<u>Progress Report</u>

BAMS3216 PROJECT

Student **:** Lee Cheng Zhan	**ID :**19WMR05929
Supervisor **:** Dr Tey Siew Kian	**Date :** 04 / 07 / 2021
Title **:** Game theory in baseball	

Chapter 1

INTRODUCTION

1.1 Background

1.1.1 What is baseball

Baseball is a bat-and-pitch game played between two opposite teams who take turns batting and pitching. The game proceeds when a player on the fielding team, called the pitcher, throws a ball and a player on the batting team tries to hit the ball. The purpose of the offensive team (batting team) is to hit the ball into the field of play to allowing its team players to run the bases, having them advance counter-clockwise around four bases to score what are called "runs". The purpose of the defensive team (fielding team) is to prevent the batters becoming runners, and to prevent runners advance around the bases.The batter have 3 chance to hit the ball when pitcher throw a ball into the strike zone(the space above home plate and between the batter's knees and the midpoint of their torso) also say as a goodball. But if the pitcher throw the ball 3 times outside the strike zone and batters can automaic go to the next base when them do not swing the ball. A run is scored when a runner legality advances around the four bases in order and touches the home plate (the place where the player started as a batter). The team that scores the most runs is the winner in the end of the games.

1.1.2 What is game theory

Game Theory objectives to help us understand the situations in which decision-makers interact. Like other sciences, game theory consists of a collection of models. A model is an abstraction that we use to understand our observations and experiences in real life, “understanding” may not not clear-cut. At least, it entails our perceiving relationships between situations and create a new principles that apply to a series of problems, so that we can incorporate the new situations we encounter into our thinking. Game theory has been widely recognized as an important tool in many fields. As of 2014, 11 game theorists have been awarded the Nobel Prize in Economics with the awarding of the Nobel Memorial Prize in Economics to game theorist Jean Tirole. John Maynard Smith was awarded the Crafoord Prize for his application of evolutionary game theory.

1.1.3 The theory of rational choice

Rational choice theory is an integral part of many models in game theory. In short, this theory holds that the decision maker chooses the best action among all available actions according to her preferences. There are no qualitative restrictions on the decision maker's preferences; her "rationality" lies in the consistency of her decisions when faced with different sets of available actions, not in the nature of her likes and dislikes.

1.1.4 The type of game theory

The type of game theory are classified as three type ,which is single, two-player and multiplayer game. The essence of the single player game is the optimization problem of the individual. Two-player games are the most common , studied, and the most basic and useful type of the game. The prisoner's dilemma, coin guessing, and Tian Ji horse race are all two-player games. There are many possibilities for two player to play the game, and the interests of the game parties may or may not be in the same direction and the multiplayer games may be have spoilers: strategic choices that do not affect their own interests, but have a significant, sometimes decisive, impact on the interests of other parties to the game.

1.1.5 Limitations of Game Theory

The biggest problem with game theory is that, like most other economic models, it depend on the assumption that people are rational actors, self-interested, and utility maximizers. Of course, we are social beings and we do cooperate and do care about the welfare of others, often at our own expense. Game theory cannot explain the fact that in some cases we may get into a Nash equilibrium and other times not, depending on the social context and who the players are.

1.2 Objectives

Sports games and tournaments have a lot of data, so it is a fruitful area to study and do betting. Baseball is a great game to analyze from a game theory perspective because the strategic and tactical decisions that are constantly being made on the field and between games are very complex and huge. In each case, there are countless participants (baseball players, coaches, team managers and owners) with different goals and payoff, making hundreds of pitch-by-pitch decisions over the course of an inning, a game and a season. Most baseball games are played in a one-on-one interaction. The pitcher throws the ball and the batter hits the ball. In that sense, in many cases it is fairly easy to judge the performance of batters and pitchers based on the outcome. Home runs, walks, strikeouts, all of these are based entirely on one-on-one matchups. All Major League Baseball (MLB) teams have their own departments dedicated to this type of analysis, and the time is ripe for analysis, as almost every aspect of the game is tracked with statistical information.

As time goes by, it has become common in sports to use mathematical methods to find the best strategy in a game. In this project, we try to use game theory to determine the best strategy in a baseball game. First, we will use the Nash equilibrium model to determine the best strategy for each pitcher and batter. In this statement, I will figure out what type of pitch the pitcher should throw, e.g., goodball or badball, and whether the batter should swing or not swing at the ball. I will try to figure this out through game theory. After that, I expect to use the data related to each player to get the payoff for each decision and make the model more accurate.

