Team Ranking

Justin Wyss-Gallifent

July 21, 2021

5.1 Introduction

Suppose there are three sports teams called T1, T2 and T3 and they play a number of games against one another with point outcomes for each game. We wish to assign some sort of numerical rank to each team so that we can decide who is best, second best, and third best, and in some sense by how much.

Just to see how this might be complicated, consider: If T1 beats T2, T2 beats T3 and T1 beats T3 then it's fairly clear how to rank the teams in terms of who is best. But what if T1 beats T2, T2 beats T3 and T3 beats T1, then who is best? Suppose T1 beats T2 by 5, T2 beats T3 by 3 and T3 beats T1 by just 1, now what? As we can tell, this isn't entirely clear at all.

The method we present is Massey's Method, developed by Kenneth Massey while he was an undergraduate.

5.2 Method

5.2.1 Building a System of Equations

What could we mean to assign each team a ranking?

One idea is that for two teams i and j with numerical rankings r_i and r_j it seems reasonable that if they play a game in which team i beats team j by p points then we could have $r_i - r_j = p$. For example if team i beats team j by 3 points then it seems reasonable that $r_i - r_j = 3$ because team i is 3 points better than team j.

For losses and ties this works just fine. For example if team i and team j tie then we can write either $r_i = r_j$ or $r_i - r_j = 0$ and if team j beats team i by 5 then we can write either $r_j - r_i = 5$ or $r_i - r_j = -5$.

By way of a juicy example let's suppose that there are four teams T1, T2, T3, T4. During a particular season we have the following where a win by a negative number really indicates a loss:

- T1 plays T2 and wins by 2.
- T1 plays T2 again and wins by 1.
- T1 plays T4 and wins by -2 (loses by 2).
- T2 plays T3 and wins by 1.
- T2 plays T4 and wins by 3.
- T3 plays T4 and wins by -3 (loses by 3).

If the rankings are r_1, r_2, r_3, r_4 then we ask that:

$$r_1 - r_2 = 2$$
$$r_1 - r_2 = 1$$
$$r_1 - r_4 = -2$$
$$r_2 - r_3 = 1$$
$$r_2 - r_4 = 3$$
$$r_3 - r_4 = -3$$

This is akin to solving the matrix equation

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ -2 \\ 1 \\ 3 \\ -3 \end{bmatrix}$$

This system has no solution.

5.2.2 Trying to Apply Least Squares

So what can we do? We can instead find the least-squares solution. What would this mean exactly? It would mean finding the team rank values such that if the teams had these rankings and played the games they did, that those game outcomes, as a vector, would be as close as possible to the actual game outcomes.

Onwards! If the above is $A\bar{r} = \bar{p}$ then we solve $A^T A\hat{r} = A^T \bar{p}$ instead:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ -1 & -1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \hat{r}$$

$$= \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ -1 & -1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -2 \\ 1 \\ 3 \\ -3 \end{bmatrix}$$

3

which gives us:

$$
\begin{bmatrix}
3 & -2 & 0 & -1 \\
-2 & 4 & -1 & -1 \\
0 & -1 & 2 & -1 \\
-1 & -1 & -1 & 3
\end{bmatrix}
\hat{r} =
\begin{bmatrix}
1 \\
1 \\
-4 \\
2
\end{bmatrix}
$$

5.2.3 Encountering the Problem

Before we proceed further let's call this new matrix equation $M\hat{r} = \bar{q}$ and make some notes about M. For simplicity suppose there are T teams and G games were played.

- The matrix M is $T \times T$.

- The entry m_{tt} equals the number of games that team t played. For example and $m_{33} = 2$ because T3 played a total of 2 games.

  This is because

  $$
  \begin{aligned}
  m_{tt} &= \left(A^T\right)_{t1} a_{1t} + \left(A^T\right)_{t2} a_{2t} + \left(A^T\right)_{t3} a_{3t} + ... + \left(A^T\right)_{tG} a_{Gt} \\
  &= a_{1t}^2 + a_{2t}^2 + a_{3t}^2 + ... + a_{Gt}^2
  \end{aligned}
  $$

  and for any $1 \le g \le G$ we have $a_{gt} = \pm 1$ if and only if, in game g, team t is involved, meaning that this sum equals the total number of games team t played.

- For $s \ne t$ the entry m_{st} equals negative of the number of games s and t played. For example $m_{21} = -2$ because T2 played T1 twice.

  This is because

  $$
  \begin{aligned}
  m_{st} &= \left(A^T\right)_{s1} a_{1t} + \left(A^T\right)_{s2} a_{2t} + \left(A^T\right)_{s3} a_{3t} + ... + \left(A^T\right)_{sG} a_{Gt} \\
  &= a_{1s}a_{1t} + a_{2s}a_{2t} + a_{3s}a_{3t} + ... + a_{Gs}a_{Gt}
  \end{aligned}
  $$

  and $a_{gs}a_{gt} = -1$ if and only if, in game g, both teams s and t are involved (one has a 1 in the row entry and the other has a -1), meaning that this sum equals the negative of the total number of games team s played against team t.

