Internal Assessment: Probability Using Bayes' Theorem and Binomial Theorem

1

Table of Contents

Background Information:

Statistics in baseball is crucial when judging a player's condition, whether they are playing well or playing badly. Statistics are also used to predict the chances and probability of future outcomes. It is frequently used in the sports world such as basketball, football or baseball. Before certain important games, some sports networks will calculate the probability of each team winning in that game to give the fans the sense of what the game might be like and how it may end up. Because I see these statistics all over the internet about MLB players, I thought it would be interesting for me to calculate the probability of a certain player getting a hit on his next at bat given several variables. I am also a big Baseball fan so I thought it would be a fun but challenging investigation to see if I can accomplish the same as some of these analysts.

In order to do this, I have chosen to study the Bayes' Theorem to find the probability that a batter will get a hit, given different variables.

The Bayes' Theorem is a probability theory which is a way of understanding how the probability that a theory is true is affected by a new piece of evidence. Bayes' Theorem help: use a known outcome to predict the sequence of events leading up to the outcome.

Below is the Bayes' Theorem, which I will be using in this investigation to find the probability that a batter will get a hit, given different variables.

Bayes' Theorem

[1] $$P(T|E) = \frac{P(E|T) \times P(T)}{P(E|T) \times P(T) + P(E|T') \times P(T')}$$

This theorem is *conditional probability* as this finds the probability that one thing is true provided that another proposition is true. In the theorem above, it shows that Event T occurs, given that Event E occurs. P(T) shows the probability of Event T occurring, $P(T')$ is 1.00 minus the probability of Event T occurring. P(E|T) is the probability that Event E occurs given that Event T occurs. And $P(E|T')$ is 1.00 minus the probability that Event E occurs given that Event T occurs.

$$\therefore P(T') = 1 - P(T)$$

[1] An Introduction to Bayes' Theorem. (n.d.). Retrieved November 27, 2016, from http://www.trinity.edu/cbrown/bayesWeb/

3

Collecting Data:

To start collecting data, I need to choose what baseball player I will be collecting data from, and choose variables that I will be introducing to the probability to make the final probability more accurate. I have chosen Ichiro Suzuki as my MLB player because he is an iconic Japanese baseball player in the Major League. He is a batter for the Miami Marlins. The weakness of this calculation that can be identified from the beginning is that not all the variables can be introduced into the theorem to further solidify the probability. If I introduce every variable possible, the calculation will be endless. Thus, I have chosen three variables to calculate the probability. Those variables are, Ichiro plays at a "Home Game", plays against the Atlanta Braves, and plays in a game the team wins in. Through introducing these variables, I will hopefully make the probability that Ichiro will get a hit more accurate.

In order to clarify and avoid confusion, I will define some baseball terminology in advance that will be used in this investigation.

"Number of At Bats" means the number of times Ichiro went up to the batter's box until he has gotten out, or got on base.

"Hit" means whenever Ichiro has gotten on base from hitting the ball with the bat.

[2]**Data:**

First Variable – Home Game

Number of At Bats in the Season	Number of Hits in the Season	Number of At Bats in Home Games	Number of Hits in Home Games
319	95	145	40

Second Variable – Playing against the Atlanta Braves

Number of At Bats against Atlanta Braves in Home Games	Number of Hits against Atlanta Braves in Home Games
15	6

Third Variable – Playing in Wins

Number of At Bats against Atlanta Braves in Home Games which Marlins win	Number of Hits against Atlanta Braves in Home Games which Marlins win
6	1

[2] Ichiro Suzuki. (n.d.). Retrieved November 27, 2016, from http://www.espn.com/mlb/player/stats/_/id/4570/ichiro-suzuki

4

[3]**Calculations:**

First Variable- Home Game

The theorem will be set up like this:

$$P(H_1|B) = \frac{P(H_1)\times P(B|H_1)}{P(H_1)\times P(B|H_1) + P(H')\times P(B|H')}$$

We will look at the probability that Ichiro will get a hit based on past performances. In this theorem, $P(H_1|B)$ stands for the probability that Ichiro will get a hit given that he is playing in a home game. $P(H_1)$ will stand for the percentage of hits that Ichiro has gotten out of the amount of "at bats" he's had. $P(B|H_1)$ stands for the percentage of hits Ichiro has gotten given that he is playing in a home game. $P(H')$ stands for the percentage of when he did not get hits out of the "at bats" he had. Finally, the $P(B|H')$ stands for the percentage of when he did not get hits given that he is playing a home game.