The Objectives we focusing in this project is to decide the best strategy for each pitcher and hitter. The objective of this project is help to the team and player achieve better pitch and bat results by finding the best strategy. In this project, we will modelling the situation by two-player games with finite games and the strategies for batter is swing and no-swing and the strategies for pitcher is pitch a goodball or badball.

1.3 Project scope

While during this project, I start by understanding the rule and background of the baseball by social media and article. In the “understanding sabermetric” , I learn about the important of the data and their abbreviation eg. OBP(on-base persentage), BA(batting average), SLG(slugging average) and how to calculate them. Base on it , I make a model to decide the ability of the batter. Beside that, in order to understand game theory, I have read some thesis and book that relevant to game theory like “an introduction to game theory” or “essentials of game theory” to understand what is game theory and their type and limitation. The Prisoners Dilemma is the method that help me to decide the best strategy. Since they are many different model in game theory learn the relevant knowledge in free learning sites like youtube or coursera also a method that how I investigate to this project.

1.4 Planning

This report is incomplete yet, still need to enhance more before submit the draft report. At first, learning python for the modelling part is significant because we need to simulate the situation by programming and do revision for the machine learning and decide which method should use to get the correlation between team run scored and the defensive data to consider which data should be use to consider the player is good at pitch or not. After that ,need to enhance the payoff prediction because the payoff now is a bit weird. Other than that, continue to learn and read the knowledge that relevant to my topic also a action that I take to improve my project.

Chapter 2
LITERATURE REVIEW

A Nash equilibrium is a solution to a non-cooperative game involving two or more players. In a Nash equilibrium, each player is assumed to know the equilibrium strategies of the other players, and no player gains anything by simply changing his or her strategy. Essentially, it tells us what we believe rational actors will do in the game. In baseball, an example is batting (Micah Melling, 2018). A pure strategy Nash equilibrium is the fairest strategy such that each player's strategy is the best response to the other players' equilibrium strategies (leading to the highest available payoff).

The Prisoner's Dilemma is an example of a Nash equilibrium, which is a standard example of an analytic game in game theory that shows why two perfectly rational people may not cooperate, even if it seems to be in their best interests to do so. It was originally proposed by Merrill Flood and Melvin Dresher in 1950 while they were working at the RAND Corporation. Albert W. Tucker formalized this game with prison sentence rewards and named it the Prisoner's Dilemma, expressed as follows :
By examining the four possible pairs of actions in the Prisoner's Dilemma (reproduced in **Table 1**), we see that (Fink, Fink) is the only Nash equilibrium.

Player 1/Player 2	Quiet	Fink
Quiet	3 , 3	2 , 5
Fink	5 , 2	3 , 3

Table 1: Prisoner's Dilemma

The action pair (Fink, Fink) is a Nash equilibrium because (i) given that player 2 chooses Fink, player 1 is better off choosing Fink than Quiet (looking at the right column of the table, we see that Fink gives player 1 a payoff of 3, while Quiet gives her a payoff of 2). and (ii) given that Player 1 chooses Fink, Player 2 is better off choosing Fink than Quiet (looking at the bottom row of the table, we see that Fink gives Player 2 a payoff of 3, while Quiet gives her a payoff of 2).

According to (Hattie James, 2015), game theory uses mathematical models to analyze decisions. Most sports are zero-sum games in which a decision by one player (or team) will have a direct effect on the opposing player (or team). This creates an equilibrium known as the Nash equilibrium, named after mathematician John Forbes Nash. This means that if a team scores a run, it is usually at the expense of the opposing team - most likely based on an error by the outfielder or a hit by the pitcher. In the case of pitching, game theory - and in particular the use of Nash equilibria - can be used to predict the strategic purpose of pitching optimization. In this case, a mixed strategy is best - in game theory, when a player intends to keep his opponent guessing.

According to (Matt at el, 2012), it considers the structure of the following pair of games, where the batter is always better off when the action is fastball/swing or curveball/no swing, and the pitcher is always better off when the action is fastball/no swing and curveball/swing. The difference between these two examples is that a pitcher who has a good curveball has an even higher expected return on the ball/swing. In situation 1, the pitcher has a strike 50% of the time and the batter swings 50% of the time. In situation 2, the pitcher throws strikes 56% of the time and the batter swings 44% of the time. A pitcher with a better out pitch will throw more fastballs to encourage the batter to actually swing at the fastball when they throw it. It also suggests that the batter observes a "signal" example after the pitch is thrown: the batter can consider the type of pitch that follows the pitcher's motion, and he uses reverse induction to solve this problem.

