DATA202/STAT483 Assignment 5

Due: Thursday, 3 October 2019, Worth 10%

Assignment Questions

Q1. (14 Marks)

a. **[8 marks]** Write a custom function called `dice.roll` to simulate the rolling of a pair of fair six-sided dice (*i.e.*, all six outcomes are equally likely). The function **must** use either `sample` or `runif` for simulating rolls of each die, and the function should be consistent with the description provided below and should perform error handling. (Be sure to include your function in a code chunk in your Rmarkdown file.)

Description

Function to simulate multiple rolls of a pair of fair six-sided dice.

Usage

```
dice.roll(n = 1, seed = 0)
```

Arguments

`n` a positive integer specifying the number of times to roll the dice.
`seed` a real number to be used for setting the seed for random number generation.

Details

`dice.roll` simulates repeated rolls of a pair of fair six-sided dice and returns information on the outcomes for each of the two dice as well as the sum of their outcomes for each roll. If either of the arguments to the function is passed a vector of length greater than 1, only the first element of the vector is used and a warning message is printed. The function returns a data frame as specified below.

Value

A data frame that consists of the variables:

`Die1` the outcomes of rolls of the first die for the `n` rolls.
`Die2` the outcomes of rolls of the second die for the `n` rolls.
`Sum` the sum of the outcomes of the two dice for each of the `n` rolls.

Example

```
dice.roll(n = 15, seed = 0)
```

b. **[2 marks]** Show output for your code when it is run for the following function specifications:

````
```{r}
dice.roll(n = 2.4, seed = 2) # Q1(b)i.
dice.roll(n = 2, seed = 'a') # Q1(b)ii.
dice.roll(n = c(2, 4), seed = c(2, 1)) # Q1(b)iii.
dice.roll(n = 6, seed = 0.7) # Q1(b)iv.
```
````

c. **[4 marks]** Use your function to simulate 10,000 rolls of the die, and use `ggplot` to produce appropriate graphical displays to show relative frequencies (*i.e.*, the distribution) of the outcomes of
   i. the first die and
   ii. the sum of the two dice.

Q2. (10 Marks)

In lecture, we described how the binomial distribution arises from a series of independent and identically distributed Bernoulli trials. In particular, if $X \sim \text{Bin}(n, p)$, then it can be represented as

$$X = W_1 + W_2 + \cdots + W_n,$$

where $W_i \sim \text{Ber}(p), i = 1, 2, \ldots, n.$

a. **[4 marks]** Write a custom function `rbinomial` that has arguments identical to the `rbinom` function but uses the property that a binomial random variable can be expressed as the sum of independent and identically distributed Bernoulli random variables to generate random Bin(*n*, *p*) observations. You may **not** use the `rbinom` function but instead should use either the `sample` or `runif` function for simulating Bernoulli outcomes. Your function does not need to perform error handling (*i.e.*, you may assume that inputs will be provided correctly). (Be sure to include your function in a code chunk in your Rmarkdown file.) Run your code for the following function specification:

````
```{r}
set.seed(0)
rbinomial(10, size = 15, prob = 0.8)
```
````

b. **[2 marks]** What are the expected value and variance for a random variable $X \sim \text{Bin}(20, 0.3)$?

c. **[4 marks]** Write code to estimate the expected value and variance for $X \sim \text{Bin}(20, 0.3)$ using n = 10,000 simulations from your function `rbinomial` (or `rbinom` if your function `rbinomial` is not working corrrectly). Comment on how well these values compare with the true expected value and variance.

Q3. (6 Marks)

In the game of Yahtzee, players roll five fair six-sided dice. A large straight occurs when five consecutive numbers appear on the five dice, regardless of order. In other words, outcomes of 5, 3, 2, 4, and 6 on the five dice would correspond to a large straight, as they include the five consecutive numbers 2 to 6. Write code to simulate 10,000 rolls of five dice. Using the relative frequency of large straights in those 10,000 rolls, estimate the probability of getting a large straight when rolling five dice.

Q4. (10 Marks)

Recall the Monty Hall Problem. How does the problem change if Monty Hall does not know which doors the car and goats are located behind? This means that it is possible that Monty could open the door with the car behind it by accident, in which case we will assume that the player neither wins nor loses and the game is replayed. In this version of the game, is it a better strategy for a contestant to change doors or stick with her or his initial choice, or does it not make a difference? Simulate 10,000 plays of the game using each strategy to answer this question.

(Assignment total: 40 Marks)