A Markov model for ranking cricket teams playing one day international matches

T Anuthrika*, S Arivalzahan, R Tharshan

Department of Mathematics and Statistics, Faculty of Science, University of Jaffna, Jaffna, Sri Lanka.

*Email: anuthv@gmail.com

Abstract— Main purpose of ranking teams in One Day International (ODI) Cricket matches is to identify challenging teams of each team. In the sports science literature, various statistical procedures have been proposed by different researchers to rank teams. International Cricket Council (ICC) is using a ranking system as the standard practice. However, it does not work well to realize some head-to head match results. This paper provides a stochastic Markov chain model to rank teams. A comprehensive analysis on head-to-head match prediction for each cricket team is given based on steady-state probabilities. Performance of the proposed model is investigated with ICC ODI rankings. It is shown that the proposed model outperforms to predict results of head-to-head ODI Cricket matches.

Keywords— Stochastic Process, Markov Model, Steady-State Probability, One Day International Cricket ranking

I. INTRODUCTION

The One Day International (ODI) team ranking is an international ranking system authorized by International Cricket Council (ICC) to show the standings of different teams from various nations by analyzing past cricket match performances. The ICC uses a ranking system developed by David Kendix to rank the international full membership teams based on a point system in which, the ranking calculation mechanism gives 66.6% weights for current year matches and 33.3% weightage for matches played in last year played. Past matches have been removed. Currently, there are ten countries namely Sri Lanka, India, Pakistan, Australia, England, South Africa, West Indies, New Zealand, Bangladesh, and Zimbabwe having the full-membership in ICC. In this system, there is a moot point at the rapid incorporation of head-to-head prediction and leads to have difficulties to identify challenging teams for a team. Therefore, it is imperative to develop an acceptable model that works well on head-to-head match prediction.

In this context, our principal goal of this paper was to propose a stochastic model in order to predict head-to-head matches results that could explicitly depict challenging teams for a team. In which, team performance is quantified by total number of wins and losses against opponent teams. An application of the stochastic process in the sport's analysis is not new. Various research papers that predict and analyze game performances in different times using the Markov chain model, have been published. Belman (1976) has introduced a Markov Chain model to Baseball match results. Ursin (2014) have developed a Markov model for baseball with applications. Norman (1999) analyzed the possibilities to use stochastic processes for statistical modeling in sports sciences, especially, Clarke and Norman (1998) utilized the stochastic techniques in various decision-making processes in cricket. Lames (1988) applied the idea of assess the performance of individual players in team games such as tennis as well as Lames and Hohmann (1997) analyzed in volleyball.

Zhang (2003) applied this concept for the table-tennis results prediction. Bukiet *et al.* (1997) found optimal batting orders using Markov chain approach in Baseball. Further, Hirotsu and Wright (2003) evaluated the baseball game using the Markov chains. They have expressed how that approach might help to select optimal hitting strategies and how much the probability of winning increases if gained strategy is followed. In addition, winning probability in any state in the game was calculated by using the Markov model. Colwell *et al.* (1991) developed a Markov chain model to outcomes of the test matches between England and Australian teams. They used three states such as team win, lost, and draw. Frequency approach was used to obtain the transition probabilities, where, transition probabilities recorded in 3×3 matrix.

The remaining part of this paper is organized as follows. Section 2 discusses the comprehension of stochastic model, details of proposed stochastic model: how the model was built, how parameters were estimated and examined the model assumptions. In section 3, empirical results of the proposed model are discussed. In the last section, section 4, findings are summarized.

II. METHODOLOGY

A. Comprehension of stochastic model

A stochastic process is a family of random variables, $\{X_t, t \epsilon T\}$, where t usually denotes the time (Jones and Smith 2001). In which, at every time t in the set T, a random number X_t is observed, where the set T is called the index set of the process, might be countable set or an interval of the real line. The transition probability for a Markov chain that jumps from state i to j, P_{ij}; is defined as

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

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

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

and this type of stochastic process is known as a discrete time Markov Chain (Ross 2010). Every P_{ij} is the row vector for $j = 1, 2, ..., N$ and these transition probabilities satisfy $\sum_{j=1}^{N} P_{ij} = 1, \forall i$. Markov property, says that, at any time n, the future state X_{n+1} is conditionally independent of the past $X_0, X_1, ..., X_{n-1}$ and just depend on the present state X_n. In other words, the future behavior of the system depends only on the current state and does not depend on any of the previous states.

