Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
`http://siba-ese.unisalento.it/index.php/ejasa/index`
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v10n3p693

A new markovian model for tennis matches

By Carrari, Ferrante, Fonseca

Published: 15 November 2017

Electronic Journal of Applied Statistical Analysis
Vol. 10, Issue 03, November 2017, 693-711
DOI: 10.1285/i20705948v10n3p693

A new markovian model for tennis matches

A. Carrari[a], M. Ferrante[a], and G. Fonseca[*b]

[a] *Università di Padova, Dipartimento di Matematica "Tullio Levi-Civita", Via Trieste 63, 35121-Padova, Italy*
[b] *Università di Udine, Dipartimento di Scienze Economiche e Statistiche, via Tomadini, 30/A, 33100-Udine, Italy*

Published: 15 November 2017

In this paper we present a generalisation of previously considered Markovian models for Tennis that overcome the assumption that the points played are i.i.d. Indeed, we postulate that in any game there are two different situations: the first 6 points and the, possible, additional points after the first deuce, with different winning probabilities. We are able to compute the winning probabilities and the expected number of points played to complete a game and a set in this more general setting. We apply our results considering scores of matches between Novak Djokovic, Roger Federer and Rafael Nadal.

keywords: Markov Chain, Tennis, Winning probability, Expected length of a tennis set.

1. Introduction

Tennis is a sport that can be nicely described with a simple mathematical model. Assuming that the probability that a player wins one point is independent of the previous points and constant during the match, the score of a single game, of a single set and of the whole match can be easily described by a set of homogeneous Markov chains. This approach leads to a series of nice theoretical results and a complete account on this approach can be found e.g. in the recent book by Klaassen and Magnus (2014).

However, the assumptions that the probability to win any point depends only on which player is on service, is independent of the previous points played and constant along the

*Corresponding author: giovanni.fonseca@uniud.it.

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

match, have been criticised by some authors (see e.g. Klaassen and Magnus (2001)). In particular, we focus on the different attitude of the serving player within a game. Commentators often say that many players, especially ones with low ranking, take on differently points played at the beginning or at the end of the game. They suffer of fear of losing or fear of not winning when they serve a game point.

Hence, we propose in this paper a simple modification of the model at the game's level. Indeed, we will assume that during any game there are two different situations: the first points and the, possible, additional points played after the (30,30) score (that in our model coincide with the "Deuce").

Under this assumption, in the present paper, we aim to provide a complete description of the winning probabilities and the expected number of points played in a game and in a set. We will consider separately the games won by the serving player and those won by the receiver (breaks) and we will be able to compute explicitly the expected length of any of these four games: A serves and wins (aA), A serves and loses (aB), B serves and wins (bB) and B serves and loses (bA). The computation of these conditional lengths is, to the best of our knowledge, original (see Ferrante and Fonseca (2014) for a similar approach to volleyball set) and is motivated by the aim to compute in Sect. 3 the expected number of points played in a set, where the exact length of the previous four types of games is needed. Indeed, all the previous results in the literature, concerning the duration of a tennis set (see e.g. Barnett et al. (2006), Carter Jr and Crews (1974)), consider the (expected) number of games needed to complete a set, which is not enough to determine the exact (expected) number of points played.

In Sect. 4 we present an application of the model using the score of 22 matches between Rafael Nadal and Novak Djokovic, 18 matches between Roger Federer and Novak Djokovic and 17 matches between Rafael Nadal and Roger Federer. We will estimate the four parameters of the model and then compare our theoretical probabilities and expected length with the corresponding estimated quantities. The choice of these 3 players, and not others, is forced by the need to obtain estimates based on a large number of observations. On the other hand, best players, as Nadal, Federer and Djokovic, change less than others their way to play within a game and hence are less appropriate to check the validity of the proposed model.

2. Winning probability and expected duration of a game

Let us start considering a single tennis game. The usual assumption is that the probability to win any point by the player on service is independent of the previous points and constant during the game. Let us denote by p this probability, but let us assume that this value does not remain the same during the game. For this reason, we will consider a second parameter $\bar{p}$, that will describe the additional played points from the "Deuce" on. To avoid trivial cases, we will assume that both p and $\bar{p}$ belong to $(0, 1)$.

The state space describing the score of a game is defined in Table 1. Note that in the present model the scores $(30, 30)$ and *Deuce* are represented by the single state 13, since they share the same mathematical properties. The graph representing the transition

Table 1: Scores and corresponding states used in equations

Score	(0,0)	(15,0)	(0,15)	(30,0)	(15,15)	(0,30)	(40,0)	(30,15)	(15,30)
State	1	2	3	4	5	6	7	8	9
Score	(0,40)	(40,15)	(15,40)	*Deuce*	Adv_A	Adv_B	Win_A	Win_B	
State	10	11	12	13	14	15	16	17	

probabilities is presented in Fig. 1, where $q = 1 - p$ and $\bar{q} = 1 - \bar{p}$:
From the graph it is immediate to define the transition matrix $P = (p_{ij})_{i,j \in S}$ and to prove that the states 16 and 17 are absorbing, while all the other states are transient. In order to compute the winning probabilities, we will need to determine the absorption probabilities in the states 16 and 17, while to investigate the expected length of a game, we will need to evaluate the mean absorption times for the conditional chains, that we will define later.

2.1. Winning probability of a game

The winning probability of the game for the player on service coincides with the absorption probability in the state 16 of the previous Markov chain starting from state 1, which can be obtained (see e.g. Norris (1998)) as the minimal, non negative solution of the linear system

$$\begin{cases} h_i = \sum_{j \in S} p_{ij} h_j & \text{for } 1 \leq i \leq 15 \\ h_{16} = 1 \ , \ h_{17} = 0 \ . \end{cases}$$

where the p_{ij} are the transition probabilities defined previously and the h_i are the conditional absorption probabilities in state 16 given the initial state equal to i.

