CIND123 Summer 2018 - Assignment #2

Paul Ycay 500709618

July 18, 2018

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

Use RStudio for this assignment. Edit the file `assignment-2.Rmd` and insert your R code where wherever you see the string "INSERT YOUR ANSWER HERE"

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

When your are done with your answers and before submitting, save the file with the following naming convention :your **Lastname_firstname**

Submit **both** the rmd and the pdf output(or word or html) files, failing to submit **both** will be subject to mark deduction.

This assignment may make use of data provided by the `ISwR` package.

```
#library(ISwR)
library(knitr)
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)
```

Sample Question and Solution

Use `seq()` to create the vector $(1, 2, 3, \ldots, 10)$.

```
seq(1,10)
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

Question 1

Consider the probability distribution associated with rolling 3 fair dice. We can label the faces of a single die using the numbers from 1 to 6. We can therefore label the simple events in this distribution by triples of numbers from 1 to 6. Let `d1`, `d2`, and `d3` represent the labels on each of the dice.

a) Set `d1` to the sequence (1,2,. . . ,6) repeated 36 times.

```
d1<-rep(c(1:6),times=36)
```

b) Set `d2` to the sequence consisting of 6 repetitions of the sequence in which each of the numbers (1,2,. . . ,6) is repeated 6 times.

```
d2<-rep(c(1:6),each=6)
```

c) Set `d3` to the sequence in which each of the numbers (1,2,. . . ,6) is repeated 36 times.

```
d3<-rep(c(1:6),each=36)
```

d) Create a new data frame `three.dice` from `d1`, `d2`, and `d3`. Write a script to confirm that there are $6 \times 6 \times 6 = 216$ rows and each row contains a unique combination of dice labels.

```
three.dice<-data.frame(d1,d2,d3)
```

e) Since the dice are fair and independent, each simple event has the same probability, namely $\frac{1}{216}$. Add the column `P` to the data frame with this value.

```
head(cbind(three.dice,p=1/216))
```

```
##   d1 d2 d3          p
## 1  1  1  1 0.00462963
## 2  2  1  1 0.00462963
## 3  3  1  1 0.00462963
## 4  4  1  1 0.00462963
## 5  5  1  1 0.00462963
## 6  6  1  1 0.00462963
```

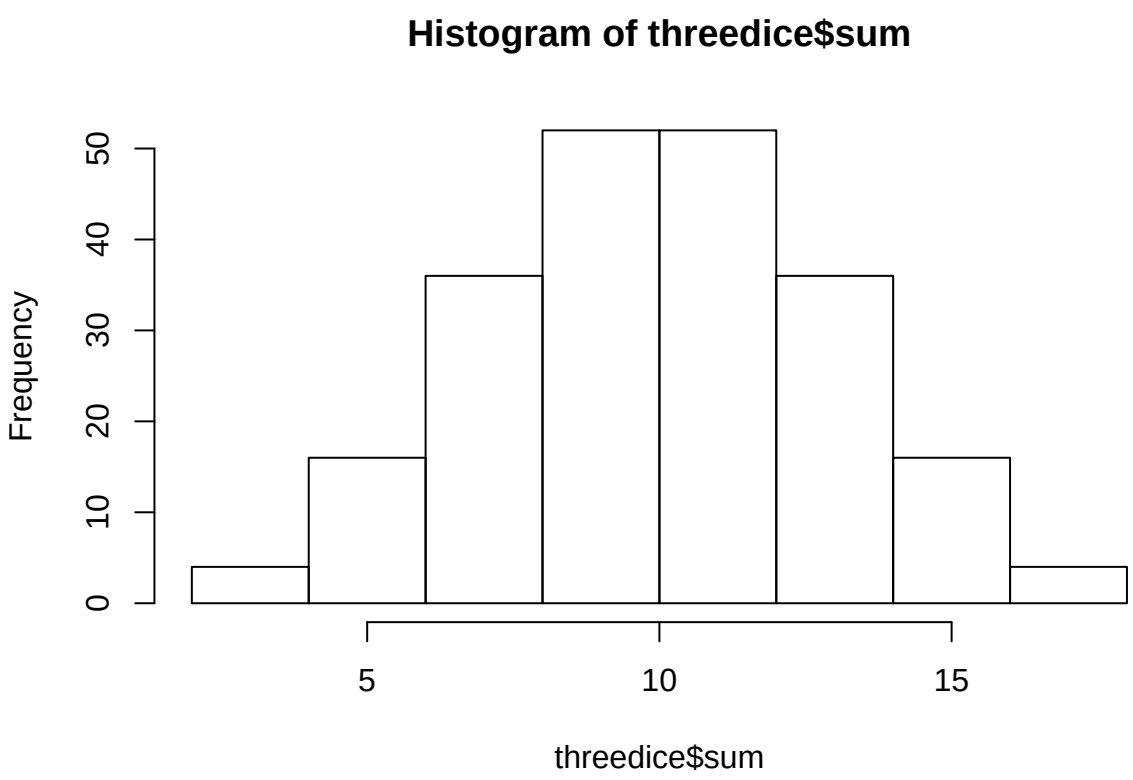
f) Add a new column `sum` equal to the sum of the dice labels. Add another new column `mean` equal to the average of the dice labels.

