Learning Goals

1. Rankings vs. Ratings
2. Calculating winning percentage and "Laplace" ratings and rankings.
3. Setting up the matrix equation for Colley's ratings for examples with a small number of teams.
4. Solving the matrix equation for Colley's ratings for examples with a small number of teams using R.

Colley's Method.

Chapter 3 from "Who's # 1" [1], chapter available on Sakai.

Colley's method of ranking, which was used in the BCS rankings prior to the change in the system, is a modification of the simplest method of ranking, the winning percentage. Colley's method gives us a rating for each team, which we can use to find a ranking for the teams.

Note A **ranking** refers to a rank-ordered list of teams and a **rating** gives us a list of numerical scores. Every rating gives us a ranking for the teams.

Winning Percentage Let t_i be the number of times Team i has played and let w_i be the number of times Team i has won, then using the winning percentage, the rating for Team i is given by

$$r_i = \frac{w_i}{t_i}.$$

Dealing with ties: Letting l_i denote the number of times Team i has lost, we get $t_i = w_i + l_i$ if we count a tie between the two teams as one game and a half of a win and a half of a loss for each. On the other hand, you may ignore ties completely when calculating t_i, w_i and l_i. The important thing in in the derivation of the Colley ratings is that $t_i = w_i + l_i$ for each team. In competitions where ties are not allowed, this is not an issue.

Some of the disadvantages of using the winning percentage to rate teams are:

- Ties in the ratings often occur.
- The strength of the opponent is not factored into the analysis.
- At the beginning of the season, the numbers do not make sense, since the ranking for each team is $\frac{0}{0}$.
- As the season progresses, a team with no wins has a rating of $\frac{0}{t_i} = 0$.

Laplace's Rule The main idea behind Colley's method starts with replacing the winning percentage by a slight modification of it called Laplace's rule of succession. The rating for Team i is then given by

$$r_i = \frac{1 + w_i}{2 + t_i}.$$

Colley uses an approximation to this rating to get a system of Linear equations from which he derives ratings r_i for the teams which incorporate strength of schedule.

Example Lets calculate the ratings and rankings given by the winning percentage and Laplace's at various stages of our running example: The 2015 Six Nations Championship.

Below we show the results for the first two rounds and our summary of the results of the first two rounds and the complete results of the tournament from previous lectures.

[1] Who's # 1, *Amy N. Langville & Carl. D. Meyer*, Princeton University Press, 2012.

2015 Six Nations Championship in Rugby

Round 1

Wales	16	vs.	England	21
Italy	3	vs.	Ireland	26
France	15	vs.	Scotland	8

Round 2

England	47	vs.	Italy	17
Ireland	18	vs.	France	11
Scotland	23	vs.	Wales	26

Recall that we summarized the results after round 2 and the final results in tables:

Feb 25	Ire.	Eng.	Wal.	Scot.	Fra.	It.
Ire.					18-11	26-3
Eng.			21-16			47-17
Wal.		16-21		26-23		
Scot.			23-26		8-15	
Fra.	11-18			15-8		
It.	3-26	17-47				

Full	Ire.	Eng.	Wal.	Scot.	Fra.	It.
Ire.		19-9	16-23	40-10	18-11	26-3
Eng.	9-19		21-16	25-13	55-35	47-17
Wal.	23-16	16-21		26-23	20-13	61-20
Scot.	10-40	13-25	23-26		8-15	19-22
Fra.	11-18	35-55	13-20	15-8		29-0
It.	3-26	17-47	20-61	22-19	0-29	

(a) As we mentioned before, fans like to make rankings for the teams part way through such tournaments in an effort to predict what the final results of the tournament will be. Let us see what rankings our statistics; the winning percentage and Laplace's rule give us after round 2 of this tournament.
Fill in the winning percentage and the ratings from Laplace's rule in the table below. Find the corresponding rankings.

After Round 2

i	Team i	$r_i = \frac{w_1}{t_i}$	Ranking Win. %	$r_i = \frac{1+w_i}{2+t_i}$.	Ranking Laplace
1	Ireland				
2	England				
3	Wales				
4	Scotland				
5	France				
6	Italy				

(b) Use the final results shown below to fill in the winning percentage and the ratings from Laplace's rule after the final round in the table below. Find the corresponding rankings.

Name	**W = Wins**	**L = Losses**	**W-L**	**W + L**
Ireland	4	1	3	5
England	4	1	3	5
Wales	4	1	3	5
Scotland	0	5	-5	5
France	2	3	-1	5
Italy	1	4	-3	5

After Final Round

i	Team i	$r_i = \frac{w_1}{t_i}$	Ranking Win. %	$r_i = \frac{1+w_i}{2+t_i}$.	Ranking Laplace
1	Ireland				
2	England				
3	Wales				
4	Scotland				
5	France				
6	Italy				

Note When using Laplace's rule, all teams/competitors start out with a rating of $1/2 = \frac{1+0}{2+0}$. The ratings move above and below 1/2 as the season progresses. Because one team's/competitor's win is another's loss, the r_i's are interdependent.