The solution can be easily calculated and we obtain that

$$h_1(p, \bar{p}) = p^2 \left[5p^2 - 4p^3 + 4(p-1)^2 p\bar{p} - \frac{2(p-1)^2 \bar{p}^2 (p(4\bar{p} - 2) - 2\bar{p} - 3)}{2\bar{p}^2 - 2\bar{p} + 1} \right] .$$

Denoting by A and B the two players and by P^G_{xY} the probability that the player Y wins a game when X serves, we obtain that:

$$P^G_{aA} = h_1(p_A, \bar{p}_A) \quad , \quad P^G_{aB} = h_1(1 - p_A, 1 - \bar{p}_A)$$
$$P^G_{bB} = h_1(p_B, \bar{p}_B) \quad , \quad P^G_{bA} = h_1(1 - p_B, 1 - \bar{p}_B)$$

Note that, since $P^G_{xX} + P^G_{xY} = 1$, $h_1(1 - p_X, 1 - \bar{p}_X) = 1 - h_1(p_X, \bar{p}_X)$ and that for $p_X = \bar{p}_X$, the previous probabilities coincide with those well known in the literature (see e.g. Newton and Keller (2005)). In Table 2 we report the values of h_1 for combinations of p and $\bar{p}$ values.

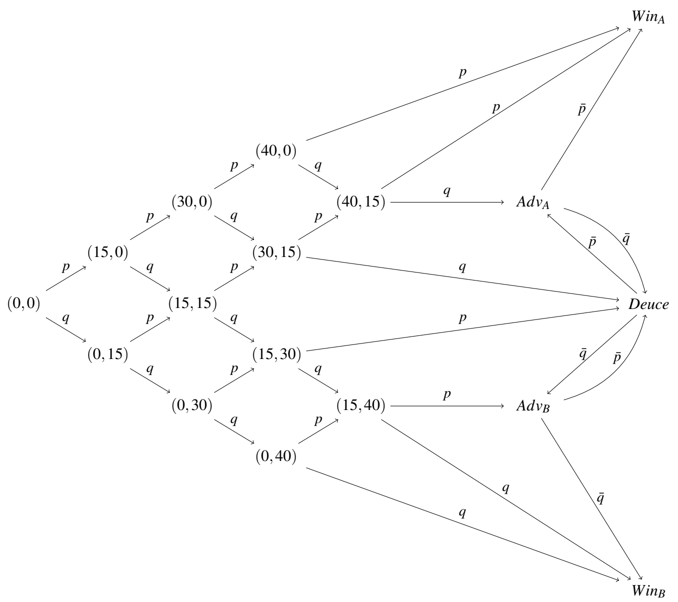

Figure 1: Graph of the Markov chain describing a tennis game

2.2. Expected length of a game

Let us now compute the expected length of a game, i.e. the expected number of points played in a game. Since in the next section we will need to know the expected length of a game won by the player serving or receiving the serve, we have to consider separately the expected length of the paths starting from 1 and ending in 16 or 17, respectively. Let us consider the case that the serving player wins the game. The computation of the mean length of such a game can be easily performed in the Markov chain framework (see e.g. Kemeny and Snell (1976), Section 3.5), by defining the Markov chain conditioned to the event {The player on service wins the game}. The conditioned transition matrix P' on the state space $\{1, \ldots, 16\}$ is given by:

$$p'_{ij} = p_{ij} \frac{h_j}{h_i} \quad \text{with } i, j \in \{1, \ldots, 16\},$$

where the h_i are the absorption probabilities in 16 computed before. So, in order to evaluate the expected duration of such a game, it will be sufficient to solve the linear

Table 2: Winning probabilities of a game

	$\bar{p}$									
p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	0.001	0.004	0.011	0.021	0.034	0.049	0.062	0.071	0.078	0.081
0.2	0.011	0.022	0.043	0.076	0.119	0.165	0.204	0.233	0.252	0.263
0.3	0.040	0.061	0.099	0.158	0.234	0.312	0.378	0.425	0.455	0.472
0.4	0.102	0.132	0.185	0.264	0.363	0.464	0.549	0.607	0.643	0.663
0.5	0.206	0.242	0.302	0.391	0.500	0.609	0.697	0.758	0.794	0.812
0.6	0.357	0.392	0.451	0.535	0.636	0.736	0.815	0.868	0.898	0.913
0.7	0.545	0.575	0.622	0.688	0.766	0.842	0.901	0.939	0.960	0.969
0.8	0.748	0.767	0.795	0.835	0.881	0.924	0.957	0.978	0.989	0.993
0.9	0.922	0.929	0.938	0.951	0.965	0.979	0.989	0.995	0.998	0.999
1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

system:

$$\begin{cases} k_i = 1 + \sum_{j \in S'} p'_{ij} k_j & \text{for } 1 \le i \le 15 \\ k_{16} = 0 \ , \end{cases}$$

(see Norris (1998) for the proof) where the k_i are the conditional expected absorption times in state 16 given the initial state equal to i. It follows that:

$$\begin{aligned} k_1(p,\bar{p}) = 4\Big[&\bar{p}^2(9 - 4\bar{p} + 12\bar{p}^3) + p^3(-5 + 26\bar{p} - 56\bar{p}^2 + 60\bar{p}^3 - 32\bar{p}^4) \\ &- 2p\bar{p}(-3 + 17\bar{p} - 14\bar{p}^2 + 6\bar{p}^3 + 12\bar{p}^4)\Big]\Big[(1 - 2\bar{p} + 2\bar{p}^2)(2\bar{p}^2(3 + 2\bar{p}) \\ &- 4p\bar{p}(-1 + 4\bar{p} + 2\bar{p}^2) - 4p^3(1 - 3\bar{p} + 3\bar{p}^2) + p^2(5 - 18\bar{p} + 24\bar{p}^2 + 4\bar{p}^3))\Big]^{-1} \\ &+ \Big[4(p^2(6 - 36\bar{p} + 89\bar{p}^2 - 92\bar{p}^3 + 48\bar{p}^4 + 12\bar{p}^5))\Big] \\ &\Big[(1 - 2\bar{p} + 2\bar{p}^2)(2\bar{p}^2(3 + 2\bar{p}) - 4p\bar{p}(-1 + 4\bar{p} + 2\bar{p}^2) - 4p^3(1 - 3\bar{p} + 3\bar{p}^2) \\ &+ p^2(5 - 18\bar{p} + 24\bar{p}^2 + 4\bar{p}^3))\Big]^{-1}. \end{aligned}$$