According to (Gelblum, Laucys, Saperstein, and Sodha, 2010), this article delves into the specific scenario that occurred in the 8th inning of the Dodgers vs. Giants game on April 18, 2010. The most interesting finding of the article talks about pitch selection and the expected balance of mixed strategies, and for simple reasons, does not include pitch location. Not surprisingly, the expected outcome of the batter's RAA (run average) was positive only when the batter correctly guessed the pitch being thrown, and the expected outcome of the pitcher was positive only when the batter incorrectly guessed the pitch about to be thrown. Based on the negative RAA payoff to the batter for guessing the wrong pitch, the batter is least negatively affected by not guessing the fastball than by not guessing the changeup or slider. Thus, mixed strategy Nash equilibrium occurred when batters Manny Ramirez guessed fastballs 6.25%, sliders 50.02%, and changeups 43.73% of the time, while pitchers threw fastballs 43.47%, sliders 34.90%, and changeups 21.63% of the time. Batters over-expect pitches that have worse returns for hitters. wRAA measures how many runs a hitter contributes, compared to an average player - so a player with a wRAA of 0 would be considered league average.

According to (Piper at el, 2012), Plate discipline statistics tell us how often a batter swings and makes contact on certain types of pitches, or how often a pitcher induces swings or contact on certain types of pitches. Some of the terminology and calculations for these data are shown below :

O-Swing% - Percentage of pitches a batter swings at outside the strike zone

Swings at pitches outside the zone / pitches outside the zone

Z-Swing% – Percentage of pitches a batter swings at inside the strike zone

Swings at pitches inside the zone / pitches inside the zone

Swing% – Percentage of pitches a batter swings at overall.

Swings / Pitches

This data is very useful in determining the type of batter or pitcher. Based on this data, we can more accurately determine the return of batters and pitchers in each game by understanding their pitching/swinging habits and frequency.

According to (Gabriel B.Costa, Micheal R. Huber and John T.Saccoman, 2009), Batting average (BA) is found by dividing the number of base hits (H) by the total number of at-bats (AB):

$$BA = \frac{H}{AB}$$

A batter’s slugging percentage (SLG) is found by dividing the total number of bases (TB) of all base hits by the total number of times at bat:

$$SLG = \frac{TB}{AB}$$

A batter's on-base percentage (OBP) is found by dividing the total number of hits plus bases on balls (BB) plus hit by pitch (HBP) by at-bats plus bases on balls plus hit by pitch plus sacrifice flies (SF):

$$OBP = \frac{H + BB + HBP}{AB + BB + HBP + SF}$$

According to (Benjamin Baumer and Andrew Zimbalist, 2013), its shows about the correlation of the data with team scores. the corresponding correlations for OBP, BA and SLG are 0.885, 0.822 and 0.910 respectively. they show the autocorrelation of many commonly used statistics, as well as the correlation with team scores. However, they also tell us that the exact statistics (such as OPS and SLG) are not necessarily reliable in terms of the strength of their correlation with team scoring, as they are highly correlated with themselves in time. Thus, knowing a player's OPS can give us a good idea of how much he contributes to a team's offense in a given season, but it does not reveal what he might do the following season.

Chapter 3
METHODOLOGY

3.1 Data

In this project, we use the official data source from Major League Baseball (*MLB)* to decide the pitcher and hitter payoff.

3.2 Determine the best strategy

In most cases, pitch selection and hitter responses will be solved by two-player mixed strategies with finite games. That means a mixed strategy involved a player select two (or more but no unlimited) strategies with probabilities between 0 and 1, and the Nash Equilibrium requires the player to be indifferent between two (or more but no unlimited) strategies, conditional on his opponent's (potentially mixed) strategy.

Let's assume the pitcher is deciding whether to pitch or strike. If a player swings at a strike, he gets a hit and wins the game; if he swings at a pitch, he misses and loses the game. Therefore, the impact on each player's team record is reflected in the table below :

Pitcher\Batter	Swing	Take (No-swing)
Strike (Goodball)	$-2,2$	$2,-2$
Ball (Badball)	$2,-2$	$-2,2$

Table 2 : the effect on each player's team's record

To solve this game, we consider the value a batter receives from swinging and not swinging, conditional on the pitcher's given strategy "p", i.e., the probability of throwing a strike. The value of a batter's swing will be the product of his payoff from Strike/Swing (2) multiplied by the probability of this outcome occurring when he swings (p), plus his payoff from Ball/Swing (-2) multiplied by the probability of this outcome occurring when he swings (1 - p).

