STAT7203: Applied Probability and Statistics
Weel 4 Exercises

1. A certain multiple choice exams has 20 questions, each providing 3 choices. To pass the exam one needs at least 10 out of 20 correct answers. Suppose a student knows the answers to 7 questions for certain, and fills in the remaining thirteen questions "at random".
   (a) What is the probability that the student will pass?
   (b) What is the probability that the student will get full marks?

2. The distribution of the leading digit D in many real datasets appears to follow *Benford's law* which has the following probability mass function:

$$\mathbb{P}(D = d) = \log_{10}\left(\frac{d+1}{d}\right), \qquad d \in \{1, 2, \ldots, 9\}.$$

   (a) Show this is a true probability mass function, that is $\sum_{d=1}^{9} \mathbb{P}(D = d) = 1$ and $\mathbb{P}(D = d) \geqslant 0$.
   (b) Compute numerically $\mathbb{E}D$, $\mathrm{Var}(D)$ and the standard deviation of D.

3. Let $X \sim \mathsf{Bernoulli}(\frac{1}{2})$ and $Y \sim \mathsf{Bernoulli}(\frac{1}{4})$.
   (a) Construct the joint probability mass function for (X, Y) such that X and Y are independent.
   (b) Construct the joint probability mass function for (X, Y) such that $\mathbb{P}(X = Y)$ is maximised.

4. Consider a group of n people. Assume that the birthdays of all individuals are independent and uniformly distributed over 365 days.
   (a) What is the expected number of people whose birthday is 1^{st} January?
   (b) What is the expected number of days that are not the birthday of any of the n people?
   (c) What is the probability that at least three people in a group of 23 people share the same birthday? Address this question by simulating this process 10000 times in R to give an estimate of the probability that this event occurs.

   Some useful R commands for this probelm:

   - `sample` can be used to sample from a discrete uniform distribution. Typing `sample(n,m,replace=TRUE)` simulates `m` random variables from the discrete uniform distribution on $\{1, 2, \ldots, n\}$. It is necessary to include the `replace=TRUE` argument as otherwise you will be sampling *without replacement* from $\{1, 2, \ldots, n\}$.
   - `table` takes as input a vector and returns a *table* whose column names are the unique values of the vector and the elements are the frequencies with which those unique values appear in the vector. Try

```
> x = sample(12,20,replace=TRUE)
> table(x)
> table(x)[5]
> as.vector(table(x))
> names(table(x))
```

- `max` takes as input a vector and returns the maximum value.