Using the same notation as before, we get the expected length of the four types of outcomes of a game

$$\begin{aligned} k^G_{aA} &= k_1(p_A, \bar{p}_A) \quad , \quad k^G_{aB} = k_1(1 - p_A, 1 - \bar{p}_A) \\ k^G_{bB} &= k_1(p_B, \bar{p}_B) \quad , \quad k^G_{bA} = k_1(1 - p_B, 1 - \bar{p}_B). \end{aligned}$$

In Table 3 we present the expected conditional length of a game won by the player on service, k^G_{aA}. Note that these are *conditional* lengths and this fact justifies some unexpected values included in the table. For example the length is maximum for $p \approx 0$ and $\bar{p} \approx 0.5$, which can be justified by the fact that, conditioned on the event $\{A \text{ wins}\}$, the path that arrives to the state 16 almost never reaches this state without reaching first the *Deuce* and here the second parameter close to 0.5 makes this part of the game as long as possible.

Table 3: Expected duration of a game won by player on service

	$\bar{p}$									
p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	5.912	6.804	7.521	8.064	8.332	8.274	7.959	7.525	7.088	6.709
0.2	5.388	6.268	7.100	7.739	8.075	8.064	7.783	7.374	6.956	6.591
0.3	5.108	5.792	6.607	7.310	7.721	7.774	7.542	7.172	6.783	6.441
0.4	4.938	5.423	6.117	6.811	7.274	7.395	7.226	6.906	6.558	6.247
0.5	4.814	5.141	5.673	6.283	6.750	6.923	6.821	6.564	6.269	6.000
0.6	4.708	4.917	5.285	5.760	6.171	6.366	6.325	6.138	5.907	5.690
0.7	4.599	4.721	4.948	5.265	5.572	5.745	5.746	5.629	5.469	5.316
0.8	4.471	4.531	4.644	4.812	4.988	5.101	5.116	5.059	4.972	4.886
0.9	4.293	4.311	4.345	4.396	4.453	4.492	4.500	4.484	4.458	4.431
1.0	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000

From the previous formulas, when the serving player has probabilities $(p, \bar{p})$ to win the points, we can easily derive the expected length of a game, computed in Table 4:

$$k_1(p,\bar{p})h_1(p,\bar{p}) + k_1(1-p,1-\bar{p})h_1(1-p,1-\bar{p}) \ .$$

3. Winning probability and expected length of a set

In order to evaluate the winning probability and expected length of a set, we will assume that every game is played independently and has the same Markovian structure defined in the previous section. The more natural approach would be to consider the Markov chain describing the set and obtaining, as done before, the required probabilities and expected values. This approach is well-known in the literature, but it doesn't work if we are interested in evaluating the expected number of points played in a set, Indeed, the Markovian approach allows us to calculate the expected number of games played in

Table 4: Expected duration of a game

	$\bar{p}$								
p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	4.455	4.489	4.533	4.567	4.580	4.586	4.569	4.539	4.511
0.2	4.977	5.086	5.201	5.301	5.355	5.349	5.280	5.193	5.097
0.3	5.455	5.639	5.831	5.992	6.075	6.048	5.929	5.763	5.593
0.4	5.808	6.044	6.287	6.483	6.565	6.513	6.351	6.119	5.898
0.5	5.969	6.220	6.467	6.673	6.750	6.673	6.467	6.220	5.969
0.6	5.898	6.119	6.351	6.513	6.565	6.483	6.287	6.044	5.808
0.7	5.593	5.763	5.929	6.048	6.075	5.992	5.831	5.639	5.455
0.8	5.097	5.193	5.280	5.349	5.355	5.301	5.201	5.086	4.977
0.9	4.511	4.539	4.569	4.586	4.580	4.567	4.533	4.489	4.455

a set, but these games have not all the same expected lengths, as shown in the previous computations.

As an alternative, we will compute the probability that a given set ends with one of the possible fourteen scores $(6,0),(6,1),\ldots,(7,6),(6,7)$, taking also into account the number of breaks won by each of the two players. Then we will evaluate the expected conditional length of any set ended with a given score and number of breaks and we will take its expectation in order to compute the average number of points played in a set.

Let us start by considering the tiebreak, the special game played when a set reaches the score $(6,6)$. Then we will analyse separately the winning probabilities of a set and its expected length.

3.1. Winning probability and expected length of a tiebreak

The tiebreak is a special type of game. It is played when the score in a set is equal to $(6,6)$ in order to determine the winner of the set. In the tiebreak 7 points, with a two-point advantage, are needed to win the game. The service rotate every two points, except for the first service, played by the player who started serving in the first game of the set. If we assume that the probabilities to win a point for the player on service are fixed during the tiebreak, the tiebreak itself can be model as a Markov chain. Due to the particular service change rule, the state space includes the following 53 states

$$\bar{S} = \Big(\{(i,j) : i,j \in \{0,\ldots,6\}, i+j \leq 10\} \setminus \{(5,5)\}\Big) \cup \{Deuce_A, Deuce_B, Adv_{AA}, Adv_{BB}, Adv_{AB}, Adv_{BA}, Win_A, Win_B\}$$

where $Deuce_A$ means that the two players scored both the same number of points, 5 or more, and A will serve the next point, Adv_{AA} means that player A is in advantage and will serve next, and so on. The transition probabilities are defined as before, paying now some attention to which player is, at every possible score, on service. For example, if $p_{(0:0),(1:0)} = p_A$, then $p_{(1:0),(2:0)} = 1 - p_B$, $p_{(2:0),(3:0)} = 1 - p_B$, $p_{(3:0),(4:0)} = p_A$, and so on. In Newton and Keller (2005), pages 266-268, the winning probabilities of the tiebreak, that we do not report here, are calculated by a recursive approach. On the contrary, the expected number of points played is not computed. To obtain this value, we have to solve, as before, a linear system of equations

$$\begin{cases} k_i = 1 + \sum_{j \in \bar{S}} p_{ij} k_j \ , & \text{for } 1 \leq i \leq 51 \\ k_{52} = 0 \ , \\ k_{53} = 0 \ . \end{cases}$$

