Probability 04

Xueyuan Li

2022-09-26

Introduction

Today we are going to discuss the question that if home field advantage exists, how much of an impact does it have on winning the world series?

The home field advantage is the edge which a team may have when playing a game at its home stadium. For example, it is the edge the Braves may have over the Yankees when the head-to-head match-up is in Atlanta. It is the advantage the Yankees may have when the head-to-head match-up is in New York.

The World Series is a first-to-4-wins match-up between the champions of the American and National Leagues of Major League Baseball. In this assignment, you are going to use simulation and analytic methods to compare the probability of winning the World Series with and without home field advantage.

Background

Setup:

We suppose that the Braves and the Yankees are teams competing in the World Series.

The table below has the two possible schedules for each game of the series. (NYC = New York City, ATL = Atlanta)

Overall advantage	Game 1	Game 2	Game 3	Game 4	Game 5	Game 6	Game 7
Braves	ATL	ATL	NYC	NYC	NYC	ATL	ATL
Yankees	NYC	NYC	ATL	ATL	ATL	NYC	NYC

PB: be the probability that the Braves win a single head-to-head match-up with the Yankees, under the assumption that home field advantage doesn't exist.

PBH: denote the probability that the Braves win a single head-to-head match-up with the Yankees as the home team (H for home).

PBA: denote the probability that the Braves win a single head-to-head match-up with the away team (A for away).

Game location	No advantage	Advantage
ATL	PB	PBˆH= PB * 1.1

Game location	No advantage	Advantage
NYC	PB	PBˆA= 1-(1-PB)*1.1

Solution

1, Compute analytically the probability that the Braves win the world series when the sequence of game locations is {NYC, NYC, ATL, ATL, ATL, NYC, NYC}. (The code below computes the probability for the alternative sequence of game locations. Note: The code uses data.table syntax, which may be new to you. This is intentional, as a gentle way to introduce data.table.) Calculate the probability with and without home field advantage when PB=0.55. What is the difference in probabilities?

```
require(dplyr)
```

```
## Loading required package: dplyr

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
require(data.table)
```

```
## Loading required package: data.table

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
library(ggplot2)
# Get all possible outcomes
apo <- fread("all-possible-world-series-outcomes.csv")

# Home field indicator
#hfi <- c(1,1,0,0,0,1,1) #{ATL, ATL, NYC, NYC, NYC, ATL, ATL}
#hfi<- c(0,0,1,1,1,0,0)   #{NYC, NYC, ATL, ATL, ATL, NYC, NYC}
```

```r
analytic_world_ser= function(hfi= c(0,0,1,1,1,0,0),
                             pb,
                             advantage_multiplier ){

  pbh <- pb*advantage_multiplier
  pba <- 1 - (1 - pb)*advantage_multiplier

 # Calculate the probability of each possible outcome
  apo[, p := NA_real_] # Initialize new column in apo to store prob
for(i in 1:nrow(apo)){
  prob_game <- rep(1, 7)
  for(j in 1:7){
    p_win <- ifelse(hfi[j], pbh, pba)
    prob_game[j] <- case_when(
        apo[i,j,with=FALSE] == "W" ~ p_win
      , apo[i,j,with=FALSE] == "L" ~ 1 - p_win
      , TRUE ~ 1
    )
  }
  apo[i, p := prod(prob_game)] # Data.table syntax
}

# Sanity check: does sum(p) == 1?
apo[, sum(p)] # This is data.table notation

# Probability of overall World Series outcomes
apo[, sum(p), overall_outcome]


}

with_adv<-analytic_world_ser(pb=0.55,advantage_multiplier=1.1) #with advantage
with_adv
```

```
##    overall_outcome       V1
## 1:               W 0.604221
## 2:               L 0.395779
```

```r
without_adv<-analytic_world_ser(pb=0.55,advantage_multiplier=1.0) #without advantage
without_adv
```

```
##    overall_outcome        V1
## 1:               W 0.6082878
## 2:               L 0.3917122
```

```r
with_adv[[1,2]]-without_adv[[1,2]]
```

```
## [1] -0.004066825
```

```r
#analytic_world_ser(pb=0.45) #change pb=0.45 and hange 0.55 to pb
```

W:prob of B win WS

L:prob of B loss WS

We set PB=0.55, and separately discuss the probability with and without home field advantage. The first table is with advantage, and the second is without advantage. We can find that the probability without advantage is about 0.6082878, and the probability with advantage is about 0.604221. And the difference between with and without advantage is -0.004066825.

