did and include a photograph of your setup! (Use a command like `\includegraphics[width=3in]{myimage.jpg}` to include an image file within LaTeX.)

Problem 5 (2pts)

Approximately how many hours did this assignment take you to complete?

My notebook URL: `https://colab.research.google.com/XXXXXXXXXXXXXXXXXXXXXXXXX`

Changelog

- 29 March 2021 - S21 initial version