Using *Mathematica*®, we obtain that the expected number of points played to complete the tiebreak is equal to

$$\begin{aligned} k_1 = [&-2 - 30p_B^2 + 71p_B^3 - 94p_B^4 + 73p_B^5 - 30p_B^6 + 5p_B^7 + \\ &+ p_A p_B(-115 + 541p_B - 1166p_B^2 + 1483p_B^3 - 1124p_B^4 + 465p_B^5 - 80p_B^6) + \\ &+ p_A^2(-25 + 500p_B - 2650p_B^2 + 6514p_B^3 - 8716p_B^4 + 6557p_B^5 - 2600p_B^6 + \\ &+ 420p_B^7) + p_A^3(40 - 885p_B + 5730p_B^2 - 16276p_B^3 + 23769p_B^4 - 18398p_B^5 + \\ &+ 7000p_B^6 - 980p_B^7) + p_A^4(-25 + 792p_B - 6224p_B^2 + 20289p_B^3 - 32532p_B^4 + \\ &+ 26320p_B^5 - 9660p_B^6 + 1050p_B^7) + p_A^5(4 - 347p_B + 3353p_B^2 - 12428p_B^3 + \\ &+ 21784p_B^4 - 18466p_B^5 + 6510p_B^6 - 420p_B^7) + p_A^6(1 + 58p_B - 716p_B^2 + \\ &+ 2996p_B^3 - 5698p_B^4 + 5040p_B^5 - 1680p_B^6)]/[-p_B + p_A(-1 + 2p_B)]. \end{aligned}$$

Note that we do not separate the case of a tiebreak won by player A or by player B. This computation would be complicate and not extremely useful. Indeed, since the number of tiebreaks played is usually scarce, separating the two cases will be of small practical interest. As a consequence, in the sequel, we will not be able to obtain the expected length of a set won by a fixed player.

3.2. Winning probability of a set

Let us now consider a set of tennis which may end also with a tiebreak. In this section we will calculate explicitly the probability that the final score of the set will be one of the seven possible pairs $(6,0), (6,1), (6,2), (6,3), (6,4)$, $(7,5), (7,6)$ and we will derive the results for the remaining cases easily. To simplify the exposition, if player A (resp. B) starts serving in the first game, we will denote by $\mathbb{P}_a$ (resp. $\mathbb{P}_b$) the conditional probability given this event. We will perform these calculations in order to evaluate the average number of points needed to complete a tennis set, and the present probabilities represent a basic ingredient.

Table 5: Expected duration of a TieBreak

	p_A										
p_B	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	∞	22.12	13.59	11.21	10.09	9.34	8.74	8.20	7.73	7.33	7.00
0.1	22.38	16.18	13.55	12.00	10.91	10.06	9.36	8.75	8.23	7.78	7.40
0.2	13.80	13.61	13.02	12.28	11.50	10.73	10.01	9.35	8.76	8.25	7.81
0.3	11.29	12.04	12.29	12.18	11.81	11.27	10.65	10.00	9.35	8.76	8.24
0.4	10.07	10.92	11.51	11.81	11.84	11.62	11.21	10.65	10.01	9.35	8.72
0.5	9.30	10.05	10.73	11.28	11.62	11.74	11.62	11.28	10.73	10.05	9.30
0.6	8.72	9.35	10.01	10.65	11.21	11.62	11.84	11.81	11.51	10.92	10.07
0.7	8.24	8.76	9.35	10.00	10.65	11.27	11.81	12.18	12.29	12.04	11.29
0.8	7.81	8.25	8.76	9.35	10.01	10.73	11.50	12.28	13.02	13.61	13.80
0.9	7.40	7.78	8.23	8.75	9.36	10.06	10.91	12.00	13.55	16.18	22.38
1.0	7.00	7.33	7,73	8.20	8.74	9.34	10.09	11.21	13.59	22.12	∞

The computation of these probabilities is a little bit tedious and we report the formulas in the final Appendix A. However with a compact notation it will be possible at least to write down the formulas in an efficient way. With some abuse of notation, we will define first the random vector (A, B), which indicates the final score of a set, and then an additional five dimensional random vector X, that denotes the final score of the set in more detail. More precisely, $X = (x_0, x_1, x_2, x_3, x_4)$ summarizes who starts serving (0 for A and 1 for B) and the number of aA, aB, bB and bA games, respectively. So, given A starts serving, the event $\{(A, B) = (6, 0)\}$ will coincide with $\{X = (0, 3, 0, 0, 3)\}$, which indicates that the set finished with three aA games and three bA games, while $\{(A, B) = (6, 1)\}$ will coincide with $\{X = (0, 4, 0, 1, 2)\} \cup \{X = (0, 3, 1, 0, 3)\}$ and so on for the other cases. In order to evaluate the probability that a set ends with a given score, we will therefore determine all the possible admissible combinations of games that leads to that score and thanks to the Markovian assumption, substitute their probabilities into the formulas. We get, for example, that

$$\mathbb{P}_a[(A, B) = (6, 0)] = \mathbb{P}_a[(6, 0)] = \mathbb{P}[X = (0, 3, 0, 0, 3)] = (P^G_{aA})^3 (P^G_{bA})^3$$

$$\mathbb{P}_b[(A, B) = (6, 0)] = \mathbb{P}_b[(6, 0)] = \mathbb{P}[X = (1, 3, 0, 0, 3)] = (P^G_{aA})^3 (P^G_{bA})^3$$

and we refer to the Appendix A for the other cases.