2, Calculate the same probabilities as the previous question by simulation.

```
sim.world.series=function(hfi= c(0,0,1,1,1,0,0),
                          pb = 0.55,
                          advantage_multiplier){
  pbh <- pb*advantage_multiplier
  pba <- 1 - (1 - pb)*advantage_multiplier

  num_win=0
  for(i in 1:7){
    if(hfi[i]){  #if hfi=1, return true
      p_win=pbh
    }else{    #if hfi=0, return false
      p_win=pba
    }
    game_outcome=rbinom(1,1,p_win)
    num_win=num_win +game_outcome  #initial num=0, if win add 1 to the num

    if(num_win==4 |(i-num_win)==4)  #if num_win=4, or num_loss=4, break
      break
  }
  return(num_win==4)

}
#if sim.world.series return 0, lose; if return 1, win
```

```
#without advantage
sim_ws_result=NA
for(k in 1:10000){
  sim_ws_result[k]=sim.world.series(advantage_multiplier = 1.0)
}

aa<-mean(sim_ws_result) #prob of win
aa
```

```
## [1] 0.6091
```

```
1-aa #prob of lose
```

```
## [1] 0.3909
```

```r
#with advantage
sim_ws_result=NA
for(k in 1:10000){
  sim_ws_result[k]=sim.world.series(advantage_multiplier = 1.1)
}

bb<-mean(sim_ws_result)#prob of win
bb
```

```
## [1] 0.6074
```

```r
1-bb #prob of lose
```

```
## [1] 0.3926
```

```r
bb-aa
```

```
## [1] -0.0017
```

We calculate the same probabilities as the previous question by simulation, and find that with advantage the probability of Braves winning is 0.6041, and the probability of losing is about 0.3959. And without advantage, the prob Braves win WS is 0.6123. And the difference of with and without advantage is 0.0068.

3, What is the absolute and relative error for your simulation in the previous question?

```r
#with advantage
analytic_out=analytic_world_ser(pb=0.55, advantage_multiplier = 1.1)
abs_error= abs(mean(sim_ws_result) - analytic_out$V1[1])
#first variable name is V1, first row[1]
abs_error
```

```
## [1] 0.003179028
```

```r
relative_error=abs_error/analytic_out$V1[1]
relative_error
```

```
## [1] 0.005261367
```

```r
#without advantage
analytic_out=analytic_world_ser(pb=0.55, advantage_multiplier = 1.0)
abs_error= abs(mean(sim_ws_result) - analytic_out$V1[1])
#first variable name is V1, first row[1]
abs_error
```

```
## [1] 0.0008877969
```

```
relative_error=abs_error/analytic_out$V1[1]
relative_error
```

```
## [1] 0.001459501
```

We use the function that we made from below question to calculate the absolute and relative error. And we can calculate the absolute error and relative error. They always change.

