# R Assignment 2

Alireza Sheikh-Zadeh, PhD

Document format: Follow the instructions given on the web page. Always review your solution word document before submission.

Plagiarism: You are not allowed to share your write-up with your peers. It's okay to advise your peers about how to solve a problem, but you never share your own write-up.

Problem 1: 28 points

Problem 2: 52 points

Format: 20 points

## Problem 1 (28 points)

Suppose you roll a die twice in succession, getting $X_1$ and $X_2$. Then divide them, getting $Y = \frac{X_1}{X_2}$. Thus, $Y$ is discrete, ranging from a minimum of $1/6$ to a maximum of 6. In R, we can generate all possible values for $Y$ as follows.

```
# run this chunk of code:
X1 = c(1,2,3,4,5,6) # possible outcome of the first try
X2 = c(1,2,3,4,5,6) # possible outcome of the second try

Y = c() # defining an empty vector Y
# Then we divide each element of X1 by all value of X2 and record them in Y
vector
for (i in X1) {
  for (j in X2) {
     Y = c(Y, i/j) # inserting X1/X2 into Y vector
  }
}
Y

##  [1] 1.0000000 0.5000000 0.3333333 0.2500000 0.2000000 0.1666667 2.0000000
##  [8] 1.0000000 0.6666667 0.5000000 0.4000000 0.3333333 3.0000000 1.5000000
## [15] 1.0000000 0.7500000 0.6000000 0.5000000 4.0000000 2.0000000 1.3333333
## [22] 1.0000000 0.8000000 0.6666667 5.0000000 2.5000000 1.6666667 1.2500000
## [29] 1.0000000 0.8333333 6.0000000 3.0000000 2.0000000 1.5000000 1.2000000
## [36] 1.0000000

# Now Y is ready to be used in part a
```

a. Report the sorted list of $Y$ values, then Report the mean of the $Y$. Present all values by 3 decimal places. (6 points)

```r
# This is sorting of the  Y values
sort(Y)
```

```
##  [1] 0.1666667 0.2000000 0.2500000 0.3333333 0.3333333 0.4000000 0.5000000
##  [8] 0.5000000 0.5000000 0.6000000 0.6666667 0.6666667 0.7500000 0.8000000
## [15] 0.8333333 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [22] 1.2000000 1.2500000 1.3333333 1.5000000 1.5000000 1.6666667 2.0000000
## [29] 2.0000000 2.0000000 2.5000000 3.0000000 3.0000000 4.0000000 5.0000000
## [36] 6.0000000
```

```r
# This is the True Mean and expected value of Y values
mean(Y)
```

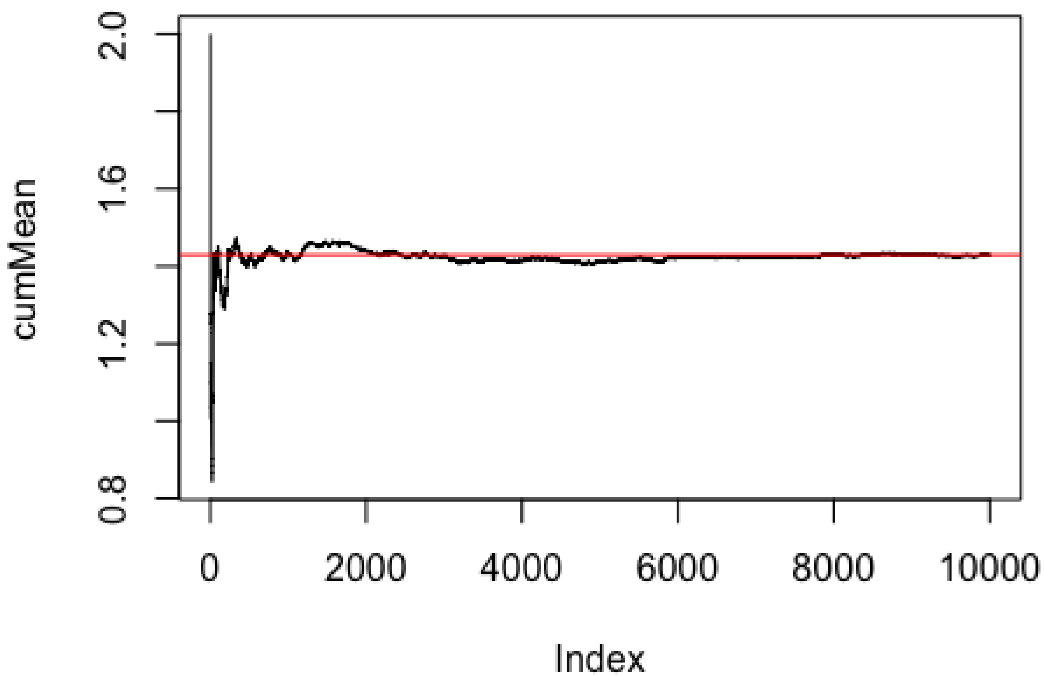
```
## [1] 1.429167
```

b. Simulate 10000 (or more) i.i.d observations $Ysim = (\frac{Xsim_1}{Xsim_2})$. Don't print the outcome of Ysim (you can only print the first 6 value). (10 points)

```r
# Simulation of 10000 outcomes for Ysim

set.seed(111)

Xsim1 <- sample(X1, 10000, replace = TRUE)
Xsim2 <- sample(X2, 10000, replace = TRUE)
Ysim <- Xsim1/Xsim2

# Did not print Ysim
```