3.3. Expected length of a set

In this subsection we will evaluate the expected number of points played in a set. We will assume that player A starts serving in the first game and we will present the computations just for the case that A wins. The results for the other cases can be easily obtained. Let D be the random duration of a set; we will evaluate its expectation by $\mathbb{E}_a[D] = \mathbb{E}_a[\mathbb{E}_a[D|(A,B)]]$, where $\mathbb{E}_a[D|(A,B)]$ will denote the conditional duration given a specific final score and the fact that player A starts serving. Since who starts serving is fixed here, we will use the compact notation for the winning probabilities $\mathbb{P}[x_1, x_2, x_3, x_4]$. Denoting $\mathbb{E}_a[D|(i,j)] = \mathbb{E}_a[D|(A,B) = (i,j)]$, we have:

$$\begin{aligned} \mathbb{E}_a[D] = &\sum_{i=0}^{4} \mathbb{E}_a[D|(6,i)] \cdot \mathbb{P}_a[(A,B) = (6,i)] + \mathbb{E}_a[D|(7,5)] \cdot \mathbb{P}_a[(A,B) = (7,5)] \\ &+ \big(\mathbb{E}_a[D|(7,6)] + \mathbb{E}_a[D|(6,7)]\big) \cdot \mathbb{P}_a[(A,B) = (6,6)] \\ &+ \sum_{i=0}^{4} \mathbb{E}_a[D|(i,6)] \cdot \mathbb{P}_a[(A,B) = (i,6)] + \mathbb{E}_a[D|(5,7)] \cdot \mathbb{P}_a[(A,B) = (5,7)]. \end{aligned} \tag{1}$$

The conditional expectations used in (1) are calculated in Appendix B.

In Table 6 we present the expected duration of a set, as a function of p_A and p_B. For simplicity, we consider just the case of the model with a unique parameter ($p_A = \bar{p}_A$ and $p_B = \bar{p}_B$).

Table 6: Expected duration of a set

	p_A								
p_B	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7
0.3	73.76	73.11	71.09	67.80	63.70	59.41	55.39	51.87	48.88
0.35	73.97	74.59	73.82	71.47	67.81	63.53	59.28	55.47	52.26
0.4	72.65	74.59	75.39	74.52	71.90	68.08	63.83	59.81	56.36
0.45	69.69	72.69	75.05	75.99	75.04	72.35	68.61	64.65	61.06
0.5	65.49	69.07	72.58	75.24	76.27	75.35	72.81	69.44	66.01
0.55	60.78	64.47	68.53	72.38	75.19	76.27	75.49	73.33	70.59
0.6	56.23	59.76	63.87	68.24	72.21	75.01	76.13	75.64	74.11
0.65	52.24	55.52	59.43	63.83	68.31	72.24	74.96	76.22	76.24
0.7	48.93	52.00	55.66	59.87	64.44	68.94	72.79	75.57	77.21

4. Comparison with real data: Nadal, Djokovic, Federer

In this section we use data from real played matches in order to check if the proposed model is able to predict the probability of winning a game and its expected length given the estimated probabilities to win a point. Data are obtained from *www.tennis.earth.com* that provides the point-by-point scores of the matches.
We consider the matches between Rafael Nadal, Novak Djokovic and Roger Federer in the period 2009–2014. These players have one of the largest number of played matches against each other. In Table 7 we summarize the dimension of the dataset.

Table 7: Number of Matches, Sets and Games

Players	Match	Set	Game
Nadal vs. Djokovic	22	63	610
Nadal vs. Federer	17	46	448
Federer vs. Djokovic	18	51	486

In the sequel we evaluate for each serving player either the case of a constant point winning probability and the case of two different *pre*-deuce and *post*-deuce point winning probabilities. For each of the three players, we calculate the relative frequencies of winning a point over all the played points, and also splitting the points in a *pre*-deuce play and *post*-deuce play, considering also the actual opponent player.
In Table 8–10 we summarize the point probabilities and we compare the probabilities of winning a game based on data ($\hat{h}_1$) and the exact probabilities (h_1) calculated using the estimated winning point probabilities ($p, \bar{p}$). Remember also that $P^G_{aB} = 1 - P^G_{aA}$.

Table 8: Probabilities of winning a game: Nadal (A) vs. Djokovic (B)

p	$\bar{p}$	h_1	$\hat{h}_1$
$p_A = 0.59459$	$\bar{p}_A = 0.59459$	$P^G_{aA} = 0.72434$	0.71237
$p_B = 0.62811$	$\bar{p}_B = 0.62811$	$P^G_{bB} = 0.79119$	0.75563
$p_A = 0.60500$	$\bar{p}_A = 0.57190$	$P^G_{aA} = 0.71529$	0.71237
$p_B = 0.63881$	$\bar{p}_B = 0.59705$	$P^G_{bB} = 0.77704$	0.75563

Similarly, we estimate the expected length of the game using the proposed Markovian model (k_1) and compare it with the mean duration of a game calculated on the played games recorded in the data ($\hat{k}_1$). Results are summarized in Tables 11–13: the two

Table 9: Probabilities of winning a game: Nadal(A) vs. Federer(B)

p	$\bar{p}$	h_1	$\hat{h}_1$
$p_A = 0.6484$	$\bar{p}_A = 0.6484$	$P^G_{aA} = 0.8271$	0.8288
$p_B = 0.6073$	$\bar{p}_B = 0.6073$	$P^G_{bB} = 0.7507$	0.7257
$p_A = 0.6414$	$\bar{p}_A = 0.6676$	$P^G_{aA} = 0.83373$	0.8288
$p_B = 0.6186$	$\bar{p}_B = 0.5858$	$P^G_{bB} = 0.73789$	0.7275

Table 10: Probabilities of winning a game: Federer(A) vs. Djokovic (B)

p	$\bar{p}$	h_1	$\hat{h}_1$
$p_A = 0.6287$	$\bar{p}_A = 0.6287$	$P^G_{aA} = 0.7923$	0.7597
$p_B = 0.6449$	$\bar{p}_B = 0.6449$	$P^G_{bB} = 0.8211$	0.8142
$p_A = 0.6357$	$\bar{p}_A = 0.6079$	$P^G_{aA} = 0.7828$	0.7597
$p_B = 0.6186$	$\bar{p}_B = 0.6321$	$P^G_{bB} = 0.8163$	0.8142

