Massey's Method

Introduction

Massey's Method, created by mathematics professor Kenneth Massey while an undergraduate student at Bluefield College in 1997, is one of the methods currently used by the Bowl Championship Series, a rating system that determines which teams in NCAA college football play in each of the bowl games. Also referred to as the Point Spread Method, this method is based on the theory of least squares and incorporates the number of games played by each team as well as the difference in points scored by teams that played against each other.

An Example

To better understand how this method works, we will begin with an example, which we will continue to refer to throughout our explanation. We consider five teams from the 2012 Division III college football season: Johns Hopkins, Franklin & Marshall, Gettysburg, Dickinson and McDaniel. Consider the following chart of these teams' wins and losses, where the score of the team in the row is listed first ("Win" and "Loss" refers to the outcome for the team in the row):

Teams	Johns Hopkins	F & M	Gettysburg	Dickinson	McDaniel
Johns Hopkins	-	Loss, 12 - 14	Win 49-35	Win 49-0	Win 49-7
F & M	Win, 14 - 12	-	Loss, 31-38	Win 36-28	Win 35-10
Gettsyburg	Loss 35-49	Win, 38-31	-	Loss 13-23	Win 35-3
Dickinson	Loss 0-49	Loss 28-36	Win 23-13	-	Win 38-31
McDaniel	Loss 7-49	Loss 10-35	Loss 3-35	Loss 31-38	-

We can see that each team played every other team once, for a total of four games played by every team. We can calculate the point differential between two teams that played a game against each other by subtracting one team's score from their opponent's score. For Massey's Method, we will want to calculate the total point differential for each team; that is, for each team, we want to sum the point differences of every game that team played. This will be an essential part of Massey's Method, which we will discuss in detail in the next section.

Let us start by calculating the total point differential for each team. For example, we begin with Gettysburg, who beat Franklin and Marshall and McDaniel, and lost to Johns Hopkins and Dickinson. We express this in the following system of equations:

Gettysburg - Johns Hopkins = 35 - 49 = -14

Gettysburg - Franklin and Marshall = 38 - 31 = 7

Gettysburg - Dickinson = 13- 23 = 10

Gettysburg - McDaniel = 35 - 3 = 32

If we add these individual point differences, we get the total point differential for Gettysburg for this series of games, and find it is 15. Doing these calculations for the other teams, we find a total point differential of 103 for Johns Hopkins, 32 for Franklin and Marshall, -40 for Dickinson, and -106 for McDaniel. We will use this later when we go back to our example. Let us first explain the method.

The Math

Massey's Method is based on the least squares solution to a system of equations. An important equation to this idea is the following: r_i - r_j = y, where r_i and r_j are the ratings of teams i and j and y is the point difference in the game played between teams i and j. Clearly, as we do not have ratings of the teams before using this method, this equation is not used exactly, but Massey employs a similar idea.

If for every game played between two teams i and j we calculate the point differential, we have the equation i - $j = s$, where i and j are the points earned by teams i and j respectfully, and s is the difference of the score. Thus, if a total of m games are played between n teams and we construct a similar equation for each game, the result is a system of equations with m equations and n unknowns. From this, we can construct an $m \times n$ matrix, (m rows because there are m equations, one for each game, and n columns, one for each team). We call this matrix X. In each row, for teams i and j that played each other, there is a 1 in the i space and a -1 in the j space, indicating that these teams played each other, with team i defeating team j, and zeros everywhere else. For our small example, we get the following matrix X:

$$X = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{array} \begin{array}{c} \begin{array}{ccccc} D & F & G & J & M \end{array} \\ \begin{pmatrix} 0 & 0 & 1 & 0 & -1 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & -1 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

The first row, for example, corresponds to the game played between Gettysburg and McDaniel. Gettysburg defeated McDaniel, thus the Gettysburg space has a 1 and the McDaniel space has a -1.

To find the rating of each team, then, we might consider the linear equation $Xr = y$, where X is our matrix, r is an $n \times 1$ vector of the ratings we are trying to find, and y is an

$m \times 1$ vector of the point differentials of every game. We write this as follows:

$$
\begin{array}{c}
\begin{array}{cc}
 & \begin{array}{ccccc} D & F & G & J & M \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{array} &
\begin{pmatrix}
0 & 0 & 1 & 0 & -1 \\
-1 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & -1 \\
-1 & 0 & 0 & 1 & 0 \\
0 & 0 & -1 & 1 & 0 \\
1 & 0 & 0 & 0 & -1 \\
1 & 0 & -1 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 \\
0 & 0 & 0 & 1 & -1 \\
0 & -1 & 1 & 0 & 0
\end{pmatrix}
\end{array}
\cdot
\begin{pmatrix} r_D \\ r_F \\ r_G \\ r_J \\ r_H \end{pmatrix}
=
\begin{pmatrix} 32 \\ 8 \\ 25 \\ 49 \\ 14 \\ 7 \\ 10 \\ 2 \\ 42 \\ 7 \end{pmatrix}
\end{array}
$$

Here, again, the game represented by the first row is that between Gettysburg and McDaniel, where Gettysburg defeated McDaniel. The number in the first row of the vector y is the point difference of this game. Gettysburg defeated McDaniel with a score of 35 to 3, thus the point differential is 32.

