Creating Markov models in a game of tennis using Maple

VEDANG JOSHI

qk18050@bristol.ac.uk

January 31, 2019

Abstract

The Markov model proposed in the study showed a game duration in tennis of 3.27 minutes consisting of an average of 5.23 points per game. The calculated result differs by 0.275% from a good game time estimate from the US Open 2014 average match times. The winning chances of players were given through a separate smaller 5×5 fundamental matrix. The server had a winning chance of 0.926 and the receiver had a winning chance of 0.074. The chances of both players losing with 3 advantages is minuscule with 2.26×10^{-3}.

1 Introduction

A Markov model describes a series of events wherein the probability of each successive event depends only on the probability of the state before that. Assuming that each point played in a tennis game is independent from the next, a Markov model can be used to approximate the length of a tennis game, set or even a match. This report focuses on a model for a tennis game, taking into consideration the event that a deuce occurs i.e. the point in the game where two consecutive points are required to win the game. An assumption made to simplify this model is the existence of two sets of situation in the game: the rst points up to (30-30) and, if applicable, the additional points played after the (30-30) score (these points will be called deuce as for any point after (30-30), two points are required to win the game). As the game has a definite end with either Player A or Player B winning, two of the rows in a transition matrix for this model will have ones, i.e. it is an absorbing Markov chain. Previous such studies [Ferrante et al., 2017] show such a method used with 17 states to describe the advent of the game. But this study uses the fundamental matrix to approximate the number of points required and thus the time required to finish the game. The player data used in the study involves the top hundred players on the ATP tour as their serve percentages are highly unlikely to change too much with each match.

2 Method and results

Assigning 0.78 as the probability of the player winning the point on serve is a good estimate as the probability of winning a point on serve for the top 100 players ranges from 82.77% to 72.59% [ATP Tour, 2019]. So the probability of losing the point is 0.22 percent in the calculations shown. Through the fundamental matrix as calculated by taking only the non-absorbing states in the transition matrix (Q) and the identity matrix ($\mathrm{I_n}$), the fundamental matrix [Kemeny and Snell, 1983] is

$$F = (I_n - Q)^{-1} \tag{1}$$

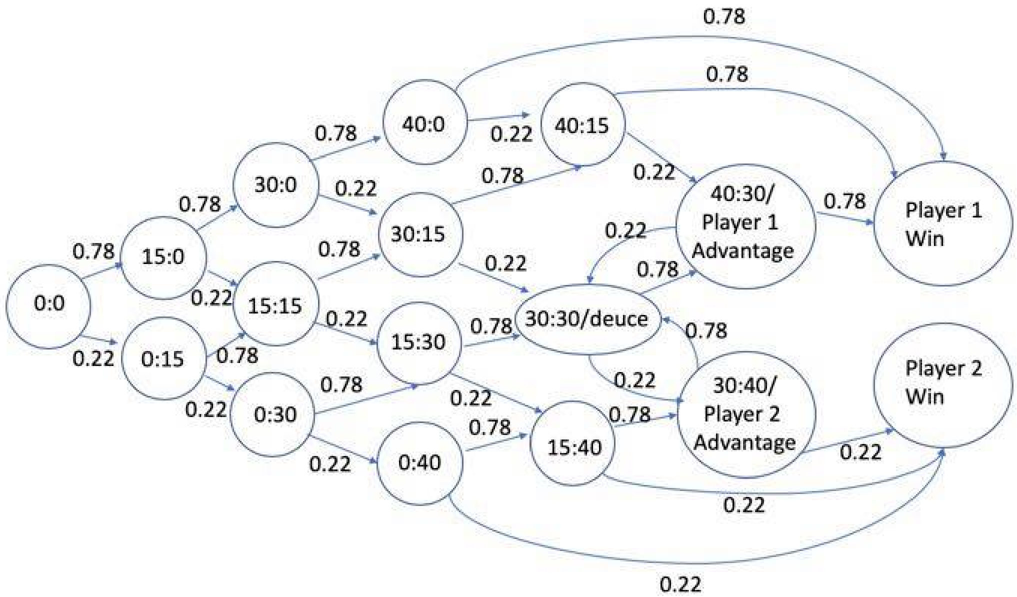

Figure 1: *Markov chain diagram*

$$
\begin{bmatrix}
0 & 0.78 & 0.22 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.78 & 0.22 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.78 & 0.22 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.78 & 0.22 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.78 & 0.22 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.78 & 0.22 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.22 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.78 & 0.22 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.78 & 0.22 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.78 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.22 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.78 & 0.22 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.78 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.22 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.78 & 0 & 0 & 0
\end{bmatrix}
$$

