A Markov Model for Ranking Cricket Teams Playing ODI Matches

Amit Kumar Yadav (190118), Ayush Nishad (190215), Nikhil Kumar Singh (190548)

INTRODUCTION

India, a country with more than a billion people and having diversity in language, food, culture, clothing etc but obsessed with one sport- Cricket. Cricket is a religion in this country and cricketers are God. This is the second most popular sport worldwide in fan following, coming only after football. With this motivation, we plan to develop a Markov Model for ranking cricket team and also try to predict the results of head-to-head matches between two teams.

The International Council for Cricket (ICC) is the global governing body for cricket. Currently, it has 106 members, and it governs and administrates the game and works with member countries to grow the sport. The concept of batting averages and bowling averages has been in cricket since the nineteenth century. These records were maintained by individuals and statisticians interested in the game. We will be looking at the ICC ODI rankings of the team. Note that the ODI has been started in the late twentieth century and the first game was played between Australia and England in 1971. India played its first ODI match in 1974 against England and has win percentage of around 55. The test player ranking was also there since the twentieth century, but ODI player rankings were introduced in the 1998. Although the official team rankings were introduced only in 2002 for 10 test teams and in 2005 the rankings were developed for top non-test nations from 11-30 to complement the top 10 test teams.

The current ICC Team Rankings method has been developed by David Kendix to rank international full membership teams playing across all three formats. In this method, the points table is adjusted every year on 1 May using the points earned in the past 36 months and points earned earlier are dropped. The ranking calculation uses 66.66 percent weightage for the points earned in the last 12 months and 33.33 percent weightage for points earned before the 1 may of last year (13-36 months). The points are given to teams after each match based on a formula. Then, the team's total points earned is divided by total number of matches played for the rating, calculation and teams are ranked on the basis of that rating. In this if the gap in the ratings

1. If the two team’s rating difference is less than 40 then points is awarded according to this table.

Match Result	Points earned
Win	Opponent’s rating + 50
Tie	Opponent’s rating
Lose	Opponent’s rating -50

2. If the two team’s rating difference is equal to 40 or more than that, then point is awarded according to the following table.

Match Result	Points Earned
Stronger team wins	Own rating + 10
Weaker team loses	Own rating -10
Stronger team ties	Own rating - 40
Weaker team ties	Own rating + 40
Stronger team loses	Own rating - 90
Weaker team wins	Own rating + 90

The points earned by matches played between 1 May and next update is added to the annual rankings to give current ICC rankings.

Generally, the team ratings for top 10 teams are in the range of 120- 90. The rating above 130 is rarely achieved and suggests a very strong performance. The rating system awards team with high consistency over a period of time and tries to reduce the impact of occasional rise or fall.

Currently, there are 20 countries in the ICC ODI ranking and they are New Zealand, England, Australia, India, South Africa, Pakistan, Bangladesh, Sri Lanka, West Indies, Afghanistan, Ireland, Scotland, Zimbabwe, Netherlands, UAE, Oman, Namibia, Nepal, US and Papua New Guinea.

In this paper, we plan to use the results of matches played in last 3 years to predict the team rankings. Also, the team rankings cannot always predict the winner of a head-to-head matches between two teams. We plan to develop a model for predicting the results of head-to-head matches and thus help teams in recognising tough opponents.

The use of Markov model has been prevalent in sports since 1976 when it was used on baseball results. Similarly in cricket, the researchers have explored Markov Chain for modelling various decision-making processes, finding optimal batting orders and calculating winning probabilities.

METHODOLOGY

A brief discussion on stochastic model.

We can model and evaluate stochastic systems that can be represented using one or a few random variables using basic probability theory. Sometimes many random variables are necessary to model a process and such random variables may evolve over time. So, using basic probability theory to model such process becomes impossible. We can model such system using stochastic processes. A stochastic process is the family of random variables {X_t: t belong to T}, for every t in set T X_t denotes the state of the system at time t.