Table 11: Expected length of a game: Nadal (A) vs. Djokovic (B)

p	$\bar{p}$	k_1	$\hat{k}_1$
$p_A = 0.59459$	$\bar{p}_A = 0.59459$	$K^G_{aA} = 6.39459$	6.28910
$p_B = 0.62811$	$\bar{p}_B = 0.62811$	$K^G_{bB} = 6.20935$	5.71368
$1 - p_A = 0.40541$	$1 - \bar{p}_A = 0.40541$	$K^G_{aB} = 6.81638$	7.01176
$1 - p_B = 0.37189$	$1 - \bar{p}_B = 0.37189$	$K^G_{bA} = 6.77603$	6.75000
$p_A = 0.60500$	$\bar{p}_A = 0.57190$	$K^G_{aA} = 6.30753$	6.28910
$p_B = 0.63881$	$\bar{p}_B = 0.59705$	$K^G_{bB} = 6.12896$	5.71368
$1 - p_A = 0.39500$	$1 - \bar{p}_A = 0.42810$	$K^G_{aB} = 6.99899$	7.01176
$1 - p_B = 0.36119$	$1 - \bar{p}_B = 0.40295$	$K^G_{bA} = 7.02878$	6.75000

parameter model results are generally closer to the empirical evidence, even if this is not true all the times.
We conclude this section considering the expected length of a set given first serving player. We can compare the average length calculated via the proposed model and the mean observed length, even if the number of sets in our dataset is quite lower with respect to the number of games.
Recalling that, using our model, we cannot calculate the expected length of a tiebreak given its winner, we use only sets in the dataset that ended without playing the tiebreak. In Tables 14–16, we present the expected lengths obtained from data ($\hat{k}$) and from the proposed model with one ($k_{(p,p)}$) and two different point probabilities ($k_{(p,\bar{p})}$). Note that, in this case, there isn't any significant difference in the results between the models with one and two point probabilities, since the difference at the game level are very small. Nevertheless, in 9 over 12 considered cases, the models estimates and the empirical mean durations are very close to each other.
In conclusion, data seem to confirm that the proposed model is a good description of a tennis match. Further comparisons with real data of matches should be performed, but, usually, the numbers of challenges between two given players are too small in order to obtain an accurate point probability estimate.

A. Appendix

The winning probabilities in Subsection 3.2 are:

$$\mathbb{P}_a[(6,0)] = \mathbb{P}[X = (0,3,0,0,3)] = (P^G_{aA})^3(P^G_{bA})^3$$

$$\mathbb{P}_b[(6,0)] = \mathbb{P}[X = (1,3,0,0,3)] = (P^G_{aA})^3(P^G_{bA})^3$$

Table 12: Expected length of a game: Nadal(A) vs Federer(B)

p	$\bar{p}$	k_1	$\hat{k}_1$
$p_A = 0.6484$	$\bar{p}_A = 0.6484$	$K_{aA}^G = 6.0857$	5.9185
$p_B = 0.6073$	$\bar{p}_B = 0.6073$	$K_{bB}^G = 6.3274$	6.1220
$1 - p_A = 0.3516$	$1 - \bar{p}_A = 0.3516$	$K_{aB}^G = 6.7383$	6.9737
$1 - p_B = 0.3927$	$1 - \bar{p}_B = 0.3927$	$K_{bA}^G = 6.8045$	6.9180
$p_A = 0.6414$	$\bar{p}_A = 0.6676$	$K_{aA}^G = 6.1207$	5.9185
$p_B = 0.6186$	$\bar{p}_B = 0.5858$	$K_{bB}^G = 6.2353$	6.1220
$1 - p_A = 0.3586$	$1 - \bar{p}_A = 0.3324$	$K_{aB}^G = 6.5629$	6.9737
$1 - p_B = 0.3814$	$1 - \bar{p}_B = 0.4142$	$K_{bA}^G = 7.0260$	6.9180

$$\begin{aligned}\mathbb{P}_a[(6,1)] &= 3 \cdot \mathbb{P}[(0,4,0,1,2)] + 3 \cdot \mathbb{P}[(0,3,1,0,3)] \\ &= 3[(P_{aA}^G)^4 P_{bB}^G (P_{bA}^G)^2] + 3[(P_{aA}^G)^3 P_{aB}^G (P_{bA}^G)^3]\end{aligned}$$

$$\begin{aligned}\mathbb{P}_b[(6,1)] &= 3 \cdot \mathbb{P}[(1,3,0,1,3)] + 3 \cdot \mathbb{P}[(1,2,1,0,4)] \\ &= 3[(P_{aA}^G)^3 P_{bB}^G (P_{bA}^G)^3] + 3[(P_{aA}^G)^2 P_{aB}^G (P_{bA}^G)^4]\end{aligned}$$

$$\begin{aligned}\mathbb{P}_a[(6,2)] &= 3 \cdot \mathbb{P}[(0,4,0,2,2)] + 12 \cdot \mathbb{P}[(0,3,1,1,3)] + 6 \cdot \mathbb{P}[(0,2,2,0,4)] \\ &= 3[(P_{aA}^G)^4 (P_{bB}^G)^2 (P_{bA}^G)^2] + 12[(P_{aA}^G)^3 P_{aB}^G P_{bB}^G (P_{bA}^G)^3] \\ &+ 6[(P_{aA}^G)^2 (P_{aB}^G)^2 (P_{bA}^G)^4]\end{aligned}$$

$$\begin{aligned}\mathbb{P}_b[(6,2)] &= 6 \cdot \mathbb{P}[(1,4,0,2,2)] + 12 \cdot \mathbb{P}[(1,3,1,1,3)] + 3 \cdot \mathbb{P}[(1,2,2,0,4)] \\ &= 6[(P_{aA}^G)^4 (P_{bB}^G)^2 (P_{bA}^G)^2] + 12[(P_{aA}^G)^3 P_{aB}^G P_{bB}^G (P_{bA}^G)^3] \\ &+ 3[(P_{aA}^G)^2 (P_{aB}^G)^2 (P_{bA}^G)^4]\end{aligned}$$

