The Colley matrix method for ranking

Ranking methods

Suppose that we have k items that we want to rank based on the outcomes of binary comparisons between pairs of items. Examples include sports competitions (where we want to rank teams or players based on the outcomes of games), wine tasting, and network routing.

In a sports competition involving k teams, it may not be feasible to have a balanced schedule where each teams plays the other $k-1$ teams an equal number of times. Therefore, in ranking the teams, we should take into account not only the number of wins by a given team but also the strength of its opponents. For example, if team i plays opponents $i_1, \cdots, i_m$ and has W_i wins then conceptually we can think of defining a strength measure for team i as

$$s_i = \phi\left(\frac{W_i}{m}, \frac{s_{i_1} + \cdots + s_{i_m}}{m}\right) \tag{1}$$

where $\phi(x, y)$ is an increasing function in both its arguments – thus for fixed $s_{i_1}, \cdots, s_{i_m}$, s_i increases as W_i increases while for fixed W_i, s_i increases as $(s_{i_1} + \cdots + s_{i_m})/m$ (the overall strength of its opponents) increases. What complicates the computation of strength ratings $s_1, \cdots, s_k$ using this formulation is the dependence of each team's strength rating on those of the other $k-1$ teams.

The Colley matrix method was developed by Wesley Colley[1] as a means of ranking American college football teams. The top division, Division I-A, consists of roughly 125 teams who each typically play a regular season schedule of between 11 and 13 games where some of these games will involve teams in lower divisions. The disparity in schedules and the fact that the best teams do not necessarily play each other means that some method of ranking teams is needed, for example, to select teams to compete in the College Football Playoff, which takes place each year in January. See `www.colleyrankings.com` for current Colley rankings for American college football and basketball.

Model and estimation

Colley's method for rating teams turns out to be essentially a method of moments procedure for a very simple model of the probability that team i defeats team j. In particular, assume that

$$P(\text{team } i \text{ beats team } j) = \frac{1}{2} + s_i - s_j \quad \text{for } 1 \le i \ne j \le k \tag{2}$$

where $s_1, \cdots, s_k$ are measures of strength for the k teams. Note that the model (2) remains unchanged if a constant is added to each s_i ($(s_i + c) - (s_j + c) = s_i - s_j$), and so to make

[1]Colley, W. N. (2002) Colley's bias free college football ranking method: The Colley matrix explained. Princeton University.

the model (2) identifiable, we impose the constraint

$$\frac{1}{k}\sum_{i=1}^{k} s_i = \frac{1}{2}; \tag{3}$$

this constraint allows us to interpret s_i as the probability that team i would beat an "average" team j with $s_j = 1/2$. (Other constraints are possible; for example, we could set $s_i = 0$ for some i, which would allow us to interpret s_j as the strength of team j relative to team i.) However, the model (2) is somewhat flawed since we require $|s_i - s_j| \le 1/2$ (for all $i \ne j$) for the model to make sense. An alternative is to define

$$P(\text{team } i \text{ beats team } j) = \psi(s_i - s_j)$$

where $\psi(x)$ is a non-decreasing function with $0 \le \psi(x) \le 1$ for all x and $\psi(0) = 1/2$; an example is the Bradley-Terry model

$$P(\text{team } i \text{ beats team } j) = \frac{\exp(s_i - s_j)}{1 + \exp(s_i - s_j)}.$$

The Bradley-Terry model has a wide variety of applications; for example, the Elo rating system for measuring the strength of chess players is based on the Bradley-Terry model.

To estimate $s_1, \cdots, s_k$ in (2), we start by defining

$$\begin{aligned} W_i &= \text{number of wins for team } i, \\ n_{ij} = n_{ji} &= \text{number of games between teams } i \text{ and } j, \text{ and} \\ n_{i\bullet} = \sum_{j \ne i} n_{ij} &= \text{total number of games played by team } i. \end{aligned}$$