c. Draw the graph of successive average (cumulative mean) of $Ysim$. Discuss your observations based on what you see in the graph of successive average and the answer of part a for the mean($Y$). (12 points)

```r
# Graph illustrating the cumalative mean of Ysim
# The cumMean hovers around 1.429.

n = 10000

cumSumY <- cumsum(Ysim)
cumMean <- c()

for (i in 1:n) {
  cumMean[i] <- cumSumY[i]/i
}

plot(cumMean, type = "l")
abline(h = 1.429167, col = "red")
```




```
head(cumMean)

## [1] 2.000000 1.250000 1.277778 1.108333 1.086667 1.005556

tail(cumMean)

## [1] 1.429341 1.429248 1.429355 1.429263 1.429136 1.429073

#----------------------------------------
# cumMean = cumsum(Ysim) / seq_along(Ysim)
```

## Problem 2 (52 points)

A small hotel has 10 rooms. From experience they know that 20% of the time, people who make reservations do not show up, so as a result, they overbook by accepting 12 reservations for a given night. Let $X$ be the number of no shows that night (people who don't show up).

a. What is the expected number of "no shows" that night? What is the standard deviation of that number? (8 points)

```
# The expected number of no shows
```

```
Ex <- 12 * .2
Ex

## [1] 2.4
```

b. Is $X$ continuous or discrete. (2 points)

#X is discrete

c. What is the range of $X$? (4 points)

## Range

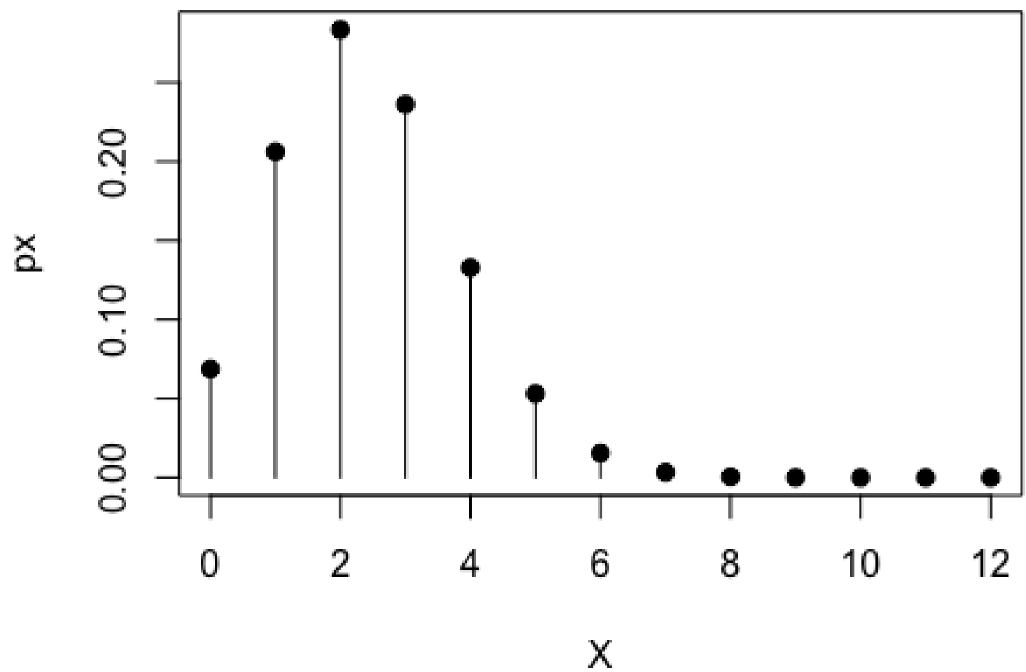
#x = 0-12

d. Find the probability of all possible value of $X$. Round by 3 decimal places. (10 points)

```
# Probability of all possible X

X <- 0:12
px <- dbinom(X, size = 12, prob = .2 )

cbind(X, px)

##        X           px
##  [1,]  0 6.871948e-02
##  [2,]  1 2.061584e-01
##  [3,]  2 2.834678e-01
##  [4,]  3 2.362232e-01
##  [5,]  4 1.328756e-01
##  [6,]  5 5.315022e-02
##  [7,]  6 1.550215e-02
##  [8,]  7 3.321889e-03
##  [9,]  8 5.190451e-04
## [10,]  9 5.767168e-05
## [11,] 10 4.325376e-06
## [12,] 11 1.966080e-07
## [13,] 12 4.096000e-09
```

e. Creat a needle plot for all peobability values of part d. (10 points)

```
# This is a needle plot.

plot(X, px, type = 'h')
points(X,px, pch=19)
```




f. What is the probability that the hotel will end up with more customers than they can handle (that is, more people with reservations than available rooms will arrive)? (4 points)

```
#If x = 0 and 1 then they will have too many customers to handle
# P(x=0)
# 0 no show > 12 shows > 10 X
# 1 no show > 11 shows > 10 X
pbinom(1, 12, .2)

## [1] 0.2748779
```

g. Simulate 10000 nights instances for the number of no-shows and call them $Xsim$; then estimate the expected value and standard deviation of the number of the simulated no-shows. What explains why the simulation results are different from the answers to a? (14 points)

```
# Simulate 10000 nights of no shows Xsim
# They are different because a is the expected number of returns while g is
the simulation of different occurences.
set.seed(123)
Xsim <- rbinom(n = 10, size = 12, prob = .2)
mean(Xsim)

## [1] 2.6
```

```
sd(Xsim)

## [1] 1.429841
```