2020-07-17 Riddler Express

Frank Cocuzza

This year, Major League Baseball announced it will play a shortened 60-game season, as opposed to the typical 162-game season. Baseball is a sport of numbers and statistics, and so Taylor wondered about the impact of the season's length on some famous baseball records.

Some statistics are more achievable than others in a shortened season. Suppose your true batting average is .350, meaning you have a 35 percent chance of getting a hit with every at-bat.

If you have four at-bats per game, what are your chances of batting at least .400 over the course of the 60-game season?2 And how does this compare to your chances of batting at least .400 over the course of a 162-game season?

```
## Define variables

avg_true <- 0.350 # True batting average
ABperGame <- 4 # At-bats per game
G_short <- 60 # Games in shortened season
G_reg <- 162 # Games in typical season
avg_target <- 0.400 # Batting average we're testing
```

Analysis

Since each at-bat can be treated as a discrete event resulting in one of 2 results: "hit" or "no hit," we can use a Bernoulli/Binomial distribution to represent this set of at-bats that comprises the shortened season.

```
# Calculate required parameters for short season
AB_short <- ABperGame*G_short # Determine # of ABs in a season
H_target_short <- AB_short*avg_target #Determine required # of hits to reach target avg

# Since we are looking for the probability that our average for a season is at least .400,
# we'll use 1 minus the cumulative probability function with our given inputs.
Prob.400_short <- 1 - pbinom(H_target_short, size=AB_short, prob=avg_true)
Prob.400_short
```

```
## [1] 0.04642618
```

```
# Calculate required parameters for regular-length season
AB_reg <- ABperGame*G_reg
H_target_reg <- AB_reg*avg_target

Prob.400_reg <- 1 - pbinom(H_target_reg, size=AB_reg, prob=avg_true)
Prob.400_reg
```

```
## [1] 0.003789922
```

```
Prob.400_short / Prob.400_reg
```

```
## [1] 12.24991
```

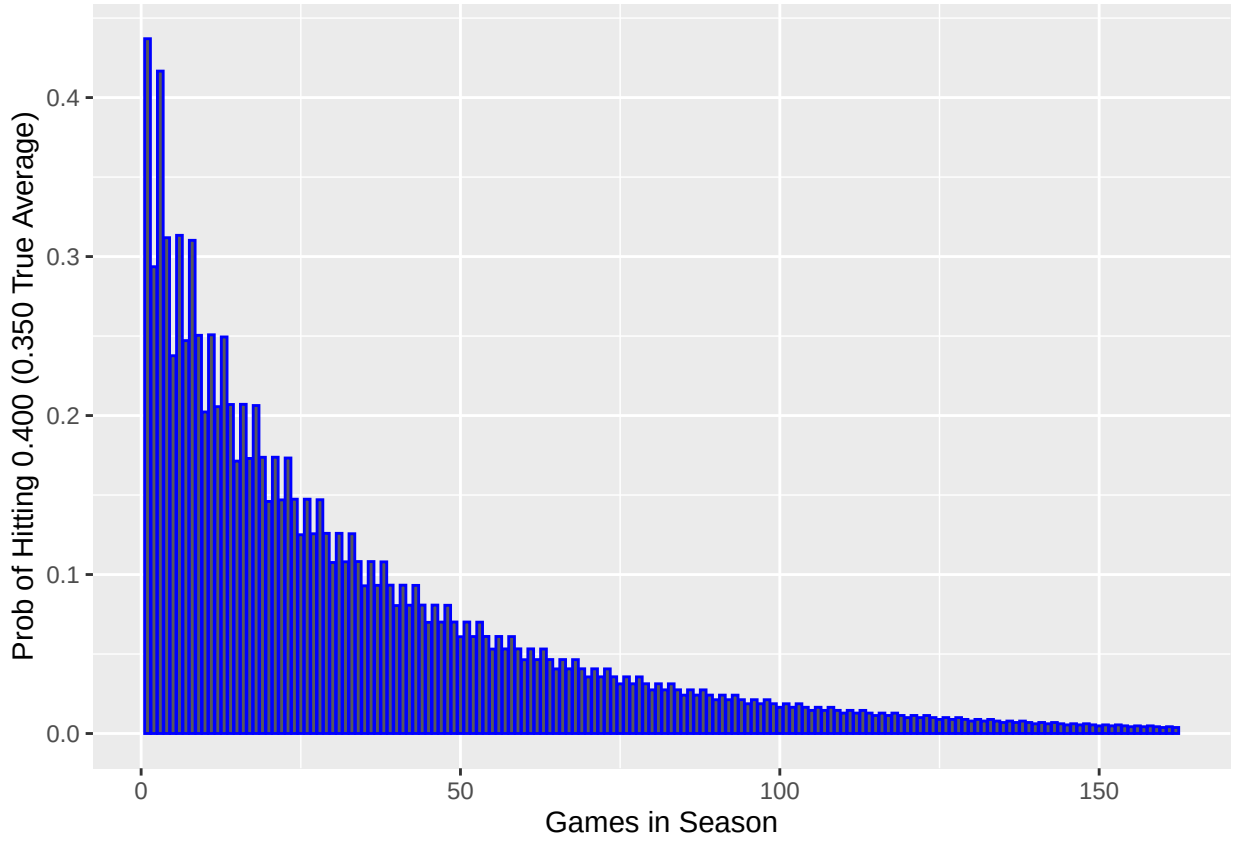
The odds that a true 0.350 batter hits 0.400 or better in a 60-game season are approximately 4.6%. This is approximately 12x more likely than the odds the same batter hits 0.400 or better in a 162-game season (~0.38%).

This relationship can be better understood if we use this same process to determine this probability for any amount of games between 1 and 162.

```
N_games <- tibble(Games = seq(from = 1, to = 162, by = 1)
                  ,AB = ABperGame*Games
                  ,Req_Hits = AB*avg_target
                  ,TrueAvg = rep(avg_true,162))

plot <-N_games %>% mutate(prob400 = 1 - pbinom(Req_Hits, size=AB, prob=TrueAvg))

ggplot(data=plot, aes(x=Games,y=prob400)) +
       geom_bar(stat = "identity", color="blue") +
       xlab("Games in Season") +
       ylab("Prob of Hitting 0.400 (0.350 True Average)")
```



Note that the probability is not a perfectly smooth curve. This is a result of the number of hits a player can get in a given season being an integer, and so the minimum number of hits required gets rounded up from the calculated, decimal value.