- The matrix M is symmetric.

- Any single row of M can be calculated from the other rows because any team's game history can be found from knowing every other team's game

4

histories. More explicitly each row is the negation of the sum of the other rows.

- Any two (or more) rows of M cannot be calculated from the remaining rows because it is impossible to know how those two teams performed without knowing something about at least one of them.

The last two points are very important, they say that the rows are linearly dependent and so the matrix is singular and therefore has infinitely many solutions (it definitely has solutions because it's a least-squares problem). Moreover they say that if we were to remove a single row that the remaining rows would be linearly independent.

Notice also that in the vector $\bar{q}$ any entry is the negative of the sum of the remaining entries since the total points won and lost among all teams is zero.

So in conclusion this new least-squares matrix equation has infinitely many solutions, and this is something we need to fix.

We could also have seen this back in the original matrix A in that each row sums to zero because each row has a 1 and a -1 in it. Consequently any column is simply the negation of the sum of the remaining columns. Therefore the columns are not linearly independent and so the least-squares method will yield multiple solutions.

5.2.4   Fixing the Problem

The way we'll solve this is to modify M and modify $\bar{q}$. We will replace the final row r_T (in this case r_4) with all 1s and the bottom entry of $\bar{q}$ with 0. This effectively adds the requirement that $r_1 + r_2 + r_3 + ... + r_T = 0$, which seems like a perfectly reasonable requirement in the sense that it only adds a requirement which doesn't get in the way of a reasonable team ranking. For the sake of simplicity we'll also call this new matrix M and new vector $\bar{q}$ and never look back.

Notice with this new row we see that for any of rows 1,2,3,...,$T-1$ that $row_i \cdot row_T = 0$ so that row T is perpendicular to each of the other rows and is therefore not a linear combination of them. Therefore the new matrix is invertible since the remaining rows we already knew to be linearly independent as noted in the bullet points above.

Thus we solve

$$
\begin{bmatrix}
3 & -2 & 0 & -1 \\
-2 & 4 & -1 & -1 \\
0 & -1 & 2 & -1 \\
1 & 1 & 1 & 1
\end{bmatrix}
\hat{r} =
\begin{bmatrix}
1 \\
1 \\
-4 \\
0
\end{bmatrix}
$$

and we find

$$\hat{r} = \begin{bmatrix} 19/26 \\ 9/26 \\ -41/26 \\ 1/2 \end{bmatrix} \approx \begin{bmatrix} 0.730769 \\ 0.346154 \\ -1.57692 \\ 0.5 \end{bmatrix}$$

So now we can assign numerical rankings to the teams as follows, in decreasing order:

Team 1 has ranking $r_1 = 0.730769$

Team 4 has ranking $r_4 = 0.5$

Team 2 has ranking $r_2 = 0.346154$

Team 3 has ranking $r_3 = -1.57692$

Note that, for example, T4 is significantly worse than the other three, all of which are rather close together.

Does this seem reasonable given their records?

5.2.5 Massey Method Summary

In summary the method is very direct:

I. Write down the system of equations corresponding to the games played and the points won or lost where the rankings are the unknowns.

II. Convert to a system of equations $A\bar{r} = \bar{p}$.

III. Write down the least-squares system $A^T A\hat{r} = A^T \bar{p}$.

IV. Change the lower row of the new matrix on the left to all 1s and the bottom entry of the new vector on the right to 0.

V. Solve.

5.2.6 Shortcut

In reality our observations earlier suggest that we don't need the intermediate step of filling in A at all.

Instead we can go straight to $M\hat{r} = \bar{q}$.

For M the final row is all 1s. Other than that row, m_{ii} equals the total number of games team i played and for $i \neq j$, m_{ij} equals the negative of the number of games i and j played.

For $\bar{q}$ the final entry is 0. Other than that entry, q_i equals the total number of points that team i earned.

Example 5.1. Suppose we have five teams who play a series of twelve games with the following outcomes:

- T1 plays T2 and wins by 3.
- T1 plays T3 and ties.
- T1 plays T4 and wins by -2.
- T2 plays T3 and wins by 1.
- T2 plays T3 and wins by 2.
- T2 plays T4 and wins by -3.
- T2 plays T5 and wins by -5.
- T3 plays T4 and ties.
- T3 plays T4 and wins by -1.
- T3 plays T5 and ties.
- T4 plays T5 and wins by 4.
- T4 plays T5 and wins by -3.