From (2), we have

$$E(W_i) = \sum_{j \ne i} n_{ij}\left(\frac{1}{2} + s_i - s_j\right) \quad \text{for } i = 1, \cdots, k$$

and so method of moments estimators of $s_1, \cdots, s_k$ are defined by

$$W_i = \sum_{j \ne i} n_{ij}\left(\frac{1}{2} + \widehat{s}_i - \widehat{s}_j\right) \quad \text{for } i = 1, \cdots, k \tag{4}$$

subject to the constraint

$$\frac{1}{k}\sum_{i=1}^{k} \widehat{s}_i = \frac{1}{2}. \tag{5}$$

(Note that (4) can be reexpressed as

$$\widehat{s}_i = \frac{W_i}{n_{i\bullet}} + \frac{1}{n_{i\bullet}}\sum_{j \ne i} n_{ij}\widehat{s}_j - \frac{1}{2},$$

which is of the form (1) with $\phi(x, y) = x + y - 1/2$.) Thus $\widehat{s}_1, \cdots, \widehat{s}_k$ satisfying (4) and (5) are defined by the matrix equation

$$\begin{pmatrix} n_{1\bullet} & -n_{12} & -n_{13} & \cdots & -n_{1k} \\ -n_{21} & n_{2\bullet} & -n_{23} & \cdots & -n_{2k} \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ -n_{k1} & -n_{k2} & \cdots & -n_{k,k-1} & n_{k\bullet} \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \widehat{s}_1 \\ \widehat{s}_2 \\ \vdots \\ \widehat{s}_k \end{pmatrix} = \begin{pmatrix} W_1 - n_{1\bullet}/2 \\ W_2 - n_{2\bullet}/2 \\ \vdots \\ W_k - n_{k\bullet}/2 \\ k/2 \end{pmatrix} \qquad (6)$$

Note that the matrix on the left hand side of (6) has $k+1$ rows and k columns. Nonetheless, the equation (6) still has a unique solution — the $k \times k$ matrix formed by the first k rows is not invertible (due to the lack of identifiability of the model (2) without the parameter constraint (3)) and adding the additional row (which reflects the constraint (5) on $\widehat{s}_1, \cdots, \widehat{s}_k$) guarantees a unique solution.

In his formulation, Colley uses a slight variation of (6) to estimate $s_1, \cdots, s_k$ by adding a constants to the diagonal of the matrix formed by the first k rows as well as to the first k elements of the right hand side of (6). Specifically, he defines $\widehat{s}_1, \cdots, \widehat{s}_k$ by the matrix equation

$$\begin{pmatrix} n_{1\bullet}+2 & -n_{12} & -n_{13} & \cdots & -n_{1k} \\ -n_{21} & n_{2\bullet}+2 & -n_{23} & \cdots & -n_{2k} \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ -n_{k1} & -n_{k2} & \cdots & -n_{k,k-1} & n_{k\bullet}+2 \end{pmatrix} \begin{pmatrix} \widehat{s}_1 \\ \widehat{s}_2 \\ \vdots \\ \widehat{s}_k \end{pmatrix} = \begin{pmatrix} W_1 + 1 - n_{1\bullet}/2 \\ W_2 + 1 - n_{2\bullet}/2 \\ \vdots \\ W_k + 1 - n_{k\bullet}/2 \end{pmatrix}.$$

This regularization guarantees the condition (5) on $\{\widehat{s}_i\}$. The difference between the two sets of estimates is typically quite small.

Example: 2015 American League baseball season

As an example, we will use data from the 2015 season of the American League (AL) of Major League Baseball (MLB). The AL consists of 15 teams divided into 3 divisions of 5 teams each; each team plays 19 games within its own division and either 6 or 7 games against teams in the other two divisions, for a total of 142 games. In addition, each team plays 20 games against teams in the National League of MLB; for the purpose of our analysis, we will consider only games played between AL teams. (One could argue that this analysis gives a better indication of the true strength of the teams as there is considerable imbalance in the 20 interleague games played by each team – for example, the Houston Astros had a losing record against AL teams but made the playoffs on the strength of its interleague record.) Table 1[2] gives the number of wins for each AL team in games played against the other 14 AL teams.

[2] These data are available at `espn.go.com/mlb/standings/grid`.

The results from the Colley matrix are given in Table 2. The ranking of teams based on their Colley matrix estimates are essentially the same as those based on the teams' winning percentage against other AL teams. Note that the Colley estimates tend to be closer to 0.5 than the teams' raw winning percentages (against AL opponents) – this seems to be a feature of Colley estimates, which may be a consequence of the fact that the model (2) makes sense only if $|s_i - s_j| \le 1/2$ for all $i \ne j$.

