Assessment 03 - The Addition Rule

Gabriele Mineo - Harvard Data Science Professional

The Cavs and the Warriors

Two teams, say the Cavs and the Warriors, are playing a seven game championship series. The first to win four games wins the series. The teams are equally good, so they each have a 50-50 chance of winning each game.

If the Cavs lose the first game, what is the probability that they win the series?

Instructions

- Assign the number of remaining games to the variable `n`.
- Assign a variable `outcomes` as a vector of possible outcomes in a single game, where 0 indicates a loss and 1 indicates a win for the Cavs.
- Assign a variable `l` to a list of all possible outcomes in all remaining games. Use the `rep` function to create a list of `n` games, where each game consists of `list(outcomes)`.
- Use the `expand.grid` function to create a data frame containing all the combinations of possible outcomes of the remaining games.
- Use the `rowSums` function to identify which combinations of game outcomes result in the Cavs winning the number of games necessary to win the series.
- Use the `mean` function to calculate the proportion of outcomes that result in the Cavs winning the series and print your answer to the console.

```
# Assign a variable 'n' as the number of remaining games.
n <- 6

# Assign a variable 'l' to a list of possible game outcomes, where 0 indicates a loss and 1 indicates a
l <- list(0:1)

# Create a data frame named 'possibilities' that contains all possible outcomes for the remaining games
possibilities <- expand.grid(rep(l, n))

# Create a vector named 'results' that indicates whether each row in the data frame 'possibilities' con
results <- rowSums(possibilities)>=4

# Calculate the proportion of 'results' in which the Cavs win the series. Print the outcome to the cons
mean(results)
```

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

The Cavs and the Warriors - Monte Carlo

Confirm the results of the previous question with a Monte Carlo simulation to estimate the probability of the Cavs winning the series after losing the first game.

Instructions

- Use the `replicate` function to replicate the sample code for `B <- 10000` simulations.
- Use the `sample`function to simulate a series of 6 games with random, independent outcomes of either a loss for the Cavs (0) or a win for the Cavs (1) in that order.
- Use the `sum` function to determine whether a simulated series contained at least 4 wins for the Cavs.

- Use the `mean` function to find the proportion of simulations in which the Cavs win at least 4 of the remaining games. Print your answer to the console.

```
# The variable `B` specifies the number of times we want the simulation to run. Let's run the Monte Car
B <- 10000

# Use the `set.seed` function to make sure your answer matches the expected result after random samplin
set.seed(1)

# Create an object called `results` that replicates the sample code for `B` iterations and tallies the
results <- replicate(B, {
  cavs_wins <- sample(c(0,1), 6, replace = TRUE)
  sum(cavs_wins)>=4
})

# Calculate the frequency out of `B` iterations that the Cavs won at least four games in the remainder
mean(results)
```

```
## [1] 0.3453
```

A and B play a series - part 1

Two teams, A and B, are playing a seven series game series. Team A is better than team B and has a p>0.5 chance of winning each game.

Instructions

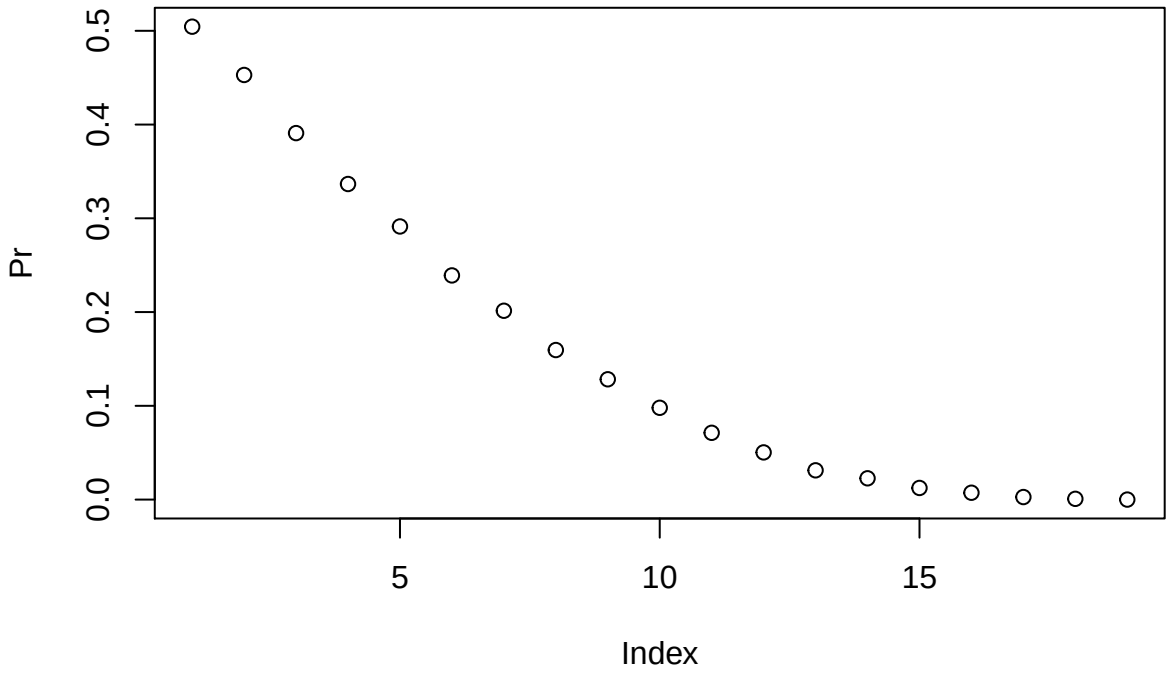
- Use the function `sapply` to compute the probability, call it `Pr` of winning for `p <- seq(0.5, 0.95, 0.025)`.
- Then plot the result `plot(p, Pr)`.

```
# Let's assign the variable 'p' as the vector of probabilities that team A will win.
p <- seq(0.5, 0.95, 0.025)

# Given a value 'p', the probability of winning the series for the underdog team B can be computed with
prob_win <- function(p){
  B <- 10000
  result <- replicate(B, {
    b_win <- sample(c(1,0), 7, replace = TRUE, prob = c(1-p, p))
    sum(b_win)>=4
    })
  mean(result)
}

# Apply the 'prob_win' function across the vector of probabilities that team A will win to determine th
Pr <- sapply(p, prob_win)

# Plot the probability 'p' on the x-axis and 'Pr' on the y-axis.
plot(Pr)
```