In this paper we will use discrete time discrete state stochastic process called Markov chain. Suppose we are in state i at time point n then the probability of jumping to state j in $n+1^{th}$ time point is called one step transition probability denoted by P_{ij}. Now, according to Markov property the future state X_{n+1} only depend on the present state i.e., X_n and is conditionally independent of past states $X_0, X_1, X_2, \ldots, X_{n+1}$ in other words the knowledge about immediate past is sufficient to predict the future state. Mathematically,

$$P_{ij} = \mathrm{P}\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1} \ldots, X_1 = i_1, X_0 = i_0\}$$

$$= P\{X_{n+1} = j | X_n = i\}$$

Where, $P_{ij} \geq 0;\ i,j \geq 0;\ \sum_{j=0}^{\infty} P_{ij} = 1, i = 0,1,2, \ldots$

If a Markov chain have N possible states, then its one step transition probability can be written in the form of N x N matrix.

$$P = (P_{ij}) = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1N} \\ P_{21} & P_{22} & \cdots & P_{2N} \\ \vdots & & \ddots & \vdots \\ P_{N1} & P_{N2} & \cdots & P_{NN} \end{pmatrix}$$

Now we have defined one step transition probability, the n step transition probability denoted by $P_{ij}^n = P(X_n = j \mid X_0 = i)$ is to probability of going from state i to state j. P_{ij}^n is equal to the element in i^{th} row and j^{th} column of P^n matrix.

If a Markov chain is irreducible and have finite states then its stationary distribution always exists and is given by:

$$\lim_{n\to\infty} P^n = \pi = \pi P = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \\ \pi_1 & \pi_2 & \cdots & \pi_n \\ \vdots & & \ddots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix}$$

where $\pi = (\pi_1, \pi_2, \pi_3, \ldots., \pi_n);\ 0 < \pi_j < 1;\ \sum_{j=1}^{\infty} \pi_j = 1$

The π_j`s are the steady state probabilities. These can be viewed as the distribution of random variables in a long run.

Model

In this paper we tried to develop a Model using Markov chain that will rank ODI cricket team. We collected data of ODI matches from 2018 to 2020 from stats.espncricinfo.com website. We defined Markov chain as X_n denoting performance of team at n^{th} day. X_n can take only two values {W, N}.

State W: Team won the match.
State N: Team did not win the match

The transition probabilities were obtained using frequentist method. First for each team the number of transitions from each state to another were obtained then transition probabilities were obtained. For each team a sequence of 'W' and 'N' was constructed using data sorted in accordance of time. For example, if a team wins the losses, then losses again and finally win two time consequently then its sequence will look like 'WNNWW'. Now number of transitions from each state to each state is calculated. The result will look like.

$$\begin{array}{ccc} & W & N \\ W & n_{11} & n_{12} \\ N & n_{21} & n_{22} \end{array}$$

For example, for the above sequence the result will look like:

$$\begin{array}{ccc} & W & N \\ W & 1 & 1 \\ N & 1 & 1 \end{array}$$

By dividing each element by the corresponding row total, the transition probability matrix P was generated. The result will look like:

$$P = \begin{array}{c|cc} & W & N \\ W & p_{WW} = \dfrac{n_{11}}{n_{11}+n_{12}} & p_{WN} = \dfrac{n_{12}}{n_{11}+n_{12}} \\ N & p_{NW} = \dfrac{n_{12}}{n_{21}+n_{22}} & p_{NN} = \dfrac{n_{22}}{n_{21}+n_{22}} \end{array}$$

For example, for the above matrix the result will look like:

$$P = \begin{array}{ccc} & W & N \\ W & 0.5 & 0.5 \\ N & 0.5 & 0.5 \end{array}$$

p_{WW} denotes the probability of winning next game given the team has won todays game. Similarly, p_{WN} denotes the probability of losing next match given that the team has won todays match.

Then the stationary distribution was obtained, and then steady state probabilities were obtained. By solving $\pi = \pi P$ and $\sum_{j \in \{W,N\}} \pi_j = 1$. On the first glance it looks like there are three equations of only two variables namely $\{\pi_W, \pi_N\}$ but $\pi = \pi P$ provides us only one equation as P is a singular matrix.

Then teams are ranked on the basis of π_W i.e., probability of winning in long run.