The one step transition probabilities of N possible states can be represented by $N \times N$ transition probability matrix P;

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

and the matrix $P = (P_{ij})$ is the transition matrix of the chain.

$P_{ij}^n = P(X_n = j | X_0 = i)$, is the probability of moving from state i to j in n steps. It can be clearlydescribed by Chapman-Kolmogorov equation as follows;

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m \forall n, m \geq 0 \text{, all } i, j$$

where $P_{ik}^n P_{kj}^m$ denotes the probability that process moves to state j in $n + m$ transitions starting from state i through a path which take it into state k at the nth transition. Then $P^{n+m} = P^n P^m$, this equation asserts that $P^n = P^{n-1}. P = P^n$ by induction for $n \geq 1$, where dot (.) denotes matrix multiplication.

A chain is to have a stationary (or steady state) distribution if there exists a vector such that given a transition probability matrix P :

$$\pi = \pi P$$

If a finite Markov chain is irreducible and ergodic then

$$\lim_{n \to \infty} P^n = \pi = \begin{pmatrix} \pi_1 & \pi_2 & ... & \pi_N \\ \pi_1 & \pi_2 & ... & \pi_N \\ ... & ... & ... & ... \\ \pi_1 & \pi_2 & ... & \pi_N \end{pmatrix}$$

where $\pi = (\pi_1, \pi_2, ..., \pi_N)$ with $0 < \pi_j < 1$ and $\sum_{j=1}^{N} \pi_j = 1$

This stationary probability vector can be viewed as the unique distribution of a random variable in a long-run.

Moreover, steady state probabilities π_j were obtained as
$$\pi_j = \lim_{n \to \infty} P_{ij}^n \forall i, j \in S.$$

B. Stochastic model building

The objective of this work was to develop a Markov model to rank ODI cricket team's that classify as win or not win in a match. By this classification, it was able to interpret the prediction through Markov chain model. However, the random walk indicates that past data cannot employed to predict the future behavior. Which means n^{th} day team performance is a random process and it is independent from past performances and just depends only immediate past. Hence it is a Markov chain model.

In data collection, secondary data of matches was gathered from 2005 to 2014 from ICC website. The counts of transitions from each state to other state are obtained and the transition probabilities were estimated using maximum likelihood criterion.

To build a Markov chain model, X_n is taken to be the nth day performance of a team and named as

State 1: Team win the match (*W*)

State 2: Team does not win the match (*N*)

So, X_n is a stochastic process that has a value from 1 to 2 on n^{th} day. Moreover, consider team performance in any sequence of letters chosen from the set $\{W, N\}$. Let n_1, n_2 denote number of letters '*W*', '*N*' respectively in the sequence. Let n_{11} denote the number of occurrences in the sequence that the letter '*W*' is immediately followed by the letter '*W*'. Let n_{12} denotes the number of occurrence in the sequence that the letter '*W*' is immediately followed by the letter '*N*'. And n_{21}, n_{22} are similarly defined.

Where, here there are two states and this sequence of results recorded in 2×2 matrix.

$$\begin{array}{cc} & \begin{array}{cc} W & N \end{array} \\ \begin{array}{c} W \\ N \end{array} & \begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{bmatrix} \end{array}$$

The transition probability matrix P was formed by dividing each element by the corresponding row total.

$$P = (\hat{p}_{ij}) = \begin{array}{c} W \\ N \end{array} \begin{bmatrix} \hat{p}_{11} = n_{11}/(n_{11} + n_{12}) & \hat{p}_{12} = n_{12}/(n_{11} + n_{12}) \\ \hat{p}_{21} = n_{21}/(n_{21} + n_{22}) & \hat{p}_{22} = n_{22}/(n_{21} + n_{22}) \end{bmatrix}$$

And also $\sum_{j=1}^{2} \hat{p}_{ij} = 1$

This expresses the fact that if the system is in one of the states at one observed value, it will with certainty be in one of the states at the next observed value. For example, *P(N, W)* means that probability of not win followed by win. This *(2×2)* matrix of transition probabilities is, called the first step transition probability matrix of the Markov chain. Moreover, each row of *P* is the probability distribution relating to a transition from state *i* to state *j*.