Therefore, we define the batter's value from swinging as:

$$\boxed{V(s) = (2)^*(P) + (-2)^*(1-p) = 4p - 2}$$

The batter's value from taking is:

$$\boxed{V(t) = (-2)^*(P) + (2)^*(1-P) = 2 - 4p}$$

Conditional on the batter's strategy "q," his probability of swinging, the pitcher's value from throwing a strike is:

$$\boxed{V(g) = (-2) * (q) + (2) * (1-q) = 2 - 4q}$$

The pitcher's value of throwing a ball when the batter swings with probability "q" is:

$$\boxed{V(b) = (2) * (q) + (-2) * (1-q) = 4q - 2}$$

If 2p - 1 > 1 - 2p (i.e., when p (the probability of a pitcher throwing a strike) > 0.5), the batter will always tend to swing, so in this case his strategy will be to always swing (i.e., set q = 1). Similarly, when p < 0.5, the batter will always take, so his strategy will be to always take (i.e., set q = 0). However, when 1-2q>2q-1 (i.e., q<0.5), the pitcher will always tend to throw strikes, so in this case his strategy will be to always throw strikes (i.e., set p=1). Similarly, when q>0.5, the pitcher will always throw a pitch, so his strategy is to throw a pitch (i.e., set p=0).

The only Nash equilibrium that exists is when p=0.5 and q=0.5. In other words, the only possible outcome is for the pitcher to throw strikes in 50% of the full cases and the batter to swing at 50% of the cases. When this happens, the expected value for both the batter and the pitcher is equal to 0. However, if we can find the probabilities of p and q from the data, it means that we can decide the best strategy for the player.

However, this is just one of those situations where the payoff varies due to the circumstances, e.g. if a good hitter with a good swing meets a bad hitter, the hitter can see the type of ball by its movement or by the last pitch the hitter can swing at (if the hitter throws a pitch on three swings (bad pitch) he will be replaced).

3.3 Methodology / Model

Since everyone have different bat and pitch skill, the payoff can not be consistant and it is hard to determine the payoff for the strategy. So we need to use the data to decide the payoff of each player. The equation of predicting the Strength of a player can be :

$$S = B_0X_0 + B_1X_1 + B_2X_2$$

S = The ability of the batter

B_0 = Correlation between team run scored and OBP

X_0 = Batter's OBP

B_1 = Correlation between team run scored and BA

X_1 = Batter's BA

B_2 = Correlation between team run scored and SLG

X_2 = Batter's SLG

To decide the ability of the Batter is weak or strong we subsitute Batter's OBP, BA and SLG by average OBP, BA and SLG with 0.323, 0.252 and 0.435 respectively and subsitute the correlation with Benjamin Baumer and Andrew Zimbalist's data which SLG are 0.885, 0.822 and 0.910 respectively :

Average S = 0.885*0.323+0.822*0.252+0.910*0.435 = 0.888849

So we can decided that if a batter ability below the average S which is 0.88849 can be consider weak and if above it can be consider strong. Now we use the Prisoners Dilemma to decide the strategy for batter that be consider strong, our payoff become :

Pitcher\Strong Batter	Swing	Take (No-swing)
Strike (Goodball)	$-3,3$	$2,-2$
Ball (Badball)	$2,-2$	$-2,2$

Table 3 : the effect on each player's team's record when strong batter is playing

we consider the value a batter receives from swinging and not swinging, conditional on the pitcher's given strategy "p", i.e., the probability of throwing a strike. The value of a batter's swing will be the product of his payoff from Strike/Swing (3) multiplied by the probability of this outcome occurring when he swings (p), plus his payoff from Ball/Swing (-2) multiplied by the probability of this outcome occurring when he swings (1 - p).

Therefore, we define the batter's value from swinging as:

$$\boxed{V(s) = (3)^*(P) + (-2)^*(1-p) = 5p - 2}$$

The batter's value from taking is:

$$\boxed{V(t) = (-2)^*(P) + (2)^*(1-P) = 2 - 4p}$$

Conditional on the batter's strategy "q," his probability of swinging, the pitcher's value from throwing a strike is:

$$\boxed{V(g) = (-3) * (q) + (2) * (1-q) = 2 - 5q}$$

The pitcher's value of throwing a ball when the batter swings with probability "q" is:

$$\boxed{V(b) = (2) * (q) + (-2) * (1-q) = 4q - 2}$$

If 5p - 2 > 1 - 2p (i.e., when p (the probability of a pitcher throwing a strike) > 0.429), the batter will always tend to swing, so in this case his strategy will be to always swing (i.e., set q = 1). Similarly, when p < 0.429, the batter will always take, so his strategy will be to always take (i.e., set q = 0). However, when 2-5q>2q-1 (i.e., q<0.428), the pitcher will always tend to throw strikes, so in this case his strategy will be to always throw strikes (i.e., set p=1). Similarly, when q>0.428, the pitcher will always throw a pitch, so his strategy is to throw a pitch (i.e., set p=0).

