

Mini Project (Problem Solving Exercises 2)

Solutions should be submitted to the assignment submission area in Canvas by **4pm on Friday 8th December 2023**. For information about the late work policy, see Canvas or the preliminary module information given in the module notes.

Your work should be typed up and any necessary graphs or code included in your typed document. **[10 marks]** will be awarded for style and presentation, to give a total maximum of **[100 marks]**.

Your typed report should clearly state your name at the top of each page, along with your student number.

Question 1 [Total 25 marks]

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

$$\mathcal{I} = \int_0^{\frac{a-1}{b}} \frac{b^a}{(a-1)!} x^{a-1} e^{-bx} dx,$$

for $x > 0$ 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:

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

where LOGIN_ID is your computer ID without the 'a' or 'b' at the start. 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:

$$a = 7; \quad b = 3; \quad f(x) = \frac{243}{80} x^6 e^{-3x}.$$

- (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 40 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. Include all necessary calculations in your solutions; also include the R code for the function that you write. **[10 marks]**

- (b) Showers of meteorites originate from the fragmentation of large meteoroids; the distribution given in this question is often used to describe the observed diameter of these fragments.

You have each been provided with some data on the diameter (in centimetres) of meteorites from a recent meteor shower (see link on Canvas). The *maximum likelihood estimate* for β can be shown to be

$$\hat{\beta} = \frac{1}{\bar{x}_{\log}},$$

where $\bar{x}_{\log} = n^{-1} \sum_{i=1}^n \log x_i$ and where $x_1, x_2, \dots, x_n$ is your sample data. Find your maximum likelihood estimate for β , and write this down in your solutions. **[4 marks]**

- (c) Using your maximum likelihood estimate for β , produce a density histogram of a sample from X of size 10,000 alongside a boxplot of this sample (remove the outliers in the boxplot if you think it is necessary). In your solutions, include the histogram and the boxplot side-by-side, in a 1×2 panel of plots. **[12 marks]**
- (d) Do some research to find the name of the distribution you have simulated from. Other than modelling the size of meteorites, are there any other typical applications of this distribution? Name at least two such applications. **[4 marks]**
- (e) Use your simulated values of X to verify the expectation and variance of the distribution you identified in part (d). **[6 marks]**
- (f) A meteorite is classed as small if it is less than 3cm in diameter, and large if its diameter exceeds 50 centimetres. Use simulation to find the probability that a randomly selected meteorite is either small or large. Include the R code in your answer to this question. **[4 marks]**

[Turn over for question 3]

Question 3 [Total 25 marks]

If the random variable X follows a *Beta distribution*, that is, $X \sim \text{Be}(a, b)$, then it has PDF

$$f_X(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a, b)}, \quad 0 < x < 1, \quad a, b \in \mathbb{Z}^+,$$

and where $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ and $\Gamma(a) = (a-1)!$.

Suppose a is the value you generated in question 1, and $b = 1$. Suppose also that $Y = -\log X$.

(a) Using R:

- (i) Produce a random sample of size 10,000 from X using a single line of code. Include this code in your solutions. **[2 marks]**
- (ii) Write an R function that produces a random sample from Y . Include this function in your solutions. **[4 marks]**
- (iii) Use your R function in part (ii) to produce a random sample of size 10,000 from Y . In your solutions, include histograms of your random samples from X and Y side-by-side, in a 1×2 panel of plots. **[4 marks]**

(b) By hand:

- (i) Derive the PDF for Y , and include full workings in your solutions. **[6 marks]**
 - (ii) Find the exact value for $P(Y > E[Y])$. **[5 marks]**
- (c) Use your simulations in part (a)(iii) to find an approximation to $P(Y > E[Y])$, and compare to your answer to part (b)(ii). Include any relevant R code/output in your solutions. **[4 marks]**