$$\begin{aligned}\mathbb{P}_a[(6,3)] &= 4 \cdot \mathbb{P}[(0,5,0,3,1)] + 24 \cdot \mathbb{P}[(0,4,1,2,2)] + 24 \cdot \mathbb{P}[(0,3,2,1,3)] \\ &+ 4 \cdot \mathbb{P}[(0,2,3,0,4)] \\ &= 4[(P_{aA}^G)^5 (P_{bB}^G)^3 P_{bA}^G] + 24[(P_{aA}^G)^4 P_{aB}^G (P_{bB}^G)^2 (P_{bA}^G)^2] \\ &+ 24[(P_{aA}^G)^3 (P_{aB}^G)^2 P_{bB}^G (P_{bA}^G)^3] + 4[(P_{aA}^G)^2 (P_{aB}^G)^3 (P_{bA}^G)^4]\end{aligned}$$

$$\begin{aligned}\mathbb{P}_b[(6,3)] &= 4 \cdot \mathbb{P}[(1,4,0,3,2)] + 24 \cdot \mathbb{P}[(1,3,1,2,3)] + 24 \cdot \mathbb{P}[(1,2,2,1,4)] \\ &+ 4 \cdot \mathbb{P}[(1,1,3,0,5)] \\ &= 4[(P_{aA}^G)^4 (P_{bB}^G)^3 (P_{bA}^G)^2] + 24[(P_{aA}^G)^3 P_{aB}^G (P_{bB}^G)^2 (P_{bA}^G)^3] \\ &+ 24[(P_{aA}^G)^2 (P_{aB}^G)^2 P_{bB}^G (P_{bA}^G)^4] + 4[P_{aA}^G (P_{aB}^G)^3 (P_{bA}^G)^5]\end{aligned}$$

Table 13: Expected length of a game: Federer(A) vs Djokovic(B)

p	$\bar{p}$	k_1	$\hat{k}_1$
$p_A = 0.6287$	$\bar{p}_A = 0.6287$	$K^G_{aA} = 6.2059$	5.8000
$p_B = 0.6449$	$\bar{p}_B = 0.6449$	$K^G_{bB} = 6.1078$	6.0197
$1 - p_A = 0.3713$	$1 - \bar{p}_A = 0.3713$	$K^G_{aB} = 6.7751$	6.7636
$1 - p_B = 0.3551$	$1 - \bar{p}_B = 0.3551$	$K^G_{bA} = 6.7455$	7.0435
$p_A = 0.6357$	$\bar{p}_A = 0.6079$	$K^G_{aA} = 6.1556$	5.8000
$p_B = 0.6186$	$\bar{p}_B = 0.6321$	$K^G_{bB} = 6.0784$	6.0197
$1 - p_A = 0.3643$	$1 - \bar{p}_A = 0.3921$	$K^G_{aB} = 6.9464$	6.7636
$1 - p_B = 0.3814$	$1 - \bar{p}_B = 0.3679$	$K^G_{bA} = 6.8593$	7.0435

Table 14: Expected length given first serving player and set winner: Nadal(A) vs. Djokovic(B)

	$\hat{k}$	$k_{(p,p)}$	$k_{(p,\bar{p})}$
aA	49.50	59.75	59.95
aB	58.89	60.55	60.69
bB	56.02	58.13	58.51
bA	64.67	62.19	62.10

$$\begin{aligned}\mathbb{P}_a[(6,4)] &= \mathbb{P}[(0,5,0,4,1)] + 20 \cdot \mathbb{P}[(0,4,1,3,2)] + 60 \cdot \mathbb{P}[(0,3,2,2,3)] \\ &\quad + 40 \cdot \mathbb{P}[(0,2,3,1,4)] + 5 \cdot \mathbb{P}[(0,1,4,0,5)] \\ &= [(P^G_{aA})^5 (P^G_{bB})^4 P^G_{bA}] + 20[(P^G_{aA})^4 P^G_{aB} (P^G_{bB})^3 (P^G_{bA})^2] \\ &\quad + 60[(P^G_{aA})^3 (P^G_{aB})^2 (P^G_{bB})^2 (P^G_{bA})^3] \\ &\quad + 40[(P^G_{aA})^2 (P^G_{aB})^3 P^G_{bB} (P^G_{bA})^4] + 5[P^G_{aA} (P^G_{aB})^4 (P^G_{bA})^5]\end{aligned}$$

$$\begin{aligned}\mathbb{P}_b[(6,4)] &= 5 \cdot \mathbb{P}[(1,5,0,4,1)] + 40 \cdot \mathbb{P}[(1,4,1,3,2)] + 60 \cdot \mathbb{P}[(1,3,2,2,3)] \\ &\quad + 20 \cdot \mathbb{P}[(1,2,3,1,4)] + \mathbb{P}[(1,1,4,0,5)] \\ &= 5[(P^G_{aA})^5 (P^G_{bB})^4 P^G_{bA}] + 40[(P^G_{aA})^4 P^G_{aB} (P^G_{bB})^3 (P^G_{bA})^2] \\ &\quad + 60[(P^G_{aA})^3 (P^G_{aB})^2 (P^G_{bB})^2 (P^G_{bA})^3] \\ &\quad + 20[(P^G_{aA})^2 (P^G_{aB})^3 P^G_{bB} (P^G_{bA})^4] + [P^G_{aA} (P^G_{aB})^4 (P^G_{bA})^5]\end{aligned}$$

Table 15: Expected length given first serving player and set winner: Nadal(A) vs. Federer(B)

	$\hat{k}$	$k_{(p,p)}$	$k_{(p,\bar{p})}$
aA	58.37	57.32	57.51
aB	58.02	61.86	61.84
bB	50.96	59.15	59.36
bA	59.80	60.00	59.98