RESULTS

Team	WW	WN	NW	NN	Steady state $\{\pi_W, \pi_N\}$	
England	0.690476	0.309524	0.866667	0.133333	0.736842	0.263158
India	0.666667	0.333333	0.73913	0.26087	0.689189	0.310811
New Zealand	0.677419	0.322581	0.529412	0.470588	0.621381	0.378619
South Africa	0.642857	0.357143	0.55	0.45	0.606299	0.393701
Pakistan	0.76	0.24	0.28	0.72	0.538462	0.461538
Bangladesh	0.56	0.44	0.5	0.5	0.531915	0.468085
Australia	0.65	0.35	0.285714	0.714286	0.449438	0.550562
Afghanistan	0.444444	0.555556	0.36	0.64	0.393204	0.606796
West Indies	0.391304	0.608696	0.368421	0.631579	0.377049	0.622951
Sri Lanka	0.421053	0.578947	0.292683	0.707317	0.335788	0.664212

The above table ranks team on the basis of their steady state winning probability. Some interesting points to observe are that a good team like England and India have very high chances of making comebacks, in fact it is more than continuing their winning streak. England team has a 0.87 probability of winning the next game after losing the previous one and Indian team has 0.74 probability of making comeback. The Pakistan has very high probability (0.76) of continuing their winning streak and very less probability (0.28) of making comebacks after losing a match.

The steady state winning probability in the above table tells us that the probability of winning a match in the long run irrespective of results of previous matches. The ranking based on this will be referred to as MC ODI ranking system in the paper further.

Now, Lets compare the ICC ODI ranking and our MC ODI ranking. As explained in the introduction the ICC ranking is based on points with different weightage given to matches played in different years. Our MC ranking system is based on steady state winning probability and can be interpreted as chance of winning a particular match against any team.

We can observe that our ranking system matches with the ICC ODI ranking system for the top 4 teams. Then the teams from 5-7 and 8-10 have slightly different order from our predicted MC rankings. We can say that our model is fairly successful in predicting the ICC ODI rankings.

Rank	**ICC ODI Ranking (2020)**	**Rating**	**MC ODI ranking**	**Steady state winning probability**
1	England	127	England	0.736842
2	India	119	India	0.689189
3	New Zealand	116	New Zealand	0.621381
4	South Africa	108	South Africa	0.606299
5	Australia	107	Pakistan	0.538462
6	Pakistan	102	Bangladesh	0.531915
7	Bangladesh	88	Australia	0.449438
8	Sri Lanka	85	Afghanistan	0.393204
9	West Indies	76	West Indies	0.377049
10	Afghanistan	55	Sri Lanka	0.335788

The above steady probability cannot predict the results of head-to-head matches between two teams accurately. For example, subcontinent team might have higher chance of beating subcontinental teams rather than SENA countries. For resolving this, we have calculated separate steady state probability for each team against all the opponents of that team.

The following table predicts the results of head-to-head matches of India. Here, we have only considered 6 teams and the rest were dropped since the number of matches played between India and these teams were less than or equal to 4. From this, we can conclude that Australia and New Zealand are the challenging teams for India and there are high chances of Indian team winning against Sri Lanka, South Africa, Pakistan and West Indies. New Zealand will be a tougher opponent for India can also be predicted from both the ranking systems. Interestingly, South Africa is just two positions below India and should be a tougher team but the head-to-head matches prediction says otherwise. Our head-to-head model can be validated from the future match results of India and South Africa.

index	**Against**	**Wining Probability**
1	**Sri Lanka**	**0.875**
2	**South Africa**	**0.8571428571428572**
3	**Pakistan**	**0.7499999999999999**
4	**West Indies**	**0.7142857142857143**
5	**Australia**	**0.6250000000000001**
6	**New Zealand**	**0.5454545454545455**

Similarly, the results for other teams can also be obtained and the results of future matches can be used to validate our model. There has been very less matches played in the last 2 years due to Covid and the data is insufficient for properly validating our model.

CONCLUSION

Our proposed Markov model for ranking ODI teams works fairly well and is mostly consistent with the ICC ODI ranking system. The interpretation of the calculated steady state probability is fairly straight and can be used for predicting results in the long run. There can be further modifications to take into account factors like home or away conditions, different match-ups etc. and make the model more accurate.

Code: *https://github.com/amit1729/IME625*

References:-

1. https://www.icc-cricket.com/about
2. https://en.wikipedia.org/wiki/ICC_Men%27s_T20I_Team_Rankings#Example
3. https://en.wikipedia.org/wiki/ICC_Men%27s_ODI_Team_Rankings
4. https://www.krick3r.com/2017/04/how-icc-rankings-annual-update-works.html
5. Anuthrika, T., Arivalzahan, S. and Tharshan, R., 2019. A Markov model for ranking cricket teams playing one day international matches. Vingnanam Journal of Science, 14(1), pp.1–6. DOI: http://doi.org/10.4038/vingnanam.v14i1.4143