Further, it is checked the stationary (or steady state) distribution and obtained steady-state probabilities π_j as
$$\pi_j = \lim_{n \to \infty} P_{ij}^n \qquad \forall i, j \in S.$$

Teams were ranked based on winning probabilities in a long run. Thus, a new ODI Markov chain ranking system has been proposed. In model validation, ICC ODI ranking and MC ODI ranking were cross-validated.

III. RESULTS AND DISCUSSION

Table1: Transition probabilities and steady state probabilities

Team	WW	WN	NW	NN	Steady state probability (π)	
Australia	0.7143	0.2857	0.6216	0.3784	[0.6851	0.3149]
South Africa	0.6864	0.3136	0.6333	0.3667	[0.6689	0.3311]
India	0.5974	0.4026	0.5849	0.4151	[0.5923	0.4077]
Pakistan	0.6239	0.3761	0.4944	0.5056	[0.5680	0.4320]
Sri Lanka	0.5778	0.4222	0.5278	0.4722	[0.5556	0.4444]
New Zealand	0.5904	0.4096	0.4023	0.5977	[0.4955	0.5045]
England	0.5161	0.4839	0.4327	0.5673	[0.4721	0.5279]
Bangladesh	0.5890	0.4110	0.2778	0.7222	[0.4033	0.5967]
West Indies	0.5584	0.4416	0.2857	0.7143	[0.3929	0.6071]
Zimbabwe	0.4318	0.5682	0.2202	0.7798	[0.2793	0.7207]

ICC ODI matches results from January 2005 to March 2014 have been used in this study. The transition probabilities and steady state probabilities of ten test match playing nations are given in Table 1. If Australian team wins a match, then probability that it will win next match is 0.7143. However, if Australian team loses a match, then the probability that team will wins next match is 0.6216, which is less than the probability of win follows by another win (0.7143). Moreover, this behavior can be observed in all ten nations. That is, for any cricket team, chance of a win follows a win is higher than a win follows a loss.

Table 2: Steady state winning probabilities

Team	Steady state winning Probability
Australia	0.685
South Africa	0.668
India	0.592
Pakistan	0.568
Sri Lanka	0.555
New Zealand	0.495
England	0.472
Bangladesh	0. 403
West Indies	0.392
Zimbabwe	0.279

Table 2, shows the steady state winning probabilities of ten teams. Australian team's steady state winning probability is 0.6851, thus, in a long run the probability of winning a match for the Australian team is 0.6851, this is irrespective of results of previous matches. South Africa's steady state winning probability is 0.6689, which is next to Australia. India, Pakistan and Sri Lanka are the succeeding countries in that order. It could be noticed that a ranking among teams which is based on steady state probabilities. The proposed ranking scheme could be called as Markov Chain ODI ranking (MC ODI ranking). Table 3 compares both the existing ICC ODI rankings in March 2014 and the proposed MC ODI rankings.

The existing ICC ODI ranking scheme is based on points, but the proposed Markov Chain ODI ranking scheme is based on steady state probabilities. In prediction, though the point of the existing ranking scheme does not have a straight forward interpretation, steady state probability of the proposed MC ODI ranking scheme has a straight forward interpretation. For instance, as per the existing ranking scheme, Australia's points is 7579, but as per the proposed MC ODI ranking scheme

Australia's steady state winning probability is 0.6851. This means in a particular match, Australian team's winning chance is 68.51%.

Table 3: ICC and MC ODI Rankings

Rank	ICC ODI Ranking (March 2014)	MC ODI Ranking
01	Australia	Australia
02	Sri Lanka	South Africa
03	India	India
04	South Africa	Pakistan
05	England	Sri Lanka
06	Pakistan	New Zealand
07	New Zealand	England
08	West Indies	Bangladesh
09	Bangladesh	West Indies
10	Zimbabwe	Zimbabwe

As it was mentioned earlier, principal goal of this paper was to propose a stochastic model in order to predict head-to-head match result that could explicitly depict challenging teams for a team. The steady state probabilities Table 2 and points of the existing ICC ODI ranking both were calculated considering all the matches (irrespective of opponent team) of a team. However, there might be some interaction effects between teams. For instance, south Asian teams could do well against the spin bowling attack. A team which has good spin bowling attack could challenge other teams but may not be able to challenge south Asian teams. Thus, one team might not perform well against a team which is below than that team in ICC ODI or MC ODI ranking. This motivated to do further precise analysis on head-to-head match prediction. Therefore, a detailed further analysis was done for all teams.