We find, however, that because there are more than likely more games played than there are teams, the matrix is overdetermined, meaning that there are more equations than there are unknowns in the system, and it almost certainly does not have a solution.

Because there is no solution, we multiply by X^T on the left on both sides and instead we consider the normal equations $X^T X r = X^T y$. We can solve for the vector r in this equation as a good estimate of the ratings. We call $X^T X$ a new matrix M and assign the variable p to $X^T y$, a $n \times 1$ vector of total point differentials, one for each team, as we calculated for our small example above. The matrix M becomes an $n \times n$ matrix with the total number of games played by each team along the diagonal; that is, for team i, entry M_{ii} is the total number of games played by team i. The entry M_{ij} for $i \neq j$ is equal to the negation of the number of games that teams i and j played against each other. We see that the rows and columns of this matrix sum to zero, so are linearly dependent:

$$\begin{array}{c c} & \begin{array}{ccccc} D & F & G & J & M \end{array} \\ \begin{array}{c} D \\ F \\ G \\ J \\ M \end{array} & \begin{pmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 \\ -1 & -1 & 4 & -1 & -1 \\ -1 & -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & -1 & 4 \end{pmatrix} \end{array}$$

We encounter one other small problem: after a few games have been played, this matrix has a rank $n-1$, which means that there are infinite solutions to this system $Mr = p$. We are looking for a single solution so we can obtain the rating of each team. To fix this problem, Massey changes the last row of the matrix to a row of all ones and the corresponding entry in the point differential vector to a zero, so that the rank is no longer less than n. This system can now be solved and we can find the ratings of each team. In our example, the new system is as follows:

$$\begin{array}{c c} & \begin{array}{ccccc} D & F & G & J & M \end{array} \\ \begin{array}{c} D \\ F \\ G \\ J \\ M \end{array} & \begin{pmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 \\ -1 & -1 & 4 & -1 & -1 \\ -1 & -1 & -1 & 4 & -1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \end{array} \cdot \begin{pmatrix} r_D \\ r_F \\ r_G \\ r_J \\ r_H \end{pmatrix} = \begin{pmatrix} -40 \\ 32 \\ 15 \\ 103 \\ 0 \end{pmatrix}$$

Advantages and Disadvantages

One advantage of this method is that M is a square matrix, unlike our original X, which makes it much easier to study. It is a straightforward method that only requires the use of the number of games played and the scores of each game played. This method does not require that every team plays the same number of games, as other methods do. It also automatically incorporates ties, as it does not utilize whether a team won or lost, but rather the point differential, so ties are already accounted for, unlike other methods that count teams' wins and losses. A disadvantage is that we need to alter the matrix M to be able to solve the system.

Possible Adjustments

As Massey already incorporates ties, we can adjust the method so that it does not account for ties. For this, Massey decides to ignore any game that ends in a tie and go about the method the same way as if these teams played one less game. Some methods ignore ties and need to be altered to account for them; we can adjust Massey's method if we want to ignore tie games. Another adjustment that can be made is to set a cap for the point differential. If one team beats another by a significant amount, the winning team's victory margin is very large, therefore perhaps causing their rating to increase drastically. One way to prevent this increase if they score much higher than their opponent is to cap the point differential. If the cap is set to a certain number, any game in which the winning team beats the opponent by a margin greater than the cap, it will only be counted as a win by whatever has been chosen for the cap.

Conclusion

Massey's method of ranking is based upon the mathematical idea of least squares and is a straightforward way to rank teams based upon the number of games played and the point differentials. It is advantageous as it is simple and easy to compute, but we must alter the original M matrix so the system has only one solution instead of infinitely many.

Additional Reading

For further reading on the Massey Method, see the links and references below:

Massey's website:

`http://www.masseyratings.com/`

Commentary from Massey about rating and ranking basketball teams:

`http://blog.press.princeton.edu/2012/02/29/march-mathness-the-massey-method/`

Langville, Amy N. , and Carl D. Meyer. *Who's # 1?: The Science of Rating and Ranking* Princeton: Princeton University Press, 2012. Print.