Figure 2: *Matrix Q*

	00-00	15-00	00-15	30-00	15-15	00-30	40-00	30-15	15-30	00-40	40-15	Deuce	15-40	Adv 1	Adv 2
00-00	1	0.78	0.22	0.6084	0.3432	0.0484	0.474552	0.401544	0.113256	0.010648	0.417606	0.330547	0.033222	0.3497	0.098633
15-00	0	1	0	0.78	0.22	0	0.6084	0.3432	0.0484	0	0.401544	0.211889	0.010648	0.253613	0.054921
00-15	0	0	1	0	0.78	0.22	0	0.6084	0.3432	0.0484	0.474552	0.751244	0.113256	0.690372	0.253613
30-00	0	0	0	1	0	0	0.78	0.22	0	0	0.3432	0.098981	0	0.152709	0.021776
15-15	0	0	0	0	1	0	0	0.78	0.22	0	0.6084	0.6122	0.0484	0.611364	0.172436
00-30	0	0	0	0	0	1	0	0	0.78	0.22	0	1.244219	0.3432	0.970491	0.541424
40-00	0	0	0	0	0	0	1	0	0	0	0.22	0.016212	0	0.061045	0.003567
30-15	0	0	0	0	0	0	0	1	0	0	0.78	0.392436	0	0.4777	0.086336
15-30	0	0	0	0	0	0	0	0	1	0	0	1.391364	0.22	1.085264	0.4777
00-40	0	0	0	0	0	0	0	0	0	1	0	0.722521	0.78	0.563567	0.767355
40-15	0	0	0	0	0	0	0	0	0	0	1	0.073691	0	0.277479	0.016212
Deuce	0	0	0	0	0	0	0	0	0	0	0	1.522533	0	1.187576	0.334957
15-40	0	0	0	0	0	0	0	0	0	0	0	0.926309	1	0.722521	0.983788
Adv 1	0	0	0	0	0	0	0	0	0	0	0	0.334957	0	1.261267	0.073691
Adv 2	0	0	0	0	0	0	0	0	0	0	0	1.187576	0	0.926309	1.261267

Figure 3: *Fundamental Matrix*

2.1 Duration of the game

The expected number of visits to from the starting state to the absorbing states can be found by adding up the top row elements of the fundamental matrix, a steady state matrix. So the estimated number of points to end the game ≈ 5.23 and as the average time taken by players is 0.625 minutes per point [Kovalchik, 2013] the total time taken to complete the game is ≈ 3.27 minutes.

The average length of a men's match is 164 minutes [Reference, 2019]. For a game considering a 4 or 5 setter; within each set there being around 10 games on average (both players gain every other game on serve until a break of serve to close out the set). These assumptions mean that the average time to play a game is around 3.64 minutes. The calculated estimate differs from the factual one by about 0.275%. So there are no obvious flaws in the model to disprove it as yet.

2.2 Chances of players losing with 3 advantages

The probabilities at deuce are the same as throughout the game. If the service game involves one or multiple deuces, the win percentage for the server is nearly identical (down by 1/10%) so it can be considered identical to having only one deuce in the game [ATP Tour, 2018]. As both players must be considered in this scenario the first part of the workings focuses when Player A loses and the second part focuses on when Player B loses. Both parts added together gives the total probability. For any player to lose with three advantages, the probability (p_l) can be shown as follows

$$p_l = 0.72^3 \times 0.22^5 + 0.72^5 \times 0.22^3 = 2.26 \times 10^{-3} \tag{2}$$

2.3 Winning chances of players

On creating a smaller transition matrix to deal with the winning chances of the players, the states are Deuce, Player A Adv, Player A win, Player B Adv and Player B win respectively as column headings for the fundamental matrix. The fundamental matrix gives a winning chance for the server (Player A) as 0.926 and a winning chance of 0.074 for Player B. This indicates that the winning probabilities for the server increase as they get into the deuce while for the receiver they drastically decrease.

