Stat 550

Exam #1: Take-home Name: Jan Domingo

For this lab, we will run some simulations using the statistical software program R. Your solutions should contain clearly identified numerical answers and well labeled tables/figures, where appropriate. Make sure to include clear and concise interpretations (that is, clear full sentences) of any values, figures, and responses to queries posed. Place all code in an appendix, at the end of your solutions document.

1. Refer to the gambler's dispute of 1654 problems on page 109 of Dobrow (problem 3.5) in which we are trying to find the probability of obtaining at least one "double six" in throwing a pair of dice 24 times. We found that probability to be

$$P(at\ least\ 1\ double\ six = 1 - \left(\frac{35}{36}\right)^{24}$$
$$= .4914$$

(a) Simulate the gambler's dispute of 1654 and calculate the estimated probability of obtaining at least one "double six" in throwing a pair of dice 24 times. Run this experiment for a simulation size of 1,000 and compute the empirical probability

See attached appendix

(b) Compare your estimated probability with the true probability

The simulation obtained the following probability: 0.4980

Therefore $|.4914 - .4980| = .0066$ absolute value difference

2. Simulating the Birthday Problem. Recall the birthday problem deals with finding the probability that two people in a group of size k will have the same birth date (assuming no leap years). We wanted to find how many people are needed in order for the probability at least two individuals share a birthday is at least 50%. We found the answer to be 23 by first finding the probability that nobody shares the same birthday and subtracting that from 1.

(a) Simulate the birthday problem and calculate the estimated probability that two people in a group size of 23 will have the same birthday. Run this simulation for 1,000, 10,000, and 50,000 times and compute the proportion of these simulations in which at least one pair of individuals shared the same birthday

We obtained the following estimated proportions for each simulation of 1,000, 10,000, and 50,000 trials respectively: 0.5220, 0.5055, 0.5104

(b) Compare each of your empirical proportions against the true proportion of

$$P(B) = 1 - \frac{365P23}{365^{23}} = .5073$$

For 1,000 trials: $|.5073 - .5220| = .0147$ absolute value difference
For 10,000 trials: $|.5073 - .5055| = .0018$ absolute value difference
For 50,000 trials: $|.5073 - .5104| = .0031$ absolute value difference

3. Refer to the dice game on page 168 of Dobrow (problem 4.3) in which you pay \$10 to play \$10 to play. If you roll a 1, 2, or 3, you lose your money. If you roll a 4 or 5, you get your money back. If you roll a 6, you win \$24. Let $w = c\,(-10, -10, -10, 0, 0, 14)$, the winnings vector for each of the six possible die outcomes.

(a) Simulate the game 1,000 times and present the distribution of your winnings x as a table

x	-10	0	14
f(x) (probability distribution function)	0.479	0.350	0.171
F(x) (cumulative distribution function)	0.479	0.829	1.000

(b) Calculate the sample mean, sample standard deviation, and the sample variance of the winnings from (a)

Sample Mean: -2.396
Sample Standard Deviation: 8.703
Sample Variance: 75.7509

(c) Compare the mean and the standard deviation from your simulation with the true mean and standard deviation of this dice game.

True mean: $-10\left(\frac{1}{6}\right) + -10\left(\frac{1}{6}\right) + -10\left(\frac{1}{6}\right) + 0\left(\frac{1}{6}\right) + 0\left(\frac{1}{6}\right) + 24\left(\frac{1}{6}\right)$ = -1
True Standard Deviation:
First, we find the variance:

$$V[X] = \left(-10-(-1)\right)^2 * \left(\frac{1}{6}\right) + \left(-10-(-1)\right)^2 * \left(\frac{1}{6}\right)$$
$$+ \left(-10-(-1)\right)^2 * \left(\frac{1}{6}\right) + \left(0-(-1)\right)^2 * \left(\frac{1}{6}\right)$$
$$+ \left(0-(-1)\right)^2 * \left(\frac{1}{6}\right) + \left(24-(-1)\right)^2 * \left(\frac{1}{6}\right) +$$
$$= 104.1667$$

Now we take the square root of the variance to find the standard deviation:

Standard Deviation: $\sqrt{V[X]} = \sqrt{104.1667} = 10.2062$

Comparing the true mean with sample mean:

$|-1-(-2.396)| = 1.396$ absolute value difference

Comparing the true standard deviation with sample deviation

$|10.2062 - 8.703| = 1.5032$ absolute value difference

```r
#R Code for Lab 3

#1
#(a)

#This function calculates the probability of a "double six"
showing up in 24 pair rolls for n amount of times
dieSim = function(n) {
  success = 0 #"Double six" roll counter
  for (i in 1:n) {
    rolls = replicate(24, sum(sample(1:6,2,replace=TRUE)))
    if (sum(rolls == 12) > 0) {
      success = success + 1 #increment counter if at least one
double six appears)
    }
  }
  return (success/n)
}

prob1000 = dieSim(1000) #Calculates the probability of "double
six" occurring in 24 die pair rolls, 1000 times


#=================================================================

#2
#a

bdayProb = function(n) {
  success = 0
  for (i in 1:n) {
    bdays = sample(1:365, 23, replace = TRUE)
    sort(bdays)
    bdayFreq = table(bdays)
    if (max(bdayFreq) >=2) { #Returns maximum frequency count
      success = success + 1
    }
  }
  return (success/n)
}

prob1000 = bdayProb(1000)
prob10000 = bdayProb(10000)
prob50000 = bdayProb(50000)


#=================================================================
```

```
#3
#a

w = c(-10, -10, -10, 0, 0, 14)
winnings = sample (w, 1000, replace = TRUE)  #Result of 1000
trials saved into winnings
table (winnings) #Displays total count for -10, 0, and 14
## winnings
## -10   0  14
## 479 350 171
table (winnings)/1000 #Displays the probability distribution
function
## winnings
##   -10     0    14
## 0.479 0.350 0.171
winProp = table(winnings)/1000
cumsum(winProp) #Displays the cumulative distribution function
(CDF)
##   -10     0    14
## 0.479 0.829 1.000
mean(winnings)  #Displays the sample mean
## [1] -2.396
sd(winnings)    #Displays the sample standard deviation
## [1] 8.703
var(winnings)   #Displays the sample variance
## [1] 75.7509
```