A and B play a series - part 2

Repeat the previous exercise, but now keep the probability that team A wins fixed at `p <- 0.75` and compute the probability for different series lengths. For example, wins in best of 1 game, 3 games, 5 games, and so on through a series that lasts 25 games.

Instructions

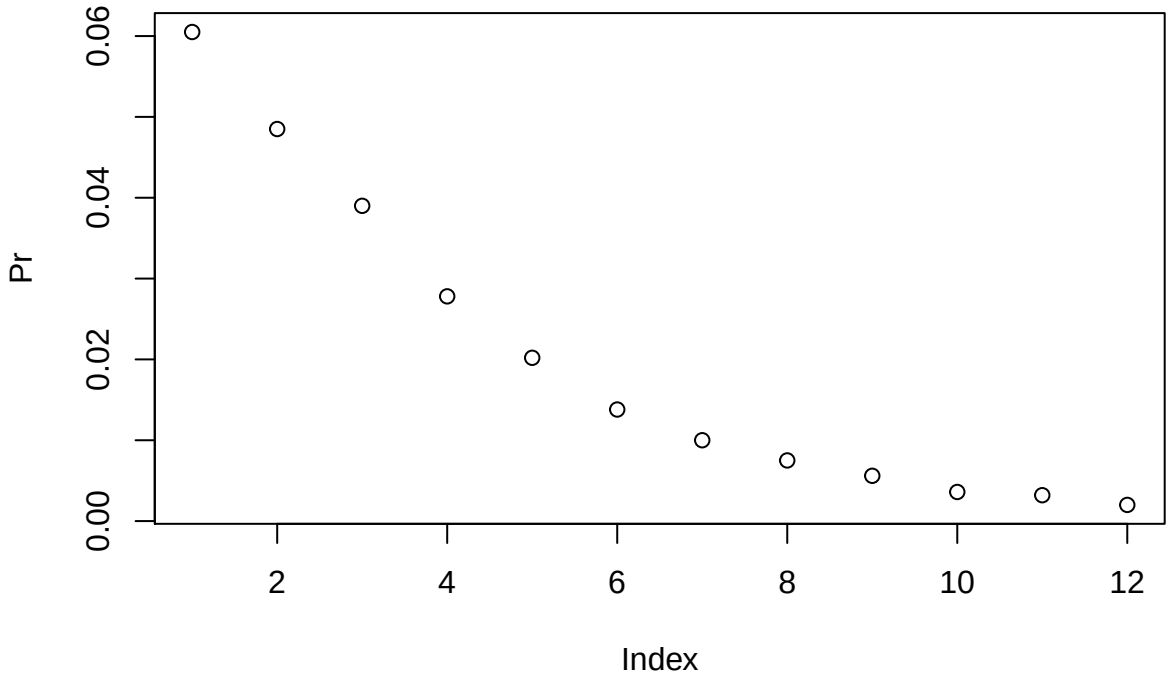
- Use the `seq` function to generate a list of odd numbers ranging from 1 to 25.
- Use the function `sapply` to compute the probability, call it `Pr`, of winning during series of different lengths.
- Then plot the result `plot(p, Pr)`.

```
# Given a value 'p', the probability of winning the series for the underdog team B can be computed with
prob_win <- function(N, p=0.75){
      B <- 10000
      result <- replicate(B, {
        b_win <- sample(c(1,0), N, replace = TRUE, prob = c(1-p, p))
        sum(b_win)>=(N+1)/2
        })
      mean(result)
    }

# Assign the variable 'N' as the vector of series lengths. Use only odd numbers ranging from 1 to 25 ga
N <- seq(2,25,2)

# Apply the 'prob_win' function across the vector of series lengths to determine the probability that t
Pr <- sapply(N, prob_win)

# Plot the number of games in the series 'N' on the x-axis and 'Pr' on the y-axis.
plot(Pr)
```


Pr
0.00
0.02
0.04
0.06
2
4
6
8
10
12
Index