Table 4: Transition probability and head-to-head steady state probability of Sri Lankan cricket team

Team	WW	WN	NW	NN	Steady state probability (π)	
India	0.3182	0.6818	0.3947	0.6053	[0.3667	0.6333]
Australia	0.3846	0.6154	0.4500	0.5500	[0.4224	0.5776]
Pakistan	0.6000	0.4000	0.3889	0.6111	[0.4930	0.5070]
South Africa	0.5000	0.5000	0.5000	0.5000	[0.5000	0.5000]
West Indies	0.4000	0.6000	0.7500	0.2500	[0.5556	0.4444]
England	0.6364	0.3636	0.5714	0.4286	[0.6111	0.3889]
New Zealand	0.7857	0.2143	0.5000	0.5000	[0.7000	0.3000]
Bangladesh	0.8095	0.1905	0.9000	0.1000	[0.8400	0.1600]
Zimbabwe	0.8889	0.1111	0.9000	0.1000	[0.9000	0.1000]
India	0.3182	0.6818	0.3947	0.6053	[0.3667	0.6333]

Separate transition probabilities and steady state probabilities were calculated for each team against all opponents of that particular team. As an illustration, consider results of comprehensive analysis of Sri Lankan team. In this respect, separate transition probabilities and steady state probabilities were calculated for Sri Lankan team against each opponent teams. Table 4 provides results of this detailed analysis. Table 5 shows separate steady state winning probabilities of the Sri Lankan team against each opponent teams. From Table 1 Sri Lankan team's common (irrespective of opponent team) steady state winning probability is 0.5556. However, in

head-to-head analysis, summarized results in Table 5 say that Sri Lankan team's steady state winning probability heavily depends on opponent team. The steady state probabilities vary with opponent team, ranging from 0.366 to 0.9. As per these results, steady state wining probability of the Sri Lankan team is less than 0.5 against India, Australia, and Pakistan and steady state probability is 0.5 against South Africa. For the rest of the teams Sri Lankan team's wining probability is greater than 0.5. Thus, India, Australia and Pakistan are the challenging teams for the Sri Lankan team.

Table 5: Head-to-head steady state winning probability for Sri Lankan team

Opponent Team	Steady state winning probability for Sri Lanka
India	0.366
Australia	0.422
Pakistan	0.493
South Africa	0.500
West Indies	0.556
England	0.611
New Zealand	0.700
Bangladesh	0.840
Zimbabwe	0.900

In the model validation in predictions, it could be used both existing ICC ODI ranking and head-to-head winning steady state probabilities of Table 5. Australia is above Sri Lanka in both rankings, that is, Australia is a challenging team to Sri Lankan team in both ranking methods. However, as per the ICC ODI ranking scheme India and Pakistan are below Sri Lankan team and they are not challenging teams to Sri Lanka. But, as per the proposed Markov chain based head-to-head prediction both India and Pakistan are challenging teams for Sri Lankan team. Therefore, future ODI match results of Sri Lanka versus India and Sri Lanka versus Pakistan could be used to validate the proposed head-to-head ranking method.

The ODI matched results from March 2014 to December 2017 has been used as a test bed to validate the proposed ranking method. There were 14 head-to-head matches between Sri Lanka and India during the above period. As per the ICC ODI ranking Sri Lanka is the favorite and as per the MC head-to-head prediction India is the favorite. Among those 14 matches India has won 12 matches and Sri Lanka has won only 2 matches. Thus, match result prediction based on ICC ODI ranking was correct two times and the proposed MC based prediction was correct on twelve times. Similarly 14 head-to-head matches played between Sri Lanka versus Pakistan was also considered as another test bed. As per the prediction based on ICC ODI ranking Sri Lankan team is the most favorite but as per the prediction based on MC, Pakistan is the most favorite.

Among those 14 matches, Sri Lanka has won only 3 and Pakistan won 11 matches. Here, again MC based head-to-head prediction worked better than ICC ODI based prediction in finding challenging teams.

Similarly, a detailed comprehensive further analysis was carried out for the rest of teams and MC based head-to-head prediction steady state probabilities were obtained. It could be noticed that except West Indies versus Bangladesh all other MC based head-to-head predictions exactly matched with ICC ODI ranking based prediction.