We can also use the jackknife to estimate the standard errors (and more generally the variance-covariance matrix) of the estimates; use of the jackknife assumes that the outcomes of the games are independent, which may not be a reasonable assumption although the degree of dependence is likely to be relatively small. Define $\widehat{\mathbf{s}}_{-ij}$ to be the vector of estimates computed by deleting a single game where team i defeated team j. Then $\widehat{\mathbf{s}}_{-ij}$ is computed as in (6) where the matrix is modified by subtracting 1 from the i and j diagonal elements and adding 1 to the (i, j) and (j, i) elements; the vector on the right hand side is modified by subtracting 1/2 from the i element and adding 1/2 to the j element. If m_{ij} is the number of games where team i defeated team j then the jackknife variance-covariance estimate is given by

$$\widehat{\mathrm{Cov}}(\widehat{\mathbf{s}}) = \frac{n-1}{n} \sum_{i \ne j} m_{ij} (\widehat{\mathbf{s}}_{-ij} - \widehat{\mathbf{s}}_{\bullet})(\widehat{\mathbf{s}}_{-ij} - \widehat{\mathbf{s}}_{\bullet})^T$$

where

$$\begin{aligned} \widehat{\mathbf{s}}_{\bullet} &= \frac{1}{n} \sum_{i \ne j} m_{ij} \widehat{\mathbf{s}}_{-ij} \\ \text{and } n &= \sum_{i \ne j} m_{ij}. \end{aligned}$$

The estimated standard errors of $\widehat{s}_1, \cdots, \widehat{s}_n$ are simply the square roots of the diagonal elements of $\widehat{\mathrm{Cov}}(\widehat{\mathbf{s}})$.

vs	BAL	BOS	CHW	CLE	DET	HOU	KC	LAA	MIN	NYY	OAK	SEA	TB	TEX	TOR
BAL	–	11	3	5	4	3	3	2	0	10	6	3	10	1	8
BOS	8	–	3	2	4	2	4	2	2	8	5	4	9	2	10
CHW	3	4	–	10	9	5	7	4	6	2	5	4	1	3	4
CLE	1	4	9	–	7	5	9	4	7	5	3	4	5	3	3
DET	3	2	10	11	–	3	9	1	11	2	2	4	3	2	2
HOU	4	4	1	2	4	–	4	10	3	4	10	12	2	6	4
KC	4	3	12	10	10	2	–	6	12	2	5	4	6	3	3
LAA	4	5	3	2	6	9	1	–	5	2	11	12	3	12	2
MIN	7	5	13	12	8	3	7	2	–	1	4	4	4	3	2
NYY	9	11	5	2	5	3	4	4	5	–	3	5	12	2	6
OAK	1	1	2	4	4	9	1	8	3	4	–	6	3	10	1
SEA	3	3	3	3	3	7	2	7	3	1	13	–	4	12	4
TB	9	10	5	2	3	5	1	3	2	7	4	3	–	2	10
TEX	6	5	3	3	5	13	4	7	3	5	9	7	5	–	2
TOR	11	9	3	4	4	3	4	5	5	13	5	2	9	4	–

Table 1: Number of wins for each AL team versus the 14 other teams; for example, Baltimore (BAL) defeated Boston (BOS) 11 times while Boston defeated Baltimore 8 times.

Team	Divison	Colley $\widehat{s}_i$	AL Win %	Win %
KC	Central	0.572	0.577	0.586
TOR	East	0.564	0.570	0.574
LAA	West	0.533	0.542	0.525
TEX	West	0.532	0.542	0.543
NYY	East	0.532	0.535	0.537
MIN	Central	0.528	0.528	0.512
CLE	Central	0.494	0.489	0.503
BAL	East	0.489	0.486	0.500
HOU	West	0.489	0.493	0.531
CHW	Central	0.478	0.472	0.469
SEA	West	0.476	0.479	0.469
TB	East	0.470	0.465	0.494
DET	Central	0.469	0.461	0.460
BOS	East	0.465	0.458	0.481
OAK	West	0.408	0.401	0.420

Table 2: Estimated strength ratings and raw winning percentages (versus AL opponents and versus all opponents) for the 15 AL teams (the five playoff teams are indicated in bold); using the jackknife, the estimated standard error of each $\widehat{s}_i$ is approximately 0.040 while the estimated standard errors of $\widehat{s}_i - \widehat{s}_j$ (for $i \neq j$) is approximately 0.060 for teams in different divisions and 0.056 for teams in the same division.