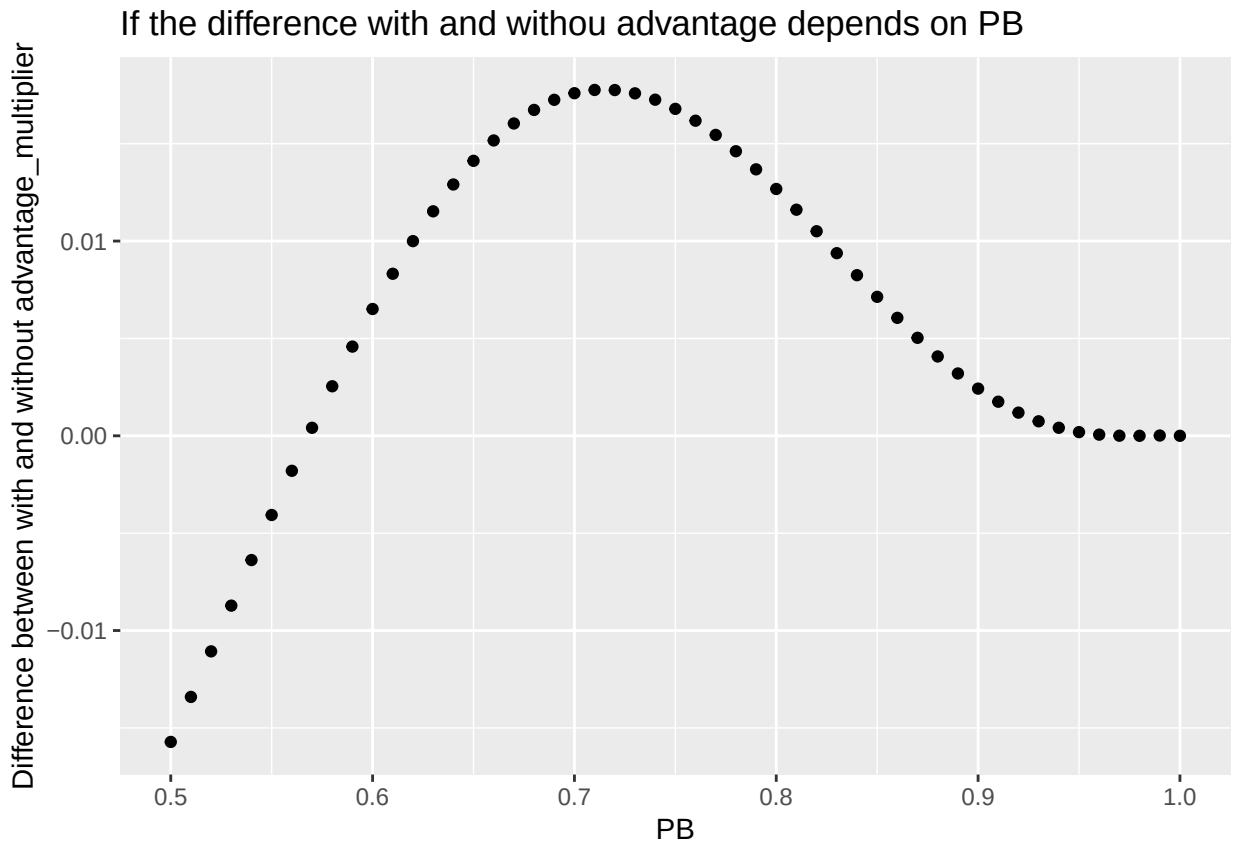
4, Does the difference in probabilities (with vs without home field advantage) depend on PB? (Generate a plot to answer this question.)

First we need to generate a plot, x-axis is PB, and y-axis is the difference between with and without home field advantage_multiplier.

```
analytic_world_ser= function(hfi= c(0,0,1,1,1,0,0),
                             pb,
                             advantage_multiplier){

    pbh <- pb*advantage_multiplier
    pba <- 1 - (1 - pb)*advantage_multiplier

 # Calculate the probability of each possible outcome
  apo[, p := NA_real_] # Initialize new column in apo to store prob
for(i in 1:nrow(apo)){
  prob_game <- rep(1, 7)
  for(j in 1:7){
    p_win <- ifelse(hfi[j], pbh, pba)
    prob_game[j] <- case_when(
        apo[i,j,with=FALSE] == "W" ~ p_win
      , apo[i,j,with=FALSE] == "L" ~ 1 - p_win
      , TRUE ~ 1
    )
  }
  apo[i, p := prod(prob_game)] # Data.table syntax
}

# Sanity check: does sum(p) == 1?
apo[, sum(p)] # This is data.table notation

# Probability of overall World Series outcomes
aa<-apo[, sum(p), overall_outcome]


}



pb=seq(0.5,1.0,by=0.01)
y<-rep(NA, length(pb))

for(i in 1:length(pb)) {
```

```
  ad<-analytic_world_ser(pb=pb[i],
                         advantage_multiplier = 1.1)[[1,2]]

  no_ad<- analytic_world_ser(pb=pb[i],
                             advantage_multiplier = 1.0)[[1,2]]

    y[i]<-round(ad-no_ad,6)
}

aa<-data.frame(pb,y)
```

```
ggplot(aa)+
  geom_point(aes(x=pb, y))+
  labs(
    x="PB",
    y="Difference between with and without advantage_multiplier",
    title="If the difference with and withou advantage depends on PB"

  )
```



We write a loop to show each PB and their matching differences between with and without home field advantage. We can find that as the PB is increasing, the difference between with and without advantage firsty increasing, but after pb=0.7, it going down. But we can still say the difference in probabilities (with vs without home field advantage) depend on PB.

