STAT7203: Applied Probability and Statistics
Week 7 Exercises

1. A continuous random variable X has probability density function

$$f_X(x) = \begin{cases} cx^2, & x \in (0,1) \\ 0, & \text{else,} \end{cases}$$

for some c.

(a) Determine the value of c.

(b) What is the cumulative distribution function of X ?

(c) What is the quantile function of X?

(d) How would you simulate from this distribution in R? Simulate 250 random variables from this distribution and compare the histogram of the simulated random variables with the known probability density function. Some useful R commands for this problem:

- `runif` simulates random variables from the continuous uniform distribution. Typing `runif(n)` returns and a vector of length `n` of random variables from the continuous uniform distribution on [0,1].
- `hist` takes as input a vector `X` and plots a histogram of the values in `X`. The `hist` function has an argument `freq` which we can set to `FALSE` so the bars of the histogram are scaled corrected to compare with the known probability density function.
- `lines(X,Y)` produces a line plot of vector `Y` versus vector `X` on the current plot.

2. In the Jelinski and Moranda model of software reliability, a program contains N bugs. Each of these bugs cause the program to fail independently. Once a bug is found, it is immediately fixed and no longer causes any issues for the program. The time T_i for bug i to cause the program to fail is assumed to have an exponential distribution with rate parameter λ.

For parts (a) - (d) of this question we shall assumed that $N = 2$.

(a) Find the cdf of the time for the program to fail for the first time, that is the cdf of $\min\{T_1, T_2\}$.

(b) Find the cdf of the time for both bugs to have occured, that is the cdf of $\max\{T_1, T_2\}$.

(c) What is the expected time for both bugs to have occured?

(d) What is the variance of the time for both bugs to have occured?
Can you generalise this to an arbitrary number of bugs in the program?
Just something to think about ...

(e) Simulate this process 10 000 times with $N = 2$ and $\lambda = 1$ to get an estimate for the mean and variance of the time for both bugs to be found. Repeat this for a program has $N = 10$ bugs and $\lambda = 1$.
Some useful R commands for this problem:

- `rexp` is the built in function to simulate exponential random variables.
- `matrix` takes an vector input and forms a matrix. The number of rows and columns can be specified with the `nrow` and `ncol` arguments. Note that the length of the input vector should be equal to `nrow` * `ncol`.
- `apply` Returns a vector of values obtained by applying a function to margins of a matrix. For example, if we have a matrix `A` typing `apply(A,1,mean)` returns the row means of `A`.