Assignment 2 – Baseball Simulation

Hoang Nguyen & Cuong Nguyen

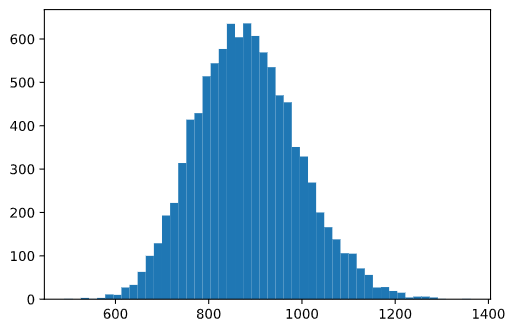
# Introduction

In baseball’s history, Roger Maris and Mickey Mantle, who were both on the New York Yankee baseball team, were racing to tie or break Ruth's record of 60 home runs in 1927. As it turned out, in 1961 Mantle only scored 54 home runs due to his bad knees, Maris, on the other hand, broke Ruth's record with 61 home runs. After breaking the record, Maris was believed to have crumbled under the pressure of the champion and did not deserve to break Ruth’s record because he was only considered as an underdog who had never hit more than 39 home runs in a year, nowhere near the 60-home-run record. We know the result of any particular at bat depends mainly on the player's ability and other factors contributed to it such as health, wind, sun, stadium lights, etc. Without that factor, the probability of hitting and failing to make the home run is 50%. Over hundreds of at bat Maris had each year, random factors average out result in homerun production increased as he became skillful and vice versa. But if those random factors do not average out how large is the aberration? In the work of Mlodinow, the author find that such rare events happen more frequently than we think. [1]

# Methodology

To demonstrate that problem further whether Maris could break Ruth record we have written a python script to discuss the probability of Maris tie or breaking Ruth’s record of 60 home run. Maris’s performance in 1960 was 1 home run for every 14.7 at bat opportunities. Meanwhile, the world’s record until 1960 was 60 home runs. Therefore, the number of opportunities needed for breaking the world’s record based on Maris’s performance is opportunities on average. In our simulation for a player with Maris’s performance we divided the game into two scenarios when the player hits at the ball (at bat):

* Let the player run (home run), the success rate is
* Count if the player successfully makes a run.
* Repeat the above steps until the player gets 60 home runs.
* Repeat the simulation 10000 times and plot the data as histogram:



*Figure 1 - With the X-axis is the number of opportunities to get 60 home runs. Y axis is the number occurrences to get 60 home runs for a specific number of opportunities.*

In the work of Mlodinow, he didn’t state the number of opportunities Maris had that year (1961). However, we find that this piece of information is crucial to determine the probability of Maris to tie or break the world's record. Therefore, we did an investigation of baseball game that year, 1961. Maris had 590 at bat (or opportunities) [2]. Hence, the probability of getting 60 home runs based on our simulation is only 0.19%. This is a very tiny probability for Maris to break the record. However, as it turned out, there is 1 player every 3 years that has the same opportunities and performance in 1961 as Maris [1]. In other words, in the period of 1927-1961, there are such candidates; and over a period of 70 years, there are such candidates. To validate Mlodinow's results, the probability of getting 60 home runs at least once over 34 seasons and 70 seasons (or 70 years) are as follow:

Probability of having at least 1 candidate to break world’s record in one season is:

Probability of having at least 1 candidate to break world’s record in T seasons (or years) is:

**Note**: the number of candidates grows as time T grows assuming all the players continue to play in every season. Also note that getting the probability of Maris scoring 60 home runs is not a numerically stable (e.g. 0.0019) due to the extremely rare event.

# Conclusion

In conclusion, these results (50.9% and 95.3%) closely match with Mlodinow’s computation despite some different assumptions in our model [1]. We conclude that Maris is very lucky that year to be the one to break the record. Not only that, we agree with the author that such a world's record can be easily broken given a large amount of time and the number of candidates.

# References

# [1] Mlodinow, L. (2008). The Drunkard's Walk: How Randomness Rules Our Lives.

# [2] Baseball Almanac, I. (n.d.). 1961 New York Yankees Hitting Stats. Retrieved from <https://www.baseball-almanac.com/teamstats/hitting.php?y=1961&t=NYA>