5, Does the difference in probabilities (with vs without home field advantage) depend on the advantage factor? (The advantage factor in PBH and PBA is the 1.1 multiplier that results in a 10% increase for the home team. Generate a plot to answer this question.)

For each advantage_multiplier, we need to compute the probability of wining WS with and without home field advantage. x-axis is advantage_multiplier y-aixs is the difference in prob of wins between with and without home field advantage.

```
# Calculate the probability of each possible outcome
 analytic_world_ser= function(hfi= c(0,0,1,1,1,0,0),
                              pb=0.55,
                              advantage_multiplier ){

  pbh <- pb*advantage_multiplier
  pba <- 1 - (1 - pb)*advantage_multiplier

 # Calculate the probability of each possible outcome
  apo[, p := NA_real_] # Initialize new column in apo to store prob
for(i in 1:nrow(apo)){
  prob_game <- rep(1, 7)
  for(j in 1:7){
    p_win <- ifelse(hfi[j], pbh, pba)
    prob_game[j] <- case_when(
        apo[i,j,with=FALSE] == "W" ~ p_win
      , apo[i,j,with=FALSE] == "L" ~ 1 - p_win
      , TRUE ~ 1
    )
  }
  apo[i, p := prod(prob_game)] # Data.table syntax
}

# Sanity check: does sum(p) == 1?
apo[, sum(p)] # This is data.table notation

# Probability of overall World Series outcomes
apo[, sum(p), overall_outcome]


 }
```

```
advantage_multiplier=seq(1.0,2.0,by=0.01)
ad<-rep(NA, length(advantage_multiplier))

for(i in 1:length(advantage_multiplier)) {
  with_ad<-analytic_world_ser(advantage_multiplier = advantage_multiplier[i])[[1,2]]
  ad[i]=round(with_ad,6)

}

ad
```

```
##   [1] 0.608288 0.607838 0.607399 0.606969 0.606549 0.606139 0.605737 0.605345
##   [9] 0.604962 0.604587 0.604221 0.603863 0.603513 0.603171 0.602837 0.602510
##  [17] 0.602190 0.601878 0.601572 0.601273 0.600981 0.600694 0.600414 0.600140
```

