Homework 4: Monte Carlo Simulations

Jenny Wang (71401898)

Due: November 10, 2019 by 11:59 pm

In this homework you will practice coding different Monte Carlo simulations. We use the example of baseball here, but you do not need to know any specifics about the game - just that the Red Sox and Astros are two US baseball teams.

Assume the baseball playoffs are about to start. During the first round of the playoffs, teams play a best of a five game series. After the first round, they play a seven game series.

Question 1

The Red Sox and Astros are playing a five game series. Assume the teams are equally good. This means each game is like a coin toss. Build a Monte Carlo simulation to determine the probability that the Red Sox win the series. (Hint: start by creating a function `series_outcome` similar to the `roulette` function from lecture and lab.)

```r
set.seed(42) # For reproducible results

n <- c(5, 7, 3) # Best out of n games
B <- 10000 # Repeat B times

series_outcome <- function(n, prob) {
  X <- sample(c(1, 0), # Let 1 represent wins and 0 represent losses
              n, replace=TRUE, # Take n samples with replacement
              prob=c(prob, 1-prob)) # Respective probabilities for wins and losses
  return(sum(X) >= ceiling(n/2)) # Best out of n games
}

sim_series <- function(n, prob, iters) {
  S <- replicate(iters, series_outcome(n, prob))
  return(length(S[S]) / length(S)) # Get probability from simulation
}

sim_series(n=n[1], prob=0.5, iters=B)
```

```
## [1] 0.5034
```

After running the simulation 10,000 times, the probability that the Red Sox win the 5-game series given that the two teams have equal chances of winning is 0.503. This makes intuitive sense: with repeated sampling, the probability of the Red Sox winning would approximate the true probability of the population. In this case, this would give the Red Sox approximately a 50% chance of winning, which was nicely demonstrated by the simulation.

Question 2

The answer to Question 1 is not surprising. What if one of the teams is better? Compute the probability that the Red Sox win the series if the Astros are better and have a 60% chance of winning each game.

```r
sim_series(n=n[1], prob=0.4, iters=B)
```

```
## [1] 0.3192
```

After running the simulation 10,000 times, the probability that the Red Sox win the 5-game series given that they have a 40% chance of winning is 0.319.

Question 3

How does this probability change if instead of five games, they play seven? How about three? What law did you learn in lecture that explains this?

```r
sim_series(n=n[2], prob=0.4, iters=B)
```

```
## [1] 0.2844
```

```r
sim_series(n=n[3], prob=0.4, iters=B)
```

```
## [1] 0.3456
```

The probability of the Red Sox winning a 7-game series is the smallest (0.284), followed by a 5-game series (0.319), followed by a 3-game series (0.346). By the law of large numbers, because the odds in favor of the Red Sox is winning is less than 50%, the more games they play, the smaller chance they would have at winning the series. At 3 games, there is more variability as to whether or not the better team would win. This variability evens out when the number of games played in a series increases, tipping the odds of winning in the Astros' favor.

Question 4

Now, assume again that the two teams are equally good. What is the probability that the Red Sox still win the series if they lose the first game? Do this for a five game and seven game series.

```r
# For a 5-game series, subtract probability of having
# 2 or fewer wins out of 4 games, since the first game
# is already lost
1 - pbinom(2, size=4, prob=0.5)
```

```
## [1] 0.3125
```

```r
# For a 7-game series, subtract probability of having
# 3 or fewer wins out of 6 games, since the first game
# is already lost
1 - pbinom(3, size=6, prob=0.5)
```

```
## [1] 0.34375
```

Given that the two teams are equally good, the probability that the Red Sox still win the series if they lose the first game is 0.312 and 0.344 for a 5- and 7-game series, respectively.