First, we know that in the entire season, Ichiro Suzuki had 319 At Bats, and out of those At Bats, he's gotten a hit 95 times. This gives us a percentage value of 0.298 for $P(H_1)$. This means that the value of $P(H')$ will be 1.000-0.298, which is equal to 0.702. When these are inserted into the theorem,

$$P(H_1|B) = \frac{0.298\times P(B|H_1)}{0.298\times P(B|H_1) + 0.702\times P(B|H')}$$

Next, we will find the percentage of hits Ichiro gets when he is playing a home game. In home games, Ichiro has had 145 At Bats, and out of those At Bats, he has had 40 hits. This gives us a $P(B|H_1)$ of 0.276. Thus, the $P(B|H')$ value will be 1.000-0.276 which is 0.724. When these are inserted into the theorem and calculated,

$$P(H_1|B) = \frac{0.298\times 0.276}{0.298\times 0.276 + 0.702\times 0.724}$$

$$P(H_1|B) = \frac{0.082248}{0.082248 + 0.508248}$$

$$P(H_1|B) = \frac{0.082248}{0.590496}$$

$$P(H_1|B) = 0.139286 \approx 0.139$$

From this, we can see that the probability that Ichiro will get a hit in the next game given that he is playing in a home game is 0.139.

[3] Ichiro Suzuki. (n.d.). Retrieved November 27, 2016, from http://www.espn.com/mlb/player/stats/_/id/4570/ichiro-suzuki

5

[4]Second Variable- Playing Against Atlanta Braves

The theorem with a newly introduced variable will be set up like this:

$$P(B|R) = \frac{P(B)\times P(R|B)}{P(B)\times P(R|B) + P(B')\times P(R|B')}$$

We will be looking at the probability that Ichiro will get a hit given that he is playing a home game against the Atlanta Braves. $P(B)$ will stand for the probability that Ichiro will get a hit given that he is playing a Home game. $P(R|B)$ stands for the percentage of hits he has gotten against Atlanta Braves in a Home game. $P(B')$ stands for the probability that Ichiro does not get a hit given that he is playing a home game. $P(R|B')$ stands for the percentage of when Ichiro does not get a hit against Atlanta Braves in a Home game.

From the previous calculation, we know that the probability that Ichiro will get a hit given that he is playing a home game is 0.139 which is the $P(B)$ value. This means that the $P(B')$ value will be 1.000-0.139 which is 0.861. When we put these values in the theorem,

$$P(B|R) = \frac{0.139\times P(R|B)}{0.139\times P(R|B) + 0.861\times P(R|B')}$$

Next, we need the $P(R|B)$ value, which is the percentage of hits Ichiro gets when playing against Atlanta Braves in a home game. In Home games against Atlanta Braves, Ichiro has had 15 At bats, and he has had 6 hits. This gives a percentage of 0.400 thus the $P(R|B)$ value is 0.400 and the $P(R|B')$ value is 0.600. When these values are inserted into the theorem,

$$P(B|R) = \frac{0.139\times 0.400}{0.139\times 0.400 + 0.861\times 0.600}$$

$$P(B|R) = \frac{0.0556}{0.5722}$$

$$P(B|R) = 0.09716882 \approx 0.0972$$

We can see that the probability that Ichiro will get a hit in the next game given that he is playing a Home Game against the Atlanta Braves is 0.0972.