```
threedice<-cbind(three.dice,p=1/216,sum=(three.dice$d1+three.dice$d2+three.dice$d3),
mean=(three.dice$d1+three.dice$d2+three.dice$d3)/3)
head(threedice)
```

```
##   d1 d2 d3          p sum     mean
## 1  1  1  1 0.00462963   3 1.000000
## 2  2  1  1 0.00462963   4 1.333333
## 3  3  1  1 0.00462963   5 1.666667
## 4  4  1  1 0.00462963   6 2.000000
## 5  5  1  1 0.00462963   7 2.333333
## 6  6  1  1 0.00462963   8 2.666667
```

g) Plot a probability histogram of `three.dice$sum`.

```
hist(threedice$sum)
```



h) Compute the probability that the sum of the dice is greater then 12 and less than 18.

HINT: Use `subset()` to select the events and sum `P`.

```
h<-subset(threedice,sum>12&sum<18)
h2<-length(h$sum)
print(h2/216)
```

```
## [1] 0.2546296
```

i) Compute the probability that the sum is even.

```
i<-subset(threedice,sum%%2==0)
i2<-length(i$sum)
print(i2/216)
```

```
## [1] 0.5
```

j) Compute the probability that the mean is exactly 4.

```
j<-subset(threedice,mean==4)
j2<-length(j$mean)
print(j2/216)
```

```
## [1] 0.1157407
```

Question 2

a)You have two groups of distinctly different items, 10 in the first group and 8 in the second. If you select one item from each group, how many different pairs can you form?

```
q2a<-10*8
q2a
```

```
## [1] 80
```

b)Evaluate the following permutation P_3^5

```
factorial(5)/factorial(2)
```

```
## [1] 60
```

c)Evaluate the following combinations $C_3^5 + C_2^5$

```
(factorial(5)/(factorial(2)*factorial(3)))+(factorial(5)/(factorial(3)*factorial(2)))
```

```
## [1] 20
```

d)In how many ways can you select five people from a group of eight if the order of selection is important?

```
factorial(8)/factorial(3)
```

```
## [1] 6720
```

e)In how many ways can you select two people from a group of 20 if the order of selection is not important?

```
(factorial(20)/(factorial(18)*factorial(2)))
```

```
## [1] 190
```

Question 3

a)Use simulation to estimate the mean and variance of a binomial random variable with size = 45 and p = 0.32. One time use 100 samples and the other time use 10000 samples.

```
q3a<-rbinom(100,45,0.32)
mean(q3a)
```

```
## [1] 14.82
```

```
var(q3a)
```

```
## [1] 9.361212
```

```r
q3b<-rbinom(10000,45,0.32)
mean(q3b)
```

```
## [1] 14.3937
```

```r
var(q3b)
```

```
## [1] 9.947095
```

b) Calculate the values using the theroy (state the value and the equation in your answer),compare the values you get with the values you got in (a), wirte one sentence sumurizing the comparision. Explain the difference between 100 samples and 10000 samples and which one seems to be more accurate and why?

For the 100 samples, the mean was 14.62 and the variance was 8.6218. For the 10000 samples, the mean was 14.3601 and the variance was 10.09664.Calculating the mean using the formula mu=np, with n=45 and p=0.32 yields 14.4; calculating the variance using variance=npq, with q=0.68 yields 9.792. It seems that the 10000 samples is more accurate since the more random values it generates, the closer estimate it is to the actual mean and variance; variability on each sampling distribution decreases due to the larger samples.

Question 4

a)If there are twelve customers entering a mall per minute on average, find the probability of having seventeen or more customers entering the mall in a particular minute.

```r
ppois(16, lambda=12, lower=FALSE)   # upper tail
```

```
## [1] 0.101291
```

b) Estimate the mean and variance of a Poisson random variable in the previuos question by simulating 100 and 10,000 Poisson random numbers.

```r
q4b1<-rpois(100,12)
mean(q4b1)
```

```
## [1] 11.9
```

```r
var(q4b1)
```

```
## [1] 15.60606
```

```r
q4b2<-rpois(10000,12)
mean(q4b2)
```

```
## [1] 11.9654
```

```r
var(q4b2)
```

```
## [1] 11.90739
```

c) Compare the mean value you got in (b),with the one stated in the question. wirte one sentence summarizing the comparision. Explain the difference between 100 samples and 10000 samples and which one seems to be more accurate and why?

In the question, the mean value was 12 and the one generated by R produced 12.31 by the 100 samples and 12.0002 by the 10000 samples. The 10000 samples mean is closer to the one stated in the question since the sampling distributions will have approximately normal distributions and the variability on each sampling distribution decreases.

END of Assignment #2.