Practical 1: The Birthday Problem

This work should be submitted electronically via NESS before ***Friday 5th October 2018 at 4pm****.*

To answer the questions you may need to refer to the R guide given in Practical 0 and the R commands given in the notes.

Marks for each question are indicated. Your submission should contain a single file which contains your function code for each part, along with any necessary plots. ***Marks will be awarded for well presented code and plots, see Practical 0 for details****.*

In 1939, a mathematician called Richard von Mises proposed the problem that has become known as the *Birthday Problem*: how many people must be in a room before the probability that at least one pair shares a birthday becomes at least 0.5? Ignoring leap years and assuming that birthdays are equally likely to fall on all 365 days in the year, we will investigate this problem by simulation.

Part 1:

Suppose the number of people in the room is 10. Write some R code which simulates a random sample of 10 birthdays (i.e. 10 numbers between 1 and 365) and then checks whether or not there is one or more shared birthday in the sample. You may find the following R commands useful:

`sample(x, n, replace = TRUE)`: generates a sample of size n from the integers in $(1, ..., x)$ with replacement.

`unique(x)`: returns a vector like x but in which any duplicate elements have been removed.

[**5 marks**]

Part 2:

We can estimate the probability of there being at least one shared birthday in a room of 10 people by simulating a large number of samples of 10 birthdays and computing the proportion with at least one shared birthday.
Write an R function to carry out this simulation procedure. It should take as an argument the number N of random samples of size 10 you want to generate (i.e. the number of repetitions of the experiment) and return an estimate of the probability.
You should test your function by varying the number of repetitions N. The exact probability is 0.117 (to three decimal places). You should see the proportion returned by your function converging towards this value as N increases.

Submit a plot of proportion vs N, for increasing values of N (e.g. 10, 50, 100, 1000, 10000) with a horizontal line showing the exact probability.

[**8 marks**]

Part 3:

Generalise the function from above so that it can estimate the probability for rooms containing n people for any n. Your function should now take two arguments: the number n of people in the room and the number N of repetitions of the experiment. It should return an estimate of the probability of one or more shared birthday.

If the number of people in the room is 366 or higher, we know that the probability of at least one shared birthday must be 1. For rooms with $n = 1, ..., 366$ people use your function to estimate the probability and produce a line plot of n on the x-axis against the estimated probability on the y-axis. Using your graph or R output, how many people are required before the probability becomes at least 0.5? Note that to avoid excessive computational time, you may need to limit N to around $N = 10,000$.

In your report for this question, just provide a listing of the R function and the plot you produced along with your answer to the Birthday Problem. If you are interested in how to compute the probabilities analytically, the Wikipedia page:

`http://en.wikipedia.org/wiki/Birthday_problem`

provides a detailed overview.

[**12 marks**]