We can immediately fill in the following:

$$
\begin{bmatrix}
3 & -1 & -1 & -1 & 0 \\
-1 & 5 & -2 & -1 & -1 \\
-1 & -2 & 6 & -2 & -1 \\
-1 & -1 & -2 & 6 & -2 \\
1 & 1 & 1 & 1 & 1
\end{bmatrix}
\hat{r} =
\begin{bmatrix}
1 \\
-8 \\
-4 \\
7 \\
0
\end{bmatrix}
$$

which we solve to get:

$$
\hat{r} =
\begin{bmatrix}
-0.002899 \\
-1.438 \\
-0.6261 \\
1.055 \\
1.012
\end{bmatrix}
$$

In decreasing order we have rankings:

$$
\text{Team 4 has ranking } r_4 = 1.055
$$
$$
\text{Team 5 has ranking } r_5 = 1.012
$$
$$
\text{Team 1 has ranking } r_1 = -0.002899
$$
$$
\text{Team 3 has ranking } r_3 = -0.6261
$$
$$
\text{Team 2 has ranking } r_2 = -1.438
$$

5.3 Commentaries

5.3.1 Ranking is Relative But...

It's important to understand that the rankings are relative to one another. For example a ranking of 4/3 has no absolute numerical meaning until it is compared to another ranking at which point we can say which is better and we can give some sense of how much better.

However the working premise was that the difference between two teams' ranking $r_i - r_j$ ought to equal the number of points that Team i wins by in a game against Team j and so it's possible to use this difference to guess at how two teams might perform against one another.

Example 5.2. In the previous example T1 never played T5. However their relative rankings are -0.002899 and 1.1012 respectively. It follows that we can guess that if they did play that the score difference would be $-0.002899 - 1.1012 = -1.104099$ meaning T5 would win by 1.104099 points.

5.3.2 Ties

Massey's Method already takes into account ties since a tie between teams i and j can be entered as $r_i - r_j = 0$ or $r_j - r_i = 0$.

5.3.3 Multiple Games

Since least-squares is sensitive to repeated equations in curve fitting if two teams play more than once then Massey's Method will account for this simply by making sure both equations have been entered into the mix.

5.3.4 Weighting Games

It's possible to give a game more significance in two straightfoward ways. First, the point difference can be increased, so for example a win by 5 points could be recorded as a win by more than 5 points if we want to give that game more significance. Alternately the game could be recorded two or more times. The difference between these is explored in the exercises.

5.3.5 Disconnected Sets of Games

The Massey Method fails to return a result when the teams may be divided into two (or more) subsets none of whom play one another.

This is because the matrix $A^T A$ (and in fact even just the matrix A) is divided into two (or more) blocks. When we create M with the lower row all being 1

the lower row of any upper block is still a linear combination of rows from the same block.

We can't replace two (or more) entire rows by 1s because then they are linearly dependent. The solution would be to replace the appropriate parts of M corresponding to the lower row of each block which has the practical result of just applying the Massey Method to each subset of games.

This is explored further in the exercises.

5.4   Matlab

Nothing new here.

5.5   Exercises

Exercise 5.1. Suppose four teams play a number of games with the following results:

- Team 1 beats Team 2 by 7 points.
- Team 1 beats Team 3 by -8 points.
- Team 2 beats Team 3 by 0 points.
- Team 2 beats Team 3 by 15 points.
- Team 2 beats Team 4 by 5 points.
- Team 3 beats Team 4 by -1 points.

(a) Find the team rankings.

(b) Even though Team 1 did not play Team 4 if they were to play what result might you expect? Hint: The working assumption is that we wished $r_i - r_j$ equals the amount that team i wins by.

Exercise 5.2. Suppose four teams play a number of games with the following results:

- Team 1 beats Team 2 by 17 points.
- Team 1 beats Team 3 by 21 points.
- Team 2 beats Team 3 by 10 points.
- Team 2 beats Team 3 by 0 points.
- Team 2 beats Team 4 by -14 points.
- Team 3 beats Team 4 by 8 points.

(a) Find the team rankings.

(b) Even though Team 2 did not play Team 4 if they were to play what result might you expect?

Exercise 5.3. If the Massey matrix equation for a set of teams is:

$$\begin{bmatrix} 3 & -1 & -2 & 0 \\ -1 & 4 & -2 & -1 \\ -2 & -2 & 4 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \hat{r} = \begin{bmatrix} 6 \\ 2 \\ 10 \\ 0 \end{bmatrix}$$

(a) How many teams are there?

(b) How many games did Team 4 play?

(c) How many games did Team 1 play and against whom?

(d) How many total games were played?

Exercise 5.4. Before replacing the final row of the Massey matrix by all 1s, the sum of the diagonal entries be even. Why?

Exercise 5.5. Suppose four teams play a number of games with the following results:

- Team 1 beats Team 2 by 5 points. Team 1 at home.