Notation, O_i: In the calculations below, we are going to denote the list of teams/opponents that team i has played so far by O_i. This list will vary depending on where we are at in the tournament. If team/opponent i has played team/opponent j twice, then team/opponent j should appear twice on the list.

For Example: In The Six Nations Tournament, the results of which are shown above, we see that Team 4 is Scotland. After round two O_4 gives a list of all teams that Scotland has played up to that point in the tournament

$$O_4 = \{\text{Team 3; Wales , Team 5; France}\}.$$

After the final round O_4 is a larger set since Scotland has played more teams:

$$O_4 = \{\text{Team 1; Ireland, Team 2; England, Team 3; Wales, Team 5; France, Team 6; Italy }\}.$$

We will continue to use the notation t_i for the number of games team i has played so far in the tournament. At any point in the tournament t_i will equal the number of elements on the list O_i.

For Example: After round two in the Six Nations Tournament above, $t_4 = 2$ and after the final round , $t_4 = 5$.

An approximation for $\frac{t_i}{2}$: Colley makes an approximation shown in our calculations below which uses the fact that the Laplace ratings for each team fluctuate around 1/2 and that they are interdependent, so one might expect them to average out to 1/2 for a large number of teams in the tournament. **The approximation says that if I sum the Laplace ratings over all of the teams played by team i at some given points in the tournament, the that sum should be roughly equal to one half of the number of games played by team i at that point in the tournament,** specifically:

$$\boxed{\frac{t_i}{2} = \sum_{k=1}^{t_i} \frac{1}{2} \approx \sum_{k \in O_i} r_k}$$

where r_k denotes the current Laplace rating for team k.

For Example: In our example above After round two

$$\sum_{k \in O_4} r_k = r_3 + r_5 = 1/2 + 1/2.$$

On the other hand $\frac{t_i}{2} = \frac{2}{2} = 1$ so Colley's approximation is exactly equal to $\frac{t_i}{2}$ here.

In our example above After the final round

$$\sum_{k \in O_4} r_k = r_1 + r_2 + r_3 + r_5 + r_6 = \frac{5}{7} + \frac{5}{7} + \frac{5}{7} + \frac{3}{7} + \frac{2}{7} = \frac{20}{7} \approx 2.86.$$

On the other hand $\frac{t_i}{2} = \frac{5}{2} = 2.5$, so the approximation is not exact in this case but is **a reasonable approximation.**

Colley Ratings Colley's ratings use the above approximation to the Laplace ratings to derive a system of linear equations. The solution to this system give Colley's ratings, which in turn give Colley's rankings. Let t_i denote the number of games Team i has played, let w_i denote the number of games that Team i has won (we consider a draw as half a win and half a loss) and l_i, the number of games they have lost. Let r_i denote the ratings we get using Laplace's rule. We have

$$\begin{aligned} w_i &= \frac{w_i - l_i}{2} + \frac{w_i + l_i}{2} \\ &= \frac{w_i - l_i}{2} + \frac{t_i}{2} \\ &= \frac{w_i - l_i}{2} + \sum_{k=1}^{t_i} \frac{1}{2} \end{aligned} \tag{0.1}$$

As noted above since all teams begin with $r_k = \frac{1}{2}$ and the ratings are distributed around this number as the season progresses, we have an approximation

$$\left(\frac{t_i}{2} =\right) \quad \sum_{k=1}^{t_i} \frac{1}{2} \approx \sum_{k \in O_i} r_k$$

where O_i denotes the set of teams that have played team i. Thus the ratings we get from Laplace's rule, $\{r_i\}$, approximately satisfy the system of n equations (n = the number of teams):

$$w_i = \frac{w_i - l_i}{2} + \sum_{k \in O_i} r_k. \tag{0.2}$$

(**NOTE:** There is one equation in this system for each team. After the final round of the Six Nations tour. above, we have six six equations. The equation(approximation) corresponding to Scotland (= Team 4) is given by $w_4 \approx \frac{w_4 - l_4}{2} + \sum_{k \in O_4} r_k.$. Plugging in the Laplace ratings, we get $0 \approx \frac{0-5}{2} + \frac{5}{7} + \frac{5}{7} + \frac{5}{7} + \frac{3}{7} + \frac{2}{7} = \frac{20}{7} = -2.5 + 2.86 = .36$ which is clearly just an approximation.)