---

[4] Ichiro Suzuki. (n.d.). Retrieved November 27, 2016, from http://www.espn.com/mlb/player/stats/_/id/4570/ichiro-suzuki

6

[5]Third Variable – Playing in a Win

The final theorem with a new variable will look like this:

$$P(R|W) = \frac{P(R)\times P(W|R)}{P(R)\times P(W|R) + P(R')\times P(W|R')}$$

With the newly introduced variable, we will now be looking at the probability that Ichiro will get a hit given that he is playing in a Home game against the Atlanta Braves that the Marlins win in. P(R) Stands for the probability that Ichiro will get a hit given that he plays a Home game against the Atlanta Braves. Thus, the $P(R')$ stands for the probability that Ichiro will not get a hit given that he plays a Home game against the Atlanta Braves. P(W|R) stands for the percentage of hits Ichiro has gotten against Atlanta Braves in Home games that his team has won. So, $P(W|R')$ stands for the percentage of when Ichiro does not get a hit against Atlanta Braves in Home games that his team has won in.

We already know P(R), as it is the value we found in the previous calculation for the second variable. P(R) is 0.0972 and the value of $P(R')$ is therefore 1.0000-0.0972 which is 0.9028.

$$P(R|W) = \frac{0.0972\times P(W|R)}{0.0972\times P(W|R) + 0.9028\times P(W|R')}$$

Next we will find the percentage of hits Ichiro gets when he is playing a home game against the Atlanta Braves and wins. In home games against Atlanta Braves that the Marlins won, out of 6 At Bats, Ichiro has gotten 1 Hit. Thus the $P(W|R)$ is 0.167 and the $P(W|R')$ is 1.000-0.167 which is 0.833. When these values are inserted into the theorem,

$$P(R|W) = \frac{0.0972\times 0.167}{0.0972\times 0.167 + 0.9028\times 0.833}$$

$$P(R|W) = \frac{0.0162324}{0.7682648}$$

$$P(R|W) = 0.02112865 \approx 0.0211$$

We can see that the probability that Ichiro will get a hit in the next game given that it is a Home game against the Atlanta Braves that the Marlins win, is 0.0211.

Looking at this value that I have found for the probability, I think that it is quite low. I feel as if a 2.11% chance of getting a hit in the next game given the different variables is significantly low however, the mathematics and the use of the Bayes' Theorem is not incorrect thus this should be an accurate measure.

[5] Ichiro Suzuki. (n.d.). Retrieved November 27, 2016, from http://www.espn.com/mlb/player/stats/_/id/4570/ichiro-suzuki

As this is conditional probabilities, the different variables or conditions being introduced should be making the probability more precise and accurate.

[6]Binomial Theorem:

To further extend this, we could look at the overall performance of Ichiro throughout the season using Binomial Theorem. The Binomial theorem is,

$$Probability = \binom{n}{k} p^k q^{n-k}$$

In this equation, "n" stands for the number of attempts and "k" stands for the number of success. "p" stands for the probability of success and "q" stands for the probability of failure, or $q = 1 - p$.

Using this binomial theorem, we will be able to figure out the probability that Ichiro will get a certain amount of hits in a certain amount of At bats in one season. With that, we can compare the probability with another elite player in the MLB by the name of Mookie Betts who is an athlete in my favorite team, who also happened to be the runner-up in the "MVP" or "Most Valuable Player" recognition race.

We will first set the number of At Bats (n) to 400. In the MLB, the ideal batting average is 0.300 which means that the batter hits 30% of the time he is at bat. 30% of 400 At Bats is 120, thus we will set the number of hits (k) as 120 for both Ichiro and Mookie Betts, to evenly explore each probability. "p" in the Binomial Theorem will be the batting average, which is the probability of the batter getting a hit, and "q" will be the probability that the batter will not get a hit ($q = 1 - p$).

Probability that Ichiro will get 120 Hits:

Ichiro's Batting Average (probability of getting a hit or "p"): P(Success) = 0.291

Ichiro's Probability of not getting a hit ("q"): P(Failure) = 0.709

$$P(Ichiro\ getting\ 120\ Hits) = \binom{n}{k} p^k q^{n-k}$$

$$P(Ichiro\ getting\ 120\ Hits) = \binom{400}{120} (0.291)^{120} (0.709)^{400-120}$$

$$\boldsymbol{P(Ichiro\ gets\ 120\ hits) = 0.0402}$$

[6] Mathwords: Binomial Probability Formula. (2016, February 21). Retrieved November 28, 2016, from http://www.mathwords.com/b/binomial_probability_formula.htm