3 Limitations

The data collection required to assess the winning chances of a player are too vast to be able to compute manually, thus the report only takes a fraction of matches played between top players. The

$$\begin{bmatrix} 0. & 0. & 0.926309378806334 & 0. & 0.0736906211936662 \\ 0. & 0. & 0.983788063337393 & 0. & 0.0162119366626066 \\ 0. & 0. & 1. & 0. & 0. \\ 0. & 0. & 0.722521315468940 & 0. & 0.277478684531060 \\ 0. & 0. & 0. & 0. & 1. \end{bmatrix}$$

Figure 4: *Fundamental Matrix for winning probabilities*

probability of holding on serve differs between the sexes and also through different eras. A winning probability of 0.78 is representative of the top 100 players and depending on the context may or may not be useful to targeted studies into such modelling projects. This model could be used to predict the outcome of a match, but it is increasingly difficult to make a balanced judgement about the outcome the closer to the start of the match you are. That is, it is impossible to say what the outcome of the match is, say, from the outcome of the first game. This is a serious limitation as a Markov model depends entirely on the result preceding the event. The study used only historical data but upon reflection a combination of historical and current data would have resulted in a better model, but the modelling required goes beyond the scope of this report. Finally the dearth of information about the different parameters as collected by systems like HawkEye® would have allowed accurate detail into the motion of the players on court the speed and spin on the ball which shaves off seconds during rallies. The model also does not take into account the time violation penalties or the time between points in this study.

4 Conclusion

A Markov model is a useful way to describe a tennis game as it operates uses discrete information i.e. information independent from preceding events which coincides with the assumption that probabilities of winning points in tennis are independent of each other. From the fundamental matrix an estimate of 3.27 minutes per game (3 s.f) was within 0.275% of the collected data estimate. The chances of a player losing with three advantages was estimated to be small, the actual probability was around 0.002. The transition matrix created for the winning chances of the players from deuce showed that the server has a 0.926 chance of winning the game as compared to the receiver (winning chance of 0.074).

References

[Ferrante et al., 2017] Ferrante, Marco Fonseca, Giovanni Carrari, Andrea. (2017). A new markovian model for tennis matches. Electronic Journal of Applied Statistical Analysis. 10. 10.1285/i20705948v10n3p693.

[ATP Tour, 2019] ATP Tour. (2019). First Serve Points Won — ATP Tour — Tennis. [online] Available at: https://www.atptour.com/en/stats/1st-serve-points-won/all/all/all/?page=1sortByField=percentageascending=False [Accessed 29 Jan. 2019].

[ATP Tour, 2018] ATP Tour. (2018). Infosys Beyond The Numbers: When Games Go To Deuce... — ATP Tour — Tennis. [online] Available at: https://www.atptour.com/en/news/isner-infosys-beyond-the-numbers-september-2018 [Accessed 29 Jan. 2019].

[Kemeny and Snell, 1983] Kemeny, J.G. and Snell, J.L. (1983). Finite Markov chains: with a new appendix" Generalization of a fundamental matrix". Springer.

[Kovalchik, 2013] Kovalchik, S. (2013). Trends in Singles Play Intensity on the ATP Tour. [online] Nessis.org. Available at: http://www.nessis.org/nessis13/kovalchik.pdf [Accessed 30 Jan. 2019].

[Reference, 2019] Reference. (2019). How Long Is a Average Tennis Game?. [online] Available at: https://www.reference.com/sports-active-lifestyle/long-average-tennis-game-6e97416e9ea45aa9 [Accessed 30 Jan. 2019].