STAT7203: Applied Probability and Statistics
Weel 5 Exercises

1. In last week's exercises we saw Benford's law which is a probability mass function on the integers $\{1, 2, \ldots, 9\}$ such that

$$f_D(d) = \log_{10}\left(\frac{d+1}{d}\right), \quad d \in \{1, 2, \ldots, 9\}.$$

We would like to be able to simulate the random variable D from this distribution.

Suppose X has a uniform distribution on the integers $\{1, 2, \ldots, 9\}$ and, conditional on $X = x$, the random variable Y has a Bernoulli$(f_D(x)/\log_{10}(2))$ distribution.

(a) Verify that $f_D(x)/\log_{10}(2) \leqslant 1$ for all $x \in \{1, 2, \ldots, 9\}$.

(b) What is the joint probability mass function of (X, Y)?

(c) Determine $\mathbb{P}(Y = 1)$.

(d) Determine the conditional probability mass function of X given $Y = 1$.

(e) This suggests we can simulate a random variable with probability mass function f_D using the following algorithm

```
Y = 0
While (Y = 0) {
  Simulate X from a uniform distribution on {1,2,...,9}
  Simulate Y from a Bernoulli distribution with success
     probability f(X)/log10(2)
}
Return X
```

In each loop a new pair of random variables (X, Y) is simulated, independent of all previously simulated random variables. Implement this algorithm in `R`. You will need to use a `while` loop. In `R`, the general form of the `while` loop is

```
while (cond) {
  expressions
}
```

where `cond` is a length one logical vector.

(f) What is the distribution of the number of pairs of random variables (X, Y) that need to be simulated in order to simulate a single random variable from Benford's law?

2. Let $(a_1, \ldots, a_n)$ be a random permutation of the integers $\{1, 2, \ldots, n\}$ with all permuations equally likely. An inversion in a permutation is an ordered pair (i, j) such that $i < j$ and $a_i > a_j$. For example, $(1, 3, 2, 4)$ has one inversion $[a_2 > a_3]$ while $(1, 4, 2, 3)$ two inversions $[a_2 > a_3$ and $a_2 > a_4]$.

   For $i < j$, let X_{ij} be the random variable such that

$$X_{ij} = \begin{cases} 1, & a_i > a_j \\ 0, & a_i < a_j \end{cases}$$

   where $(a_1, \ldots, a_n)$ is a random permutation.

   (a) What is the expected value of X_{ij}?

   (b) What is the expected number of inversions in a random permutation?

   Aside: The sorting algorithm Bubblesort sorts a list by resolving inversions one by one. The above analysis essentially determines the expected number of swaps performed by Bubblesort.