As per ICC ODI ranking West Indies is the favorite team for head-to-head prediction with Bangladesh team (see, ICC ODI ranking in Table 3). However, steady state winning probability for West Indies against Bangladesh is 0.4375, which is less than 0.5. Thus, as per the MC based head-to-head prediction, Bangladesh is the most favorite team. Here, again three ODI matches between West Indies and Bangladesh played between March 2014 and December 2017 have been used as the test bed. Bangladesh has won all three ODI matches. Thus, MC based head-to-head prediction worked better than ICC ODI based prediction. Table 6 summarizes the cross-validation significant results on our test beds.

IV. CONCLUSION

The ICC ODI ranking system is suspected to provide the reliable realization on some head-to-head match result for a person who is fascinated in Cricket. So, this paper proposed a new head-to-head prediction method so-called Markov chain based head-to-head prediction that can more effectively describe challenging teams for each team in ODI Cricket matches. To show the superiority of the proposed model with the ICC ODI ranking system, it was used real-data from 2005 to 2014 (March) for model fitting and 2014 (April) to 2017 for the model validation. In which, teams were ranked based on steady state probabilities of each team respective to opponent teams, and head-to-head proposed model working efficiency was investigated by the cross-validation with ICC ranking. The proposed model outperforms well than ICC to predict head-to-head matches in ODI cricket matches.

ICC ODI ranking scheme is based on points, but, in prediction, those points do not have a straight forward interpretation, however, steady state probabilities of the proposed MC ODI ranking scheme has a meaningful interpretation. It is believed that this proposed ranking system can be developed to predict the cricket players' performances against each team as well as teams in the international test Cricket matches.

Table 6: Significant results

Team	Opponent	ICC ODI ranking favorite	MC head-to-head prediction favorite	Results	Percentage of correct Prediction	
					In favor of ICC ODI ranking	In favor of MC head-to-head prediction
Sri Lanka	India	Sri Lanka	India	India won 12 out of 14 Matches	14.29%	85.71%
Sri Lanka	Pakistan	Sri Lanka	Pakistan	Pakistan won 11 out of 14 matches	21.43%	78.57%
West Indies	Bangladesh	West Indies	Bangladesh	Bangladesh won all 3 matches	0%	100%

V.   REFERENCES

Bellman, R. (1976). *Dynamic Programming and Markovian Decision Processes, with Application to Baseball*, Departments of Electrical Engineering, Mathematics and Medicine, UCLA. 12A(6), pp. 633-651

Bukiet, B. Harold, E and Palacios, J. A. (1997). *Markov chain approach to baseball Operations Research* 45(1), pp.14-23

Clarke, S. R. and Norman, J.M (1998). *When to rush behind in Australian Rules football: a Dynamic Programming approach*, Journal of Operational Research Society, 49(5), pp.530-536

Colwell, D. Jones, B. and Jillet, J. (1991). *A Markov Chain in Cricket. The Mathematical Gazette,* 75, pp.183-185.

Hirotsu, N. and Wright, M. (2003). *A Markov chain approach to optimal pinch-hitting strategies in a designated hitter rule baseball game*, Journal of the Operations Research Society of Japan, 46, pp.353-371.

Jones, P. and Smith, P.   (2001). *Stochastic processes an introduction*, Arnold texts in statistics, London: Arnold, 7, pp.259

Lames, M. (1988). *Sportspiele: animieren – trainieren,* Techniktraining im Tennis durch Computer simulation In R. Ahrensburg: Czwalina, pp.181-191.

Lames, M. and Hohmann, A. (1997). *Integrative Aspekte in Theorie und Praxis der Rückschlagspiele*. Zur Leistungsrelevanz von Spielhandlungen im Volleyball, pp. 121-128. In B. Hoffmann and P. Koch (Ed.) Hamburg: Czwalina.

Norman, J. M. (1999). *Markov process applications in sport*. In IFORS conference. Beijing, China, 50, pp. 536-545.

Ross, S. M. (2010). *Introduction to Probability Models,* 10[th]edn. Academic Press

Ursin, D. J. (2014). *A Markov model for baseball with applications*, Thesis and Dissertation. University of Wilconsin-Milwaukee, pp.964

Zhang, H. (2003). *LeistungsdiagnostikimTischtennis*, Doctoral dissertation. University of Potsdam.