CSC380 **Homework 2** **Fall 2024**

This homework is due on Friday, September 27^{th} at 11.59pm. Late submissions are not accepted.

Submission Guidelines

- You must make two submissions:
  - Your complete homework as a SINGLE PDF file by the stated deadline to the gradescope. **Include your code and output of the code as texts in the PDF.**
  - Your codes to a separate submission: a single notebook file including codes for questions 8 through 14.
- For your PDF submissions:
  - Select the page number for the answer to each question in the Gradescope.
  - You may submit typed or handwritten/scanned answers. If you decide to submit handwritten answers then please ensure that it is easily readable.
  - You can easily scan and upload your answers as a PDF using the Gradescope mobile app.

Questions below are related to the following table. Let $A, B, C \in \{0, 1\}$ be three binary random variables with the following joint probability distribution:

a	b	c	$f(a, b, c)$
0	0	0	0.01
0	0	1	0.07
0	1	0	0.02
0	1	1	0.10
1	0	0	0.02
1	0	1	0.30
1	1	0	0.04
1	1	1	0.44

1. **[6pts]** By direct calculation, compute the marginal $P(A, B)$ (recall that $P(A, B)$ is represented by 4 numbers: $P(A = 0, B = 0)$, $P(A = 0, B = 1)$, $P(A = 1, B = 0)$, $P(A = 1, B = 1)$).
2. **[6pts]** By direct calculation compute the marginals $P(A)$ and $P(B)$.
3. **[6pts]** Are the random variables A and B independent? Why or why not?
4. **[6pts]** Compute the conditional distribution $P(A, B \mid C)$. Note that this includes computing $P(A, B \mid C = 0)$ as well as $P(A, B \mid C = 1)$, each of which is represented by 4 numbers (in total 8 numbers).

5. **[10pts]** Let

$$f(x,y) = \begin{cases} c(x+y^2) & for\ \ 0 \le x \le 1\ \ and\ \ 0 \le y \le 1 \\ 0 & otherwise \end{cases}$$

   Find $P(X \le 1/2 \mid Y = 1/2)$.

6. **[10pts]** Assume there are 100 boys and 150 girls in a classroom. We are going to select 70 students at random from the classroom without replacement. Let X denote the number of boys that are selected and let Y denote the number of girls that are selected. Find the expectation of $X - Y$.

7. **[6pts]** Assume X, Y are two random variables that have a negative correlation. Determine the relationship between $Var(X+Y)$ and $Var(X-Y)$, i.e. find out whether the former is larger or smaller than the latter.

Questions below are related to random variable $X \sim N(3, 16)$. You need to use `scipy.stats`. Paste your relevant code for each question separately.

8. **[6pts]** Find $P(X > -2)$
9. **[6pts]** Find x such that $P(X > x) = .05$
10. **[6pts]** Find $P(0 \leq X < 4)$

In continuous probability, we often need to solve messy integrals. For example, in this class we might need to use integrals to evaluate the probability of an event under a cumulative distribution function (CDF). Rather than solve this by hand, we can approximate it using discrete intervals. This problem will explore discrete approximation of integrals using a Gaussian model. Recall that the probability density function of a Gaussian random variable $X \sim N(\mu, \sigma^2)$ is,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

In the questions below, we will use Python to form a discrete approximation of this continuous distribution, and evaluate associated probabilities.

11. [**10pts**] Form a discrete approximation of the Normal PDF with mean $\mu = 70$ and standard deviation $\sigma = 2$. To do this, create an array x of evenly spaced values in the range $[68, 76]$ at increments of 2 excluding 76 (this array will include 68 and 74). The function `numpy.arange` might be helpful. Create an array p containing values of the PDF at each location x. Plot the result as a bar chart (use `matplotlib.pyplot.bar`). In the same figure, overlay a PDF curve (use `matplotlib.pyplot.plot`) at more finely spaced intervals (e.g. 0.01). Paste your code.

12. [**10pts**] The bar chart above is a discrete approximation of the continuous PDF. We will use it to approximate $P(68 < X \leq 76)$. Recall that $X \sim N(\mu, \sigma^2)$, so

$$P(68 < X \leq 76) = \int_{68}^{76} f(x)\, dx.$$

We will approximate this integral using a Riemann sum. Let N be the number of grid points in your array x. The spacing between grid points is Δx and let the i^{th} point of array p be p_i. The Riemann sum approximation is,

$$P(68 < X \leq 76) \approx \sum_{i=1}^{N} p_i \, \Delta x$$

Find the value of the approximation to $P(68 < X \leq 76)$. Paste your code.

13. [**6pts**] Now, reduce the spacing $\Delta x = 0.01$ and recompute the discrete approximation of $P(68 < X \leq 76)$. Paste your code and argue: How do the two approximations compare? What is the practical downside of smaller spacing?

14. [**6pts**] Repeat the steps above to show the distribution over the range $[20, 120]$ and compute $P(20 \leq X < 120)$. What is the value? This interval should contain almost all of the probability in this distribution, i.e. the event is almost certain. Paste your code.