Baseball and Statistics: Rethinking Slugging Percentage

Tanner Mortensen

December 5, 2013

Abstract

In baseball, slugging percentage represents power, speed, and the ability to generate hits. This statistic does not account for the relative skill of the opponent. We develop a new statistic that will more accurately rank players based on their opponents performance in addition to their own. We assign weights to the slugging percentage of batters and pitching effectiveness of pitchers and use linear algebra to determine these weights.

Introduction

Behind every hit, strikeout, run, win, loss and World Series victory lies statistics. These statistics track every player's performance through the highs and lows of their respective seasons and careers as a whole. Statistics are used for determining awards in batting and pitching based on recorded actions throughout the season. Several of these awards take into account the slugging percentage or pitching effectiveness of a certain player for the season in question. However, the effectiveness and current performance of the opposing players are not taken into account when awarding honors such as the Golden Glove, Rookie of the Year and Most Valuable Player. By assigning weights to the opposing batters and pitchers through performance, a player's statistical value and rating can be determined using linear algebra. More specifically, the weighted slugging percentage and the weighted pitching effectiveness are found by finding a specific eigenvector of a matrix. Using these weighted statistics, comparisons can be drawn between current leaders of slugging percentage against leaders of weighted slugging percentage of the same season.

Background

Major League Baseball (MLB) consists of 30 teams, each with a set roster of batters and pitchers for that particular season. Each player, through plays during the season, has numbers and statistics attached to them; plate appearances, at-bats, number of times on-base and hits. In particular, hits can be subdivided into several different types based on the number of offensive bases gained. Unlike batting average, the slugging percentage takes into account the number of bases gained with each hit. Every single hit by the batter is counted as 1, every double as 2, every triple as 3 and every home-run as 4. The slugging percentage sp of batter i can be represented as

$$sp_i = \frac{tb_i}{ab_i} = \frac{(1 \times S) + (2 \times D) + (3 \times T) + (4 \times H)}{ab_i}$$

where tb_i represents the total bases of player i, ab_i represents the at-bats of player i, and where S, D, T, and H represent the total number of singles, doubles, triples and home-runs hit.

An at-bat is defined as "an official turn at batting charged to a baseball player except when the player walks, sacrifices, is hit by a pitched ball, or is interfered with by the catcher" [1]. This differs from another term, plate appearance, which can be defined as "a statistic in baseball that is earned when a player completes a turn at batting with a hit, walk, out or reaching base on an error. Plate appearances do not occur when the batters time is interrupted before completion by way of events such as an existing runner being caught stealing or being picked off to end the inning or if the batter is replaced by a pinch hitter"[2].

For example, if a batter strikes out, it would be marked as a plate-appearance and an at-bat. However, if a player was to get hit by a pitch and advance to first base, it would be marked only as a plate-appearance and not an at-bat.

Let N_b be the number of batters in a given league and let N_p be the number of pitchers in the same league. Therefore, the slugging percentage of batter i for an entire season can be written

$$sp_i = \frac{1}{ab_i} \sum_{j=1}^{N_p} tb_{i,j}$$

where $tb_{i,j}$ represents the total bases of i against j. This defines the slugging percentage for batter i against

all pitchers j that i faces throughout the season.

Pitchers use another statistic to keep track of the slugging percentages of the batters they face. The opponent slugging percentage, *osp*, measures the how many total bases the pitcher has given up over the total number of at bats. For pitcher j, we have

$$ops_j = \frac{otb_j}{oab_j}$$

where otb_j stands for the total bases against pitcher j and oab_j stands for the number of at-bats against pitcher j.

In order to monitor how well the pitcher is performing, the statistic of pitching effectiveness is introduced. Pitching effectiveness (pe) of pitcher j can be defined as

$$pe_j = \frac{(oab_j - \frac{1}{4}(otb_{i,j}))}{oab_{i,j}}$$

To account for every batter faced during the season, pitching effectiveness can be written as a summation.

$$pe_j = \frac{1}{oab_{i,j}} \sum_{i=1}^{N_b} (oab_{i,j} - \frac{1}{4}(tb_{i,j}))$$

where $oab_{i,j}$ represents the number of times pitcher j has faced batter i and $otb_{i,j}$ represents the number of total bases pitcher j gave up to batter i.

Existing Metrics

Most baseball statistics do not account for the skill level of the opponent when calculating performance for a player. This follows suit with many other sports and competitive games. However, some methods do exist that account for opponents skill levels in several different statistical ways.

In the area of college football, Kenneth Massey, during his undergraduate studies at Bluefield College, developed Massey's Method. Massey's Method includes the mathematical theory of least squares and it's application to statistics. Massey's least squares method centers around the equation

$$r_i - r_j = y_k$$

where y_k is the margin of victory for game k and r_i, r_j are the ratings for teams i and j.