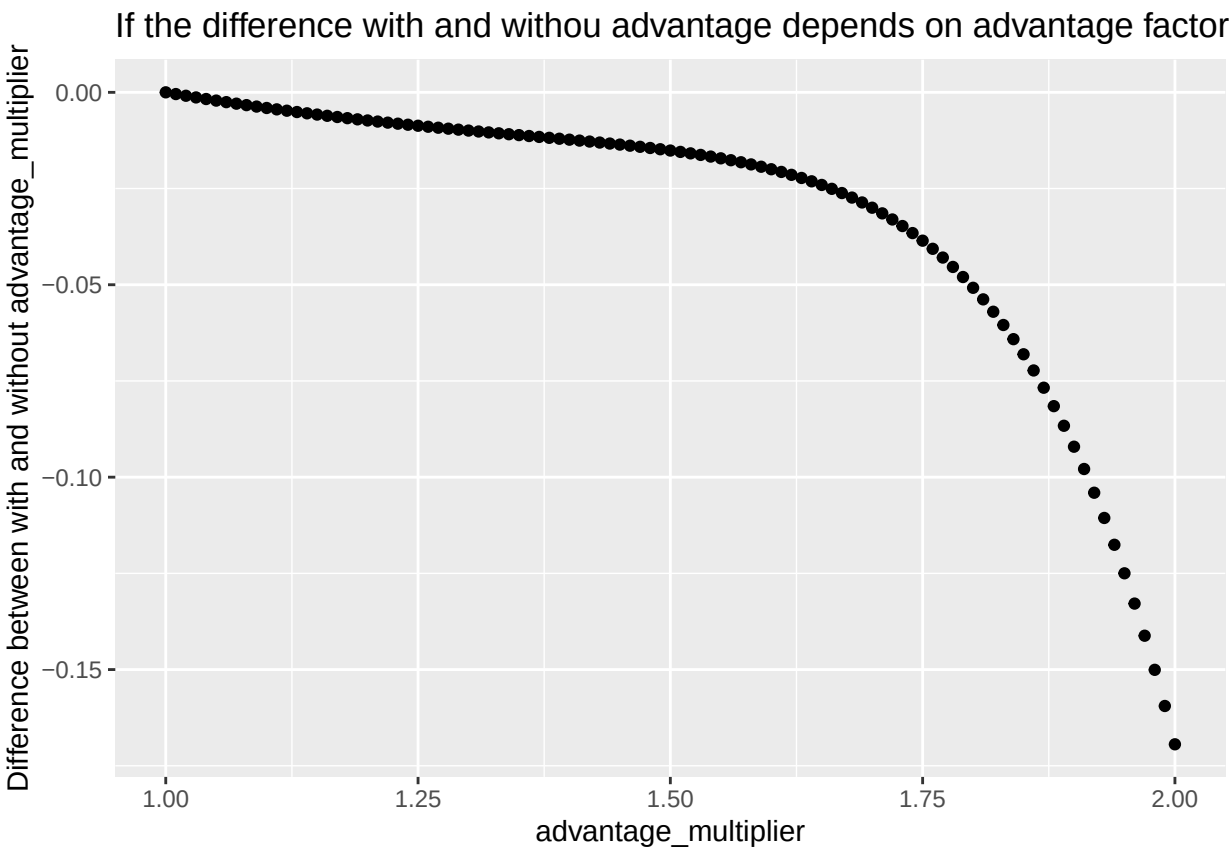
```
##  [25] 0.599870 0.599606 0.599347 0.599093 0.598842 0.598596 0.598353 0.598113
##  [33] 0.597875 0.597639 0.597405 0.597171 0.596938 0.596704 0.596469 0.596232
##  [41] 0.595992 0.595747 0.595498 0.595242 0.594979 0.594708 0.594426 0.594132
##  [49] 0.593825 0.593503 0.593164 0.592806 0.592426 0.592023 0.591594 0.591136
##  [57] 0.590646 0.590121 0.589558 0.588954 0.588304 0.587604 0.586852 0.586041
##  [65] 0.585168 0.584226 0.583212 0.582119 0.580941 0.579672 0.578305 0.576833
##  [73] 0.575249 0.573545 0.571712 0.569742 0.567626 0.565353 0.562914 0.560298
##  [81] 0.557494 0.554490 0.551274 0.547832 0.544151 0.540217 0.536015 0.531529
##  [89] 0.526743 0.521639 0.516201 0.510409 0.504245 0.497686 0.490714 0.483304
##  [97] 0.475435 0.467083 0.458221 0.448826 0.438868
```

```
no_ad<- analytic_world_ser(advantage_multiplier=1.0)[[1,2]]
no_ad
```

```
## [1] 0.6082878
```

```
diff<-ad-no_ad

aa<-data.frame(advantage_multiplier,diff)
```

```
ggplot(aa)+
  geom_point(aes(x=advantage_multiplier, y=diff))+
  labs(
    y="Difference between with and without advantage_multiplier",
    title="If the difference with and withou advantage depends on advantage factor"
  )
```



We set the pb=0.55, and change the advantage_multiplier in sequence of 1.0-2.0. We can find that as the advantage_multiplier increasing, the difference between with and without advantage is decreasing from 0.0 to about -0.20. So the difference in probabilities (with vs without home field advantage) depends on the advantage factor

---

Conclusion

So we discuss the probability of Braves wins WS with and without home field advantage. We can find that advantage does impact the probability of winning. In the questing 4 and 5, we discuss the difference in probabilities with and without home field advantage depends on both pb and the advantage factor.