Mini Project

This mini project is worth 25% of the MAS2906 module marks. Solutions should be submitted to the assignment submission area in Canvas by **4pm** on **Friday 9th December**. 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. Marks will be awarded for style and presentation.

You may work individually or as a pair. Your typed report should clearly state your name/both names at the top of each page, along with your student number(s).

Question 1 [Total 25 marks]

In this question, you are required to use Monte Carlo Integration to approximate $\int_0^1 f(x)dx$, where

$$f(x) = abx^{a-1}(1-x^a)^{b-1}$$

for $0 \le x \le 1$ and where $a > 0$ and $b > 0$.

(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(3:14, 1); b = sample(3:14, 1)
```

where STUDENT_ID is your student ID (excluding any letters that might appear at the start). If you are working as a pair, just use the STUDENT_ID for one student in the pair. In your solutions, write down your values for a and b and the resulting integrand. For example, if your generated values for a and b were 10 and 12 respectively, you would write down:

$$a = 10; \quad b = 12; \quad f(x) = 120x^9(1-x^{10})^{11}.$$ **[1 mark]**

(b) Use R to produce a plot of $f(x)$. Use the `abline` command to indicate a suitable *simulation grid* on your plot. Include this plot in your solutions, with appropriately labelled axes and title. **[7 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]**

Turn over for questions 2 and 3

Question 2 [Total 25 marks]

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

$$f_X(x) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, \quad \text{for } x > 0,$$

where σ is a fixed constant.

(a) Find the cumulative distribution function (CDF) for X and write a function 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. **[12 marks]**

(b) Let d be the last digit of your student ID. Let $\sigma = 1 + (d/2.5)$. Write down your value of d, and plot a density histogram of a sample of X of size 10,000 using your value of σ. Include this histogram in your solutions; again, make sure axes are appropriately labelled. If you are working as a pair, just use the ID for one student in the pair. **[5 marks]**

(c) Check that your function has worked by superimposing a curve on your histogram showing the PDF. **[2 marks]**

(d) The distribution you have simulated from is often used in the physical sciences to model wind speeds and wave heights; it is also used to model noise variation in magnetic resonance imaging (MRI). Can you find the name of this distribution? **[2 marks]**

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

Question 3 [Total 25 marks]

(a) Write an R function to sample from the log-normal distribution $LN(\mu, \sigma^2)$ (see section 5.1 of the notes and example 5.3). Make sure your function takes three input arguments: the number of realisations to generate (`N`), the mean (`mu`) and the standard deviation (`sigma`). Include your R code in your solutions. Note: You must *not* use any of R's built-in functions associated with the log-normal distribution! **[5 marks]**

(b) For $\sigma = 2$ generate samples from the log-normal distribution for a number of different values for μ in the range $\mu \in (0, 5)$. For each value of μ compute the log of the sample mean. Plot a graph of the log of the sample mean versus μ, and include this plot in your solutions. **[8 marks]**

(c) Repeat part (b), plotting the log of the sample median versus μ. **[8 marks]**

(d) What can you conclude about the mean and median of the $LN(\mu, \sigma^2)$ distribution? **[4 marks]**

Turn over for question 4

Question 4 [Total 10 marks]

Suppose a random variable X follows a gamma distribution, that is, $X \sim Ga(\alpha, \lambda)$; then the probability density function (PDF) is

$$f_X(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad \alpha, \lambda > 0,$$

and where $\Gamma(k) = (k-1)!$ for positive integers k.

(a) Let _your_ value for α be the value for a you generated in question 1 part (a), and let _your_ value for λ be the value for b you generated in question 1 part (a). In your solutions write down your PDF for X. **[1 mark]**

(b) Find the PDF for $Y = X/\sigma$, where σ is the value you generated in question 2 part (b). *Note: If your value for $\sigma = 1$, then use the transformation $Y = X/2$.* **[6 marks]**

(c) Use R to find $P(Y < E[Y])$. **[3 marks]**