NBA Hackathon Problem 1

Chia-Wei Hsu, Zeyu Li

August 4, 2017

(a)

Approach 1: Exact answer

```
no_closs <- function(win){
  ans <- choose((win + 1), (82 - win) ) * 0.8^win * 0.2^(82 - win)

  return(ans)
}
### Probability of no 2 consecutive losses
x <- sum(sapply(41:82, no_closs))
x

## [1] 0.05881686
```

The exact probability that the Warriors would never lose consecutive games at any point during an 82-game season is as follows,

$$\sum_{x=41}^{82} \binom{x+1}{82-x} 0.8^x 0.2^{82-x}$$

Above sum is approximately 0.0588169.

Approach 2: Simulation

Below code chunk is the function that repeats n times of simulating 900 samples of 82-game seasons.

```
season_simulation <- function(n = 1000){
    # simulate 900 82-game seasons
    sample <- 900
    # create a container for storing experiment outcomes
    container <- matrix(NA, ncol = sample, nrow = n)

    for(j in 1:n){
        for(i in 1:sample){
            # set-up
            sample.space <- c(0,1) # win -- 1; loss -- 0
            theta <- 0.8 # Warriors have 80% chance of winning each game
            N <- 82 # 82 games
```

```r
            # Similar to biased coin flips
            flips <- rbinom(n = N, size = 1, prob = theta)
            # next check for consecutive losses (consecutive 0s)
            r0 <- rle(flips) # list running lengths
            x <- r0$length[r0$values == 0] # count running length of 0s
            boo <- sum(x != 1) == 0 # consecutive occurrence of 0 or not;
            # FALSE : there are consecutive loss in a simulated 82-game season
            # TRUE : there are no consecutive loss in a simulated 82-game season
            container[j,i] <- boo
        }
    }

    return(container)
}
```

Repeat the process for 1000 times.

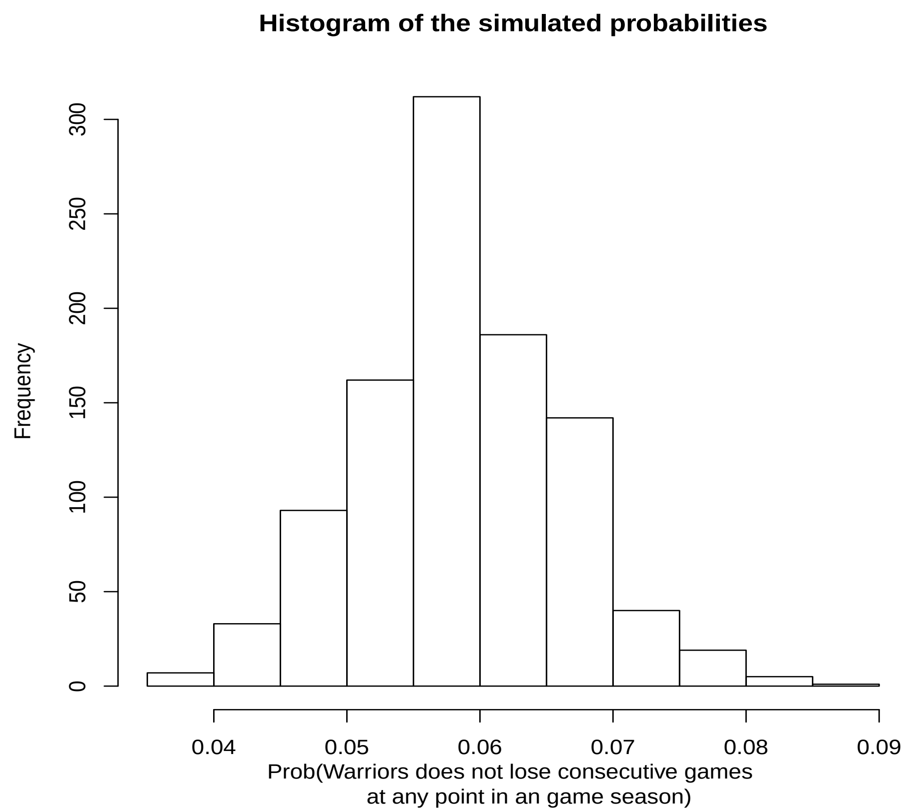
```r
# (a)
set.seed(123456789) # set starting point
sim <- season_simulation(n = 1000)
y <- rowSums(sim)/ncol(sim) # Prob(no consecutive losses)
p <- mean(y)
se <- sd(y)/sqrt(900)
p

## [1] 0.05891111

se

## [1] 0.0002582845
```

```r
hist(y,
     main = "Histogram of the simulated probabilities",
     xlab = "Prob(Warriors does not lose consecutive games
     at any point in an game season)")
```




```
# 95% confidence interval
upp <- p + 1.96 * se
low <- p - 1.96 * se
upp_per <- upp*100
low_per <- low*100
```

(Note: code for simulation see separate attached file). We build a simulator that simulates many 82-game seasons. Each 82-game season is binomially distributed with a success probability of 0.8 ($p = 0.8$) and a failure probability of 0.2 ($q = 1 - p = 0.2$). We draw random samples of size 900 from the binomial distribution and count the number of samples where the Warriors does not lose consecutive games at any point in a given 82-game season. This way we obtain an approximated probability. We repeat this process n times (we choose n to be 1000 to ensure that law of large numbers applies). After repeating for 1000 times, we obtain an averaged probability of 0.0589111 and a standard error (variability) of 2.582845×10^{-4}. We are 95% confidence that the true probability lies within the interval $(0.0584049, 0.0594173)$.

(b)

The null hypothesis is that the Warriors would never lose consecutive games at any point during an 82-game season, that is $\mathbb{P}(\textit{no consecutive losses}) = 100\%$. From above simulations, we can calculate a 95% confidence interval, which is $(5.8404873\%, 5.9417349\%)$. Since it does not contain 100%, we reject the null hypothesis.

(c)

```
# (c)
# At least game win percentage
least_per <- function(percentage){
  ans <- 0
  for(i in 41 :82){
    ans <- ans + choose((i + 1), (82 - i) ) * percentage^i *
          (1-percentage)^(82-i)
  }
  ans - 0.5
}
z <- uniroot(least_per,c(0,1))
z <- z$root
z_per <- z*100
### At least winning percentage
#0.903776165, this is tested by a lot of numbers to get different answer from
#the uniroot function
```

This question is easy to derive from the exact solution in the form presented in (a). The answer is calculated to be 90.3772241%.