- Team 1 beats Team 3 by 10 points. Team 3 at home.

- Team 2 beats Team 3 by -10 points. Team 2 at home.

- Team 2 beats Team 3 by 7 points. Team 3 at home.

- Team 2 beats Team 4 by -14 points. Team 2 at home.

- Team 3 beats Team 4 by 21 points. Team 3 at home.

(a) Find the team rankings, ignore the "at home" aspect.

(b) Suppose when a team plays at home it has a point advantage. Suppose history suggests that to factor this into the scores if a team wins at home the score should be lowered by 25% before calculating the ranking. Find the new team rankings; do they change?

Exercise 5.6. It seems reasonable that if T1 beats T2 by k points, T2 beats T3 by k points and T3 beats T1 by k points that all three teams should have the same ranking. Show that this is the case.

Exercise 5.7. Suppose two teams play two games. In the first game Team 1 beats Team 2 by α points. In the second game Team 2 beats Team 1 by β points. Both α and β could be zero or negative. Find the rankings of the two teams.

Exercise 5.8. Suppose three teams play three games with the following results:

- Team 1 beats Team 2 by 17 points.
- Team 2 beats Team 3 by 22 points.
- Team 3 beats Team 1 by α points.

(a) Assuming α is unknown find the team rankings.

(b) What would α need to be so that Team 3 ranks higher than Team 1?

Exercise 5.9. Is it better for a team to beat another team twice by one point each time or once by two points? Provide evidence to support your assertion. Note: This question is given in an easier version in the next question; this question is more exploratory.

Exercise 5.10. Is it better for a team to beat another team twice by one point each time or once by two points?

(a) Find the rankings in the following scenario:

- T1 beats T2 by 3 point,
- T2 beats T3 by 2 points,
- T1 beats T3 by 2 points.

(b) Find the rankings in the following scenario:

- T1 beats T2 by 3 point,
- T2 beats T3 by 2 points,
- T1 beats T3 by 1 point.
- T1 beats T3 by 1 point.

(c) What is your conclusion?

Exercise 5.11. This problem explores the same idea as the previous problem.. It looks at the difference between winning 1 game by n points versus winning n games by 1 point.

(a) Suppose T1 beats T2 by n points, T2 beats T3 by 1 point and T3 beats T1 by 1 point. Find the team rankings $\bar{r}_n$ and find $\lim_{n \to \infty} \bar{r}_n$.

(b) Suppose T1 beats T2 by 1 point but does this n times, T2 beats T3 by 1 point and T3 beats T1 by 1 point. Find the team rankings $\bar{r}_n$ and find $\lim_{n \to \infty} \bar{r}_n$. This should simplify a lot!

(c) What do you notice about your two limits? Which seems better for T1?

Exercise 5.12. Suppose three teams T1,T2,T3 play games as follows:

- T1 beats T2 by α.

- T2 beats T3 by β.

- T3 beats T1 by γ.

Note that α, β, γ may be zero or negative.

(a) Show that if the three are to be ranked equally that it must be true that $\alpha = \beta = \gamma$.

(b) Is it possible for exactly two of the teams to be ranked the same if all of α, β, γ are different from one another? If not, explain why not. If so, give an example.

(c) Is it possible to choose α, β, γ so that the spread of rankings is arbitrarily large? Explain.

Exercise 5.13. Suppose six teams play games as follows:

- T1 beats T2 by 2.

- T2 beats T3 by 3.

- T4 beats T5 by 2.

- T5 beats T6 by 3.

(a) Show that the Massey method as given will fail to rank the teams.

(b) Rank the teams by applying the Massey method to each subset T1, T2, T3 and T4, T5, T6.

(c) Suppose later on T3 plays a game against T4 and wins by 1. Apply the Massey Method to rank all six teams together.

(d) Does the way the rankings change after the T3-T4 game make sense intuitively? Explain.

Exercise 5.14. The Massey method modifies $A^T A \bar{r} = A^T \bar{p}$ by replacing the lowest row of $A^T A \bar{r}$ by all 1s and by replacing the lowest entry of $A^T \bar{p}$ by 0. This is the same as insisting that $r_1 + ... + r_n = 0$. What would happen if we had set this equal to something else? Modify the first question and instead use

$r_1 + \ldots + r_n = 1$ and then try it with $r_1 + \ldots + r_n = 4$. What do you find in each case? What do you think the general pattern is?

Exercise 5.15. Suppose a collection of more than two teams play a series of games. If the games and/or scores between exactly two of the teams changes can this affect the ranking value of teams other than those two? Provide evidence to support your assertion.

Exercise 5.16. Give some nontrivial examples of situations where the team rankings produced by the Massey method do not require least-squares to solve; that is, there is an exact solution to the initial desired system of equations. Is it still necessary to insist that the sum of all of the rankings is zero?