STA 141A - Spring 2025 - Homework 3

Instructor: Dr. Akira Horiguchi

Student name: ABCDE FGHIJ; Student ID: 123456789

Due date: Apr 30, 2025 at 9 PM (PT)

The assignment must be done in an R Markdown or Quarto document, or in LaTeX. The assignment must be submitted by the due date above by uploading:

1. a .pdf file in GRADESCOPE (if you can knit/compile your .rmd to a .html file only, please save the created .html file as a .pdf file (by opening the .html file -> print -> save to .pdf)).

Email submissions will not be accepted.

Each answer has to be based on `R` code that shows how the result was obtained. The code has to answer the question or solve the task. For example, if you are asked to find the largest entry of a vector, the code has to return the largest element of the vector. If the code just prints all values of the vector, and you determine the largest element by hand, this will not be accepted as an answer. No points will be given for answers that are not based on `R`. This homework already contains chunks for your solution (you can also create additional chunks for each solution if needed, but it must be clear to which tasks your chunks belong).

There are many possible ways to write `R` code that is needed to answer the questions or do the tasks, but for some of the questions or tasks you might have to use something that has not been discussed during the lectures or the discussion sessions. You will have to come up with a solution on your own. Try to understand what you need to do to complete the task or to answer the question, feel free to search the Internet for possible solutions, and discuss possible solutions with other students. It is perfectly fine to ask what kind of an approach or a function other students use. However, you are not allowed to share your code or your answers with other students. Everyone has to write the code, do the tasks and answer the questions on their own.

During the discussion sessions, you may be asked to present and share your solutions.

Figure 1: If you cannot type math in TeX, you can replace this image with a screenshot of your handwritten solution.

1. PMF, PDF, CDF

Let X be a discrete random variable with pmf $f_X(x) = (1/2)^x$ for $x = 1, 2, 3, 4, 5, \ldots$.

a) Draw or plot the pmf of the random variable X.

b) Draw or plot the cdf of the random variable X.

c) Compute the expected value $E(X)$. You can use R for this.

d) Let C be the event that $X \in \{1, 3\}$, and let D be the event that $X \geq 4$. Compute $P(C) + P(D)$.

- *Hint*: the solution can be written in just one line of math.

e) Let A be the event that X is even, and let B be the event that X is odd. Are A and B independent? Are A and B disjoint? What is $P(A) + P(B)$? Justify your answer.

2. Expected value and variance

Let X be a random variable with pdf $f_X(x) = |x|, x \in (-1, 1)$. (Here $|x|$ denotes the absolute value of x.)

a) Calculate the expected value $E(X)$ by hand.

b) Calculate the expected value $E(3X + 5)$ by hand. *Hint*: use your answer from part a).

c) Calculate the variance $Var(X)$ by hand.

d) Calculate the variance $Var(3X + 5)$ by hand. *Hint*: use your answer from part c).

3. Monte Carlo integration

Suppose we want to compute the expected value

$$\mathbb{E}g(X) = \int g(x) f_X(x)\, dx$$

for a continuous random variable X with pdf $f_X(x)$, where $g(x)$ is a function of x. The exact value of this expected value can be difficult to compute if the function g or the pdf f_X is complicated. However, by noting that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} g(x_i) = \int g(x) f_X(x)\, dx$$

for samples $x_1, x_2, \ldots, x_N$ drawn according to pdf f_X, we can approximate the expected value $\mathbb{E}g(X)$.

For this problem, suppose the random variable X follows a standard uniform distribution, and $g(x) = x^2$.

a) Compute the expected value $\mathbb{E}g(X)$ by hand.

b) Estimate the expected value $\mathbb{E}g(X)$ using R and this Monte Carlo method described above.

- *Hint*: Step 1 is to draw N samples $x_1, x_2, \ldots, x_N$ according to the distribution of X. Step 2 is to compute the sample mean of $g(x_1), g(x_2), \ldots, g(x_N)$.

c) Repeat part b) for four values of N: 10^2, 10^5, 10^8, and 10^{15}. As N increases, what happens to the estimate? What happens to the computation time?

4. Variance and covariance

Let Z_1, Z_2, Z_3, Z_4 be random variables each having respective pdfs f_1, f_2, f_3, f_4. Suppose each pdf has the same variance σ^2. For parts a) and b), assume Z_1, Z_2, Z_3, Z_4 are mutually independent.

a) Compute the variance of $Z_1 + Z_2 + Z_3 + Z_4$.

b) Compute the variance of $(Z_1 + Z_2 + Z_3 + Z_4)/4$. How does it compare to the variance of just one of these random variables Z_i?

For parts c) and d), assume the covariance between any pair of these random variables Z_1, Z_2, Z_3, Z_4 is $\sigma^2/2$. That is, assume $Cov(Z_i, Z_j) = \sigma^2/2$ whenever $i \neq j$.

c) Compute the variance of $Z_1 + Z_2 + Z_3 + Z_4$.

d) Compute the variance of $(Z_1 + Z_2 + Z_3 + Z_4)/4$. How does it compare to the variance of just one of these random variables Z_i?

For parts e) and f), assume the covariance between any pair of these random variables Z_1, Z_2, Z_3, Z_4 is $-\sigma^2/2$. That is, assume $Cov(Z_i, Z_j) = -\sigma^2/2$ whenever $i \neq j$.

c) Compute the variance of $Z_1 + Z_2 + Z_3 + Z_4$.

d) Compute the variance of $(Z_1 + Z_2 + Z_3 + Z_4)/4$. How does it compare to the variance of just one of these random variables Z_i?