Performance Streaks in Sports

Dr. Thomas Beatty and Christopher Bradway

Florida Gulf Coast University
Mathematics

Abstract

The purpose of our paper is to use the negative binomial experiment, and its ability to calculate probability in order to explore the phenomenon of streaks in sports. A binomial experiment is a series of Bernoulli trials. The Bernoulli trial is a method used to test the probability of an occurrence that has a "success or failure" outcome. We will apply our method to hitting streaks in Major League Baseball, where we will explore the level of consistency needed to accomplish one of the longest held records in sports; Joe DiMaggio's 56 consecutive game hitting streak which has stood for 78 years. DiMaggio's feat remains the pinnacle for exceptional consistency at the plate, however, using the negative binomial experiment we have discovered other players who have overcome much greater odds in achieving their own hitting streaks.

Introduction

In 1941 Yankee's center fielder Joe DiMaggio was batting with such consistency that he went on a 56 consecutive game hitting streak. Fans knew they were witnessing something special, but few probably thought they were witnessing a record setting performance that would not be eclipsed for more than 78 years. Many have chased the record but none have come close to surpassing DiMaggio's consistency, the next closest being Pete Rose's 44 consecutive games with a hit. When we decided to study performance streaks the age of this long standing record drew our attention. We decided to apply our model, the negative binomial experiment, to Dimaggio's record performance and to those who have attempted to chase it down.

Joe DiMaggio

Method

As previously stated, we have chosen to use the negative binomial experiment to calculate the probability of a streak occurring. While researching the topic we found that other methods had been tested, but by the admission of the authors, their methods were "not a good fit." We feel the nature of the negative binomial experiment will give us a much better result since the binomial experiment is made up of a series of Bernoulli trials. Bernoulli trials are used to test the probability of an event based on a "success or failure" outcome, meaning every attempt either produces a success or failure. This is easily applied to sports, since almost every event in sports is binomial in nature. With the consecutive game hitting streak we will have a success if the batter hits safely in at least one of his at bats per game.

Negative Binomial Distribution

P = constant probability of success (batting average)
λ = block length (avg. number of at bats per game)
n = number of block successes (games with at least one safe hit)
$(1-p)^{\lambda}$ = probability of failure
$1-(1-p)^{\lambda}$ = probability of success
$(1-(1-p)^{\lambda})^{n}$ = n consecutive block successes
Failure occurs at the $(n+1)^{\text{st}}$ block, follows a geometric distribution.
Our probability mass function (pmf) is $f(n;p,\lambda)$ where $f(n;p,\lambda)=(1-p)^{\lambda}(1-(1-p)^{\lambda})^{n}$
Our cumulative distribution function (cdf) is $F(n;p,\lambda)$, where $F(n;p,\lambda)=\Sigma_{k=0}^{n}f(k;p,\lambda)$ or after calculating the sum we get $F(n;p,\lambda)=1-(1-(1-p)^{\lambda})^{n}$. Now let $S(n;p,\lambda)$ be a streak of at least length n then we get $1-F(n-1;p,\lambda)=(1-(1-p)^{\lambda})^{n}$.

Model Testing and Results

For testing our model we will use $S(n;p,\lambda)$. We will then calculate our results using the player's batting average, at bats, and consecutive games with a hit.

Model Testing

First we create a 3-D plot showing how changes in p and λ will effect S.

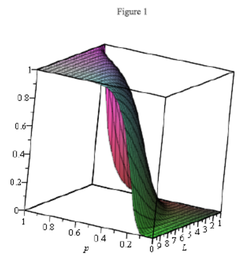
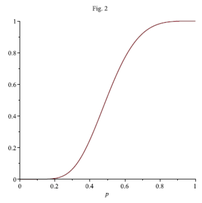


Figure 1: $S(10;p,\lambda)p=[0,1]and\lambda=[1,10]$



Results

We can tell by the graph of the cross section that S is very sensitive to change when p=[.25, .6]. Which means the likelihood of a ten game hitting streak goes up dramatically over that range. Applying our model to the statistics of the players with the longest consecutive game hitting streaks in baseball history, we come to the following results:

Player	Year	Streak Length (Games)	Batting Average	At Bats / Game	Probability (%)
Joe DiMaggio	1941	56	.408	3.98	0.0524
Pete Rose	1978	44	.302	4.12	0.000911
George Sisler	1922	41	.420	4.13	0.929
Ty Cobb	1911	40	.420	4.05	0.837
Paul Molitor	1987	39	.353	3.94	0.0359
Jimmy Rollins	2005/6	38	.283	4.32	0.00254
Tommy Holmes	1945	37	.352	4.13	0.0981
Ty Cobb	1917	35	.383	3.87	0.2396
George Sisler	1924/5	35	.325	4.27	0.0588
Luis Castillo	2002	35	.305	4.15	0.0125
Chase Utley	2006	35	.309	4.11	0.0137
Dan Uggla	2011	33	.233	3.73	0.000014

Figure 3: Prob. of Hitting Streaks

Conclusion

The choice of using the negative binomial experiment to calculate the probability of a performance streak has proven to be effective and accurate. Using the batting averages, at bats per game, and consecutive game hitting streaks of the players on our list, we can reorganize our table based on least probability of achieving each given hitting streak.

Hitting Streaks by Unlikelihood

Player	Year	Streak Length (Games)	Batting Average	At Bats / Game	Probability (%)
Dan Uggla	2011	33	.233	3.73	0.000014
Pete Rose	1978	44	.302	4.12	0.000911
Jimmy Rollins	2005/6	38	.283	4.32	0.00254
Luis Castillo	2002	35	.305	4.15	0.0125
Chase Utley	2006	35	.309	4.11	0.0137
Paul Molitor	1987	39	.353	3.94	0.0359
Joe DiMaggio	1941	56	.408	3.98	0.0524

Joe DiMaggio's consistency of getting a hit in 56 consecutive games is obviously an amazing feat. One which has not been challenged in 78 years. However, when we look at our probabilistic calculations we realize that Dan Uggla overcame much greater odds and the least probability of accomplishing his 33 game hitting streak.

Figure 4: Dan Uggla

References

- http://www.baseball-almanac.com
- Wearden, Andrew. 2016. https://www.saberballblog.com/2016/06/19/probability-why-hitting-streaks-are-impressive-and-why-theyre-not/
- Beltrami, Edward and Mendelsohn, Jay. https://sabr.org/research/more-thoughts-dimaggio-s-56-game-hitting-streak

Acknowledgements

Thank you Dr. Thomas Beatty for mentoring me through my project, and all you have done for me during my time at FGCU. Thank you Dr. Jaffar Ali Shahul Hameed for guiding me through the process. This research work is submitted as a partial fulfilment of the course MAT6930 Research Seminar.