A.A. Markov created Markov Chains, which randomly determine processes and functions. The Markov rating method uses the concept of voting, in which a weaker opponent votes for a stronger opponent in a matchup. These votes are tallied and performance statistics are taken from the number and strength of the opponent votes. First used in decyphering poetry and works of literature, Markov Chains were then applied to NCAA basketball and March Madness.

The ranking and rating of players has even been expanded to sports and games you would not ordinarily expect. A prime example of this would be Elo's system of ranking and rating chess players. Arpad Elo, a physics professor and avid chess player, created a system where a players deviation from their previous performance is heavily taken into account when predicting current performance. Elo's system is represented as

$$r_{new} = r_{old} + K(S - \mu)$$

where r_{old} represents the player's older record, K is a constant set as 10 by Elo, S is the statistics based off the player's most recent performance and r_{new} represents the players new record.

Finally, Joe Scott, in his paper "Implicitly Defined Baseball Statistics", created weighted statistics for batting average and pitching effectiveness. These weighted statistics took into account the relative skill level of the opponent and how they impacted the overall ranking and ratings of players at the end of the season.[4]

Weighted Statistics

To account for the opposing pitcher or batter in each player's statistics, weights are assigned to the pre-existing formulas. The implementation of these weights will more accurately depict the performance of each batter and pitcher based on who they have faced during the season. The weighted slugging percentage wsp for batter i can be expressed as

$$wsp_i = \frac{1}{ab_i} \sum_{j=1}^{N_p} wpe_j(tb_{i,j}) \tag{1}$$

wpe_j, or weighted pitching effectiveness for pitcher j can be defined as

$$wpe_j = \frac{1}{oab_j} \sum_{i=1}^{N_b} wsp_i(oab_{i,j} - \frac{1}{4}(otb_{i,j})) \tag{2}$$

We place the weights for the slugging percentages for batters 1 through N_b in a $N_b \times 1$ vector matrix of size called wsp. Likewise we place the weighted pitching effectiveness weights for pitchers 1 through N_p in a vector matrix of size $N_p \times 1$ called wpe.

$$wsp = \begin{bmatrix} wsp_1 \\ wsp_2 \\ wsp_3 \\ \vdots \\ wsp_i \end{bmatrix} \qquad wpe = \begin{bmatrix} wpe_1 \\ wpe_2 \\ wpe_3 \\ \vdots \\ wpe_i \end{bmatrix}$$

By combining these two vectors, the total weighted vector w can be represented as,

$$w = \begin{bmatrix} wsp \\ wpe \end{bmatrix} \tag{3}$$

Therefore, systems (1) and (2) can be expressed as the matrix equation,

$$w = \begin{bmatrix} O_b & M \cdot TB \\ N \cdot (AB - \frac{1}{4}TB^T) & O_p \end{bmatrix} \cdot w \tag{4}$$

where O_b is a $N_b \times N_b$ zero matrix, O_p is a $N_p \times N_p$ zero matrix, TB is a $N_b \times N_p$ matrix such that $(\mathbf{TB})_{i,j} = tb_{i,j}$, AB is a $N_b \times N_p$ matrix such that $(\mathbf{AB})_{i,j} = ab_{i,j}$,

$$M = \begin{bmatrix} \frac{1}{ab_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{ab_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{ab_{N_b}} \end{bmatrix},$$

$$N = \begin{bmatrix} \frac{1}{oab_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{oab_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{oab_{N_p}} \end{bmatrix}$$

Therefore, (4) can be expressed as the following linear system:

$$w = Cw \tag{5}$$

where

$$C = \begin{bmatrix} O_b & M \cdot TB \\ N \cdot (AB - \frac{1}{4}TB^T) & O_p \end{bmatrix} \tag{6}$$

Non-trivial solutions to system (5) are unlikely, as this would imply a $\lambda = 1$. We look for eigenvalues λ such that

$$\lambda w = \mathrm{C}w$$

with λ being a non-negative, real number.

In order to find a unique non-negative, real eigenvector that represents the weights for each player, we use the Perron-Frobenius Theorem:

Perron-Frobenius Theorem: *Let A be an irreducible non-negative $n \times n$ matrix. Then A has a real eigenvalue λ_1 with the following properties:*

1. $\lambda_1 > 0$
2. λ_1 *has a corresponding positive eigenvector.*

Matrix C, which represents the 2012 MLB season, is an irreducible non-negative 1407×1407 matrix. Therefore, C has a real, non-negative eigenvalue with a corresponding real, non-negative eigenvector.

Results

The following fictional league demonstrates the weighted and non-weighted slugging percentages and pitching effectivenesses. Consider a league which has three batters and two pitchers. In the chart below, the first number represents the number of total bases earned by a batter against a certain pitcher and the second number represents the number of at-bats that same batter had against that same pitcher.