8

[7]Probability that Mookie Betts will get 120 Hits:

Mookie Betts' Batting Average ("p"): P(Success) = 0.318

Mookie Betts' Probability of not getting a hit ("q"): P(Failure) = 0.682

$$P(Mookie\ Betts\ getting\ 120\ Hits) = \binom{n}{k} p^k q^{n-k}$$

$$P(Mookie\ Betts\ getting\ 120\ Hits) = \binom{400}{120} (0.318)^{120} (0.682)^{400-120}$$

$$\boldsymbol{P(Mookie\ Betts\ gets\ 120\ hits) = 0.0322}$$

Conclusion:

My aim in this investigation was to use the Bayes' Theorem to find the probability that a Baseball player would get a hit in his next At Bat. I chose Ichiro Suzuki to be the subject of this, who is a baseball player in the MLB playing for the Miami Marlins. I chose three variables to look at to incorporate into the theorem to find the conditional probability of Ichiro getting a hit in the next game. These three variables were that Ichiro plays in a Home game, he plays against the rivals of the Miami Marlins, the Atlanta Braves, and that he plays in a game in which the Marlins win. Thus through introducing each condition or variable, it is supposed to make the probability more accurate.

After I collected the data needed for the calculations, I found the conditional probability of Ichiro getting a hit in the next game given that is it a Home game against the Atlanta Braves that the Marlins win in, is 0.0211. This value is actually relatively small as the hitting average of Ichiro is 0.298. The difference of the two is significantly large, however, I believe this final value found through the theorem is accurate because several variables are taken into account.

I found that the largest limitation to the Bayes' Theorem is that I am unable to test every variable possible in the world to find the most accurate conditional probability. As I previously stated in the beginning, if I were to include every variable in this theorem to find a more accurate probability, the investigation would be endless. There are some variables that I would like to test if I have the resources to collect such data. These variables would be, the weather of the day of the game or the time of day of the game etc. With these variables, I will be able to make my probability more accurate.

[7] MLB Player Batting Stats - 2016. (n.d.). Retrieved November 28, 2016, from http://www.espn.com/mlb/stats/batting/_/year/2016/seasontype/2

9

When we looked at the extension of Bayes' Theorem and used Binomial Theorem to find the probability of Ichiro getting 120 hits out of 400 At Bats, we found that the probability Ichiro will get 120 hits out of 400 At bats was 0.0402. Compared to that, another elite batter in my favorite team of the MLB, Mookie Betts, had a probability of 0.0322 of getting 120 hits out of 400 At bats. From this, we can see that Ichiro is also an elite batter as his probability of getting an ideal batting average is higher than those who are competing for a "Most Valuable Player" award.

Another interesting investigation related to this investigation that could be done is using statistics to find where the ball hit by the baseball player could land on the field. By using past data on where the ball has landed for a hit, an investigation could find where the next hit may land on the field. Once again, this uses statistics that I have used above and more resources to further push the investigation.

This investigation is important because it can be very useful in the baseball team in my school. I am a baseball player in my school and our managers, coaches and players all care about statistics. The Bayes' Theorem and the Binomial theorem can be very useful when planning a batting line-up. If we use the Bayes' Theorem and deduce that someone has a better probability of getting a hit in his next bat than another person, the coach will use the player with the higher probability in the batting line-up to heighten the chance of winning. In the real life, finding the probability of players getting hits using Bayes' Theorem and Binomial theorem will be very tactical and useful.

References/Bibliography:

1. Ichiro Suzuki. (n.d.). Retrieved November 27, 2016, from http://www.espn.com/mlb/player/stats/_/id/4570/ichiro-suzuki
2. An Introduction to Bayes' Theorem. (n.d.). Retrieved November 27, 2016, from http://www.trinity.edu/cbrown/bayesWeb/
3. MLB Player Batting Stats - 2016. (n.d.). Retrieved November 28, 2016, from http://www.espn.com/mlb/stats/batting/_/year/2016/seasontype/2
4. Mathwords: Binomial Probability Formula. (2016, February 21). Retrieved November 28, 2016, from http://www.mathwords.com/b/binomial_probability_formula.htm