Mini Project (first attempt students/deferrals only)

Solutions should be submitted to the assignment submission area in Canvas by **4pm** on **Friday 16th August 2024**.

Your work should be typed up and any necessary graphs or code included in your typed document. Marks will be awarded for style and presentation (e.g. for including suitable labels on plots; any such marks are included in the submarks specificed for each question, as appropriate).

There are **four questions** to complete with a total of **50 marks**, covering all aspects of the module.

Question 1 [Total 25 marks]

In this question, you are required to use Monte Carlo integration to approximate

$$\mathcal{I} = \int_0^1 abx^{a-1}(1-x^a)^{b-1}\,dx,$$

for $0 \leq x \leq 1$ and where $a, b \in \mathbb{Z}^+$.

(a) Each student will need to generate their own unique values for a and b. To generate *your* values for a and b, type the following code into R:

```
set.seed(STUDENT_ID)
a = sample(2:10, 1); b = sample(3:12, 1)
```

where STUDENT_ID is your student number (with all non-numeric parts excluded). In your solutions, write down your values for a and b and the resulting integrand $f(x)$. For example, if your generated values for a and b were 7 and 3 respectively, you would write down:

$$a = 7; \quad b = 3; \quad f(x) = 21x^6(1-x^7)^2.$$

(b) Use R to produce a plot of $f(x)$, experimenting with xlim and ylim to make sure your plot is nicely displayed. Also use the abline command to indicate a suitable *simulation grid* over which you will perform your Monte Carlo integration. Include this plot in your solutions, with appropriately labelled axes and title. **[5 marks]**

(c) Using a `for` loop over $N = 10^6$ replications, estimate your integral using Monte Carlo Integration. Include your R code in your solutions, which should be nicely formatted/indented. Where necessary, include comments in your code so it is clear what you are trying to do. In your solutions, make it obvious what your estimated integral is. **[17 marks]**

(d) Use a built-in command available in R to verify your answer to part (c). Include the R code in your solutions document. *Hint: the function over which you are integrating is actually the density function for a standard probability distribution. Can you spot which distribution this is?* **[3 marks]**

Question 2 [Total 25 marks]

Suppose that X is a random variable with probability density function (PDF)

$$f_X(x) = \beta x^{-(\beta+1)}, \text{ for } \beta > 0 \text{ and } x > 1.$$

(a) Write a function in R to generate simulated values of X using the inverse CDF method. In your solutions, include all necessary calculations and your R function. **[10 marks]**

(b) Let d be the last digit of your student number. Let $\beta = 2 + (d/2.5)$. Write down d and β, and plot a density histogram of a sample from X of size 10,000 using your value of β. Include this histogram in your solutions, along with a line of code that shows your function from part (a) being implemented. Again, make sure axes are appropriately labelled. Superimpose a curve on your histogram showing the PDF. **[9 marks]**

(c) There are many real-life applications of the distribution you have simulated from, including modelling the size of human settlements (cities, towns, villages). What is the name of this distribution? **[2 marks]**

(d) Use your simulated values of X to verify the expectation and variance of the distribution you identified in part (c). **[4 marks]**