Total Bases/At Bats	Pitcher 1	Pitcher 2
Batter A	0/9	2/5
Batter B	0/10	8/12
Batter C	5/2	0/11

Following the notation in the weighted statistics section, we obtain the following matrices:

$$AB = \begin{bmatrix} 9 & 5 \\ 10 & 12 \\ 2 & 11 \end{bmatrix}, TB = \begin{bmatrix} 0 & 2 \\ 0 & 8 \\ 5 & 0 \end{bmatrix}$$

$$M = \begin{bmatrix} \frac{1}{14} & 0 & 0 \\ 0 & \frac{1}{22} & 0 \\ 0 & 0 & \frac{1}{12} \end{bmatrix}, N = \begin{bmatrix} \frac{1}{21} & 0 \\ 0 & \frac{1}{28} \end{bmatrix}$$

Using these matrices, we construct C using (6),

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.1429 \\ 0 & 0 & 0 & 0 & 0.3636 \\ 0 & 0 & 0 & 0.3846 & 0 \\ 0.4286 & 0.4762 & 0.0357 & 0 & 0 \\ 0.1607 & 0.3571 & 0.3929 & 0 & 0 \end{bmatrix}$$

We find one non-negative eigenvalues of C that yields a unique eigenvector, giving the weights for each player, seen in the table below.

Name of Batter	Slugging Percentage(Ranking)	Weighted Slugging Percentage(Ranking)
Batter A	0.1429(3)	0.1637(3)
Batter B	0.3636(2)	0.4167(1)
Batter C	0.3846(1)	0.3823(2)

Name of Pitcher	Pitching Effectiveness(Ranking)	Weighted Pitching Effectiveness(Ranking)
Pitcher 1	0.9524(1)	0.5297(2)
Pitcher 2	0.9107(2)	0.6106(1)

For the 2012 MLB Season, the following five batters are the non-weighted slugging percentage leaders along side their weighted slugging percentages:

Name of Batter	Team	SP(Ranking)	Weighted SP(Ranking)
Giancarlo Stanton	Miami Marlins	0.608(1)	0.05622(13)
Miguel Cabrera	Detroit Tigers	0.606(2)	0.07803(2)
Ryan Braun	Milwaukee Brewers	0.595(3)	0.05455(19)
Josh Hamilton	Texas Rangers	0.595(4)	0.02355(155)
Mike Trout	Anaheim Angels	0.564(5)	0.0310307(111)

For the same season, the following five batters are the weighted slugging percentage leaders along with their non-weighted slugging percentages:

Name of Batter	Team	Weighted Slugging Percentage(Ranking)	SP(Ranking)
Jason Kipnis	Cleveland Indians	0.083501(1)	0.379(122)
Miguel Cabrera	Detroit Tigers	0.07803(2)	0.606(2)
Alex Gordon	Kansas City Royals	0.0704(3)	0.455(60)
Adam Dunn	Chicago White Sox	0.0657(4)	0.468(49)
Alejandro De Aza	Chicago White Sox	0.0650(5)	0.410(90)

For the 2012 MLB Season, the following five pitchers are the weighted pitching effectiveness leaders:

Name of Pitcher	Team	Weighted Pitching Effectiveness(Ranking)
Zach McAllister	Cleveland Indians	0.30897(1)
Jose Quintana	Chicago White Sox	0.25911(2)
Corey Kluber	Cleveland Indians	0.24189(3)
Hector Santiago	Chicago White Sox	0.20923(4)
Deunte Heath	Chicago White Sox	0.17203(5)

Future Research

By tracking the performance of an opponent and using weights, more complete results will be available. This system of weights and adjusted rankings may be applied to many other fields, both baseball related and non-baseball related. In the field of baseball statistics, this idea can be applied to assigning weights to teams as a whole rather than individual players. Weights may also be used in other statistics such as on-base percentage, runs batted in, and stolen bases to more accurately depict the value of each statistic based on the relative skill level of the opponent.

The implication of weights in statistics are not limited to the sport of baseball. Any head-to-head competition has the potential for implicitly defined weights. Given a set number of players or teams in which competitions are held where points are scored against one another, weights can take each player or teams record in addition to their past opponents records into account when determining their current ranking and rating in the given league. Much like the weighted baseball statistics, these weights have the potential to change trades between teams, order of play for members of the teams and awarding of honors at the end of the season, all due to taking the opponents skill into account when calculating performance of players.

References

[1] "At-Bats" Def. 1. Merriam Webster. n.d. Web. 10 Sept. 2013.

[2] "Plate Appearances" Def. 1. Sporting Charts. n.d. Web. 10 Sept. 2013.

[3] "Retrosheet." Retrosheet. N.p., n.d. Web. 12 Dec. 2013.

[4] Scott, Joe. Implicitly Defined Baseball Statistics. Thesis. Georgia College, 2012. N.p.: n.p., n.d. Print.

[5] "MLB.com: The Official Site of Major League Baseball." MLB.com: The Official Site of Major League Baseball. N.p., n.d. Web. 1 Dec. 2013.

[6] Langville, Amy N., and C. D. Meyer. Who's #1?: The Science of Rating and Ranking. Princeton [N.J.: Princeton UP, 2012. Print