The payoff for weak batter :

Pitcher\Weak Batter	Swing	Take (No-swing)
Strike (Goodball)	$-1,1$	$2,-2$
Ball (Badball)	$2,-2$	$-2,2$

Table 4 : the effect on each player's team's record when weak batter is playing

we consider the value a batter receives from swinging and not swinging, conditional on the pitcher's given strategy "p", i.e., the probability of throwing a strike. The value of a batter's swing will be the product of his payoff from Strike/Swing (1) multiplied by the probability of this outcome occurring when he swings (p), plus his payoff from Ball/Swing (-2) multiplied by the probability of this outcome occurring when he swings (1 - p).

Therefore, we define the batter's value from swinging as:

$$\boxed{V(s) = (1)^*(P) + (-2)^*(1-p) = 3p - 2}$$

The batter's value from taking is:

$$\boxed{V(t) = (-2)^*(P) + (2)^*(1-P) = 2 - 4p}$$

Conditional on the batter's strategy "q," his probability of swinging, the pitcher's value from throwing a strike is:

$$\boxed{V(g) = (-1) * (q) + (2) * (1-q) = 2 - 3q}$$

The pitcher's value of throwing a ball when the batter swings with probability "q" is:

$$\boxed{V(b) = (2) * (q) + (-2) * (1-q) = 4q - 2}$$

If $3p - 2 > 1 - 2p$ (i.e., when p (the probability of a pitcher throwing a strike) > 0.6), the batter will always tend to swing, so in this case his strategy will be to always swing (i.e., set $q = 1$). Similarly, when $p < 0.6$, the batter will always take, so his strategy will be to always take (i.e., set $q = 0$). However, when $2-3q>2q-1$ (i.e., $q<0.6$), the pitcher will always tend to throw strikes, so in this case his strategy will be to always throw strikes (i.e., set $p=1$). Similarly, when $q>0.6$, the pitcher will always throw a pitch, so his strategy is to throw a pitch (i.e., set $p=0$).

3.4 Discussion

At first, we study about the type of game theory and consider that , pitch selection and batter responses will be solved by two-player mixed strategies with finite games. The strategy for the pitcher is throw a goodball or badball and the strategy for the batter is swing or do not swing. During the investigate of choosing the best strategy we found that we can decide the best strategy if know the probability of strategy that opponent take eg. when pitcher have the higher probabiulity to throw a goodball then the best strategy for batter is swing. But the payoff will not be constant while playing. So we try to classified the player to strong and weak, strong player can get high payoff in the game and weak player get low payoff. While during this proccess we found that we do not have a method to decide strong player will have how many payoff and I consider strong player add 1 and weak player -1.

When decide the player is strong or weak, 3 variable(OBP, BA, SLG) been used because they have high correlation with team run scored but it may be have a high probability misjudgment due to too less variable. Classified the player to two category only also might let the result been misjudgment.

References

Benjamin Baumer and Andrew Zimbalist, 2013, The Sabermetrics Revolution, *University of pennsylvania press.*

Gabriel B.Costa, Micheal R. Huber and John T.Saccoman, 2009, Understanding Sabermetrics, *McFarland*

Gelblum, Laucys and Saperstein and Sodha., 2010, The Game Theory of Baseball, *University library of berkeley.*

Hattie James., 2015, How Game Theory Is Applied To Pitch Optimization, *Fan Graphs Community Research.*

Issah Musah and Douglas Kwasi Boah and Baba Seidu., 2020, A Comprehensive Review of Solution Methods and Techniques for Solving Games in Game Theory, *Journal of Game Theory.*

Jackson, Matthew O., 2011. A Brief Introduction to the Basics of Game. Available at SSRN: https://ssrn.com/abstract=1968579 or http://dx.doi.org/10.2139/ssrn.1968579

Jim Albert., 2010, Sabermetrics: The Past, the Present, and the Future, *ResearchGate, 10.5948/UPO9781614442004.002*

Kevin Leyton-Brown and Yoav Shoham., 2008, Essentials of Game Theory: A Concise, Multidisciplinary Introduction, *Morgan & Claypool.*

Martin J.Osborne., 2000 , An Introduction to Game Theory, *oxford university press*

Matt Swartz., 2012, Game Theory and Baseball Part 1-5, *Fan Graphs Community Research.*

Micah Melling,. 2018, Game Theory Applications in Baseball, *Baseball Data Science*

Piper Slowinski., 2012, Plate Discipline (O-swing%, Z-swing%, etc) , *Fan Graphs Community Research.*