By definition, the ratings from Laplace's rule satisfy

$$r_i = \frac{1 + w_i}{2 + t_i}. \tag{0.3}$$

or equivalently:

$$r_i(2 + t_i) - 1 = w_i. \tag{0.4}$$

Substituting the expression for w_i from Equation **??** into Equation **??**, we get a system of equations which are approximately satisfied by the ratings we get from Laplace's rule of the form:

$$(2+t_i)r_i - 1 = \frac{w_i - l_i}{2} + \sum_{k \in O_i} r_k$$

Adding 1 to both sides of each equation, we get:

$$(2+t_i)r_i = 1 + \frac{w_i - l_i}{2} + \sum_{k \in O_i} r_k$$

and subtracting $\sum_{k \in O_i} r_k$ from both sides of Equation i, we get

$$(2+t_i)r_i - \sum_{k \in O_i} r_k = 1 + \frac{w_i - l_i}{2}. \tag{0.5}$$

Example: After the final round of the six nations cup above, the equation corresponding to Scotland (Team 4) in this system is given by:

$$(2+t_4)r_4 - (r_1 + r_2 + r_3 + r_5 + r_6) = 1 + \frac{w_4 - l_4}{2}$$

which is

$$7r_4 - r_1 - r_2 - r_3 - r_5 - r_6 = 1 + \frac{0-5}{2} = -3/2$$

It can be shown that this system of equations has a unique solution and **Colley's ratings** are the ratings which actually satisfy these equations. if we have n teams in a conference, we can write this system of equations in Matrix form as

$$\begin{pmatrix} 2+t_i & -n_{12} & -n_{13} & \dots. & -n_{1n} \\ -n_{21} & 2+t_2 & -n_{23} & \dots. & -n_{2n} \\ \vdots & \vdots & \vdots & \dots. & \vdots \\ -n_{n1} & -n_{n2} & -n_{n3} & \dots. & 2+t_n \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

where $b_i = 1 + \frac{w_i - l_i}{2}$ and n_{ij} denoted the number of times Team i and Team j have faced each other.

In summary, we have :

Colley's Method

Colley's ratings are the solutions $\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$ of the linear system

$$C\mathbf{r} = \mathbf{b}$$

where
C is an $n \times n$ matrix called the Colley Matrix where

$$C_{ij} = \begin{cases} 2+t_i & i = j \\ -n_{ij} & i \neq j \end{cases}$$

t_i	total number of games played by team i
n_{ij}	number of times team i faced team j
$\mathbf{b}_{n\times 1}$	$n \times 1$ matrix on the right with $b_i = 1 + \frac{1}{2}(w_i - l_i)$
w_i	total number of wins accumulated by team i.
l_i	total number of losses accumulated by team i.
$\mathbf{r}_{n\times 1}$	general rating vector produced by the Colley system.
n	number of teams in the conference = order of C.

Example Write out the matrix equation $C\mathbf{r} = \mathbf{b}$ for the Colley method for the Six Nations Tournament example above after round 2 and solve the system of equations using R. Derive the corresponding Colley Rankings for the teams after round 2.

Example Let's look at some data from 2013 NCAA Mens Basketball, Division 1. Below we look at the data from the games played in the America East conference from Jan 02, 2013 to Jan 10 2013. This data can be found on the ESPN website. The teams in the conference are as follows:

i	Team i	Abbreviation
1	Stony Brook	STON
2	Vermont	UVM
3	Boston University	BU
4	Hartford	HART
5	Albany	ALBY
6	Maine	ME
7	Univ. Maryland, Bal. County	UMBC
8	New Hampshire	UNH
9	Binghampton	BING

The following is a record of their games and results (W/L) from Jan 02, 2013 to Jan 10, 2013:

Date	Teams	Winner
Jan 02, 2013	BING vs HART	HART
Jan 02, 2013	UVM vs UNH	UVM
Jan 02, 2013	BU vs ME	ME
Jan 02, 2013	ALBY vs UMBC	ALBY
Jan 05, 2013	STON vs UNH	STON
Jan 05, 2013	UVM vs ALBY	UVM
Jan 05, 2013	BU vs HART	HART
Jan 05, 2013	ME vs UMBC	ME
Jan 07, 2013	BING vs ALBY	ALBY
Jan 08, 2013	UVM vs BU	BU
Jan 09, 2013	BING vs STON	STON
Jan 09, 2013	ME vs HART	HART
Jan 09, 2013	UMBC vs UNH	UMBC

Example Write out the matrix equation $C\mathbf{r} = \mathbf{b}$ for the Colley method for this example on Jan 09. Solve for the ratings using R and convert to the Colley rankings.

i	Team i	Abbreviation	Colley Rank
1	Stony Brook	STON	
2	Vermont	UVM	
3	Boston University	BU	
4	Hartford	HART	
5	Albany	ALBY	
6	Maine	ME	
7	Univ. Maryland, Bal. County	UMBC	
8	New Hampshire	UNH	
9	Binghampton	BING	

Properties of Colley Ratings

- The Colley ratings are generated using win loss information only. Hence they are unaffected by teams that purposefully run up the score against weak opponents. If one wishes to take the point differential into account, one can use the Massey method (see Who's Number One).
- Each team starts with a rating of $\frac{1}{2}$ and as the season progresses, the ratings bounce back and forth above and below $\frac{1}{2}$. The average of all team ratings is $\frac{1}{2}$ (check the examples above). If one team increases its rating, the rating of another team must decrease to keep this balance.
- Because it uses only win loss information, the Colley method can be used in a wider variety of situations, in particular in non-sporting examples.
- A draw between two teams is counted as neither a win nor a loss, but in the above system, it is counted as a game. If we remove draws from the data, the Colley ratings may change, hence it is important to decide what to do with draws before ranking the teams.