Table 16: Expected length given first serving player and set winner: Federer(A) vs. Djokovic(B)

	$\hat{k}$	$k_{(p,p)}$	$k_{(p,\bar{p})}$
aA	58.37	58.32	58.39
aB	58.02	60.45	60.29
bB	50.96	57.63	57.62
bA	59.80	61.15	61.12

$$\begin{aligned}
\mathbb{P}_a[(7,5)] = \mathbb{P}_b[(7,5)] &= \mathbb{P}[(x_0,6,0,5,1)] + 25\cdot\mathbb{P}[(x_0,5,1,4,2)] \\
&+ 100\cdot\mathbb{P}[(x_0,4,2,3,3)] + 100\cdot\mathbb{P}[(x_0,3,3,2,4)] \\
&+ 25\cdot\mathbb{P}[x_0,3,4,1,5)] + \mathbb{P}[(x_0,1,5,0,6)] \\
&= [(P^G_{aA})^6(P^G_{bB})^5 P^G_{bA}] + 25[(P^G_{aA})^5 P^G_{aB}(P^G_{bB})^4(P^G_{bA})^2] \\
&+ 100[(P^G_{aA})^4(P^G_{aB})^2(P^G_{bB})^3(P^G_{bA})^3] \\
&+ 100[(P^G_{aA})^3(P^G_{aB})^3(P^G_{bB})^2(P^G_{bA})^4] + 25[(P^G_{aA})^2(P^G_{aB})^4 P^G_{bB}(P^G_{bA})^5] \\
&+ [P^G_{aA}(P^G_{aB})^5(P^G_{bA})^6].
\end{aligned}$$

When the final score is $(7,6)$, we have the two cases:

$$\mathbb{P}_a[(7,6)] = \mathbb{P}_a[A \text{ wins the tiebreak}|(6,6)]\cdot\mathbb{P}_a[(6,6)]$$

and

$$\mathbb{P}_b[(7,6)] = \mathbb{P}_b[A \text{ wins the tiebreak}|(6,6)]\cdot\mathbb{P}_b[(6,6)]$$

We obtain

$$\begin{aligned}\mathbb{P}_a[(6,6)] = \mathbb{P}_b[(6,6)] &= \mathbb{P}[(x_0,6,0,6,0)] + 26\cdot\mathbb{P}[(x_0,5,1,5,1)]\\ &+ 125\cdot\mathbb{P}[(x_0,4,2,4,2)] + 200\cdot\mathbb{P}[(x_0,3,3,3,3)]\\ &+ 125\cdot\mathbb{P}[(x_0,2,4,2,4)] + 26\cdot\mathbb{P}[(x_0,1,5,1,5)] + \mathbb{P}[(x_0,0,6,0,6)]\\ &= (P^G_{aA})^6(P^G_{bB})^6 + 26[(P^G_{aA})^5P^G_{aB}(P^G_{bB})^5P^G_{bA}]\\ &+ 125[(P^G_{aA})^4(P^G_{aB})^2(P^G_{bB})^4(P^G_{bA})^2] + 200[(P^G_{aA})^3(P^G_{aB})^3(P^G_{bB})^3(P^G_{bA})^3]\\ &+ 250[(P^G_{aA})^2(P^G_{aB})^4(P^G_{bB})^2(P^G_{bA})^4] + 26[P^G_{aA}(P^G_{aB})^5P^G_{bB}(P^G_{bA})^5]\\ &+ (P^G_{aB})^6(P^G_{bA})^6\end{aligned}$$

while $\mathbb{P}_a[A$ wins the tiebreak$|(6,6)]$ and $\mathbb{P}_b[A$ wins the tiebreak$|(6,6)]$ can be found in Newton and Keller (2005), pages 266-268. To obtain the remaining cases, let us consider a set which ends with the score (β,α), with $\beta<\alpha$. Denoting

$$\mathbb{P}_a[(\alpha,\beta)] = S_{\alpha,\beta}(P^G_{aA},P^G_{aB},P^G_{bB},P^G_{bA})$$
$$\mathbb{P}_b[(\alpha,\beta)] = T_{\alpha,\beta}(P^G_{aA},P^G_{aB},P^G_{bB},P^G_{bA})$$

it is easy to see that

$$\mathbb{P}_a[(\beta,\alpha)] = S_{\alpha,\beta}(P^G_{aB},P^G_{aA},P^G_{bA},P^G_{bB}).$$
$$\mathbb{P}_b[(\beta,\alpha)] = T_{\alpha,\beta}(P^G_{aB},P^G_{aA},P^G_{bA},P^G_{bB})$$

Note that the functions S and T depend on the final score.

B. Appendix

The conditional expectations in (1) are:

$$\mathbb{E}_a[D|(6,0)]\cdot\mathbb{P}_a[(A,B)=(6,0)] = \mathbb{P}[3,0,0,3](3k^G_{aA}+3k^G_{bA})$$

$$\begin{aligned}\mathbb{E}_a[D|(6,1)]\cdot\mathbb{P}_a[(A,B)=(6,1)] &= 3\mathbb{P}[4,0,1,2](4k^G_{aA}+k^G_{bB}+2k^G_{bA})+\\ &+ 3\mathbb{P}[3,1,0,3](3k^G_{aA}+k^G_{aB}+3k^G_{bA})\end{aligned}$$

$$\begin{aligned}\mathbb{E}_a[D|(6,2)]\cdot\mathbb{P}_a[(A,B)=(6,2)] &= 3\mathbb{P}[4,0,2,2](4k^G_{aA}+2k^G_{bB}+2k^G_{bA})\\ &+ 12\mathbb{P}[3,1,1,3](3k^G_{aA}+k^G_{aB}+k^G_{bB}+3k^G_{bA})\\ &+ 6\mathbb{P}[2,2,0,4](2k^G_{aA}+2k^G_{aB}+4k^G_{bA})\end{aligned}$$

$$\begin{aligned}\mathbb{E}_a[D|(6,3)]\cdot\mathbb{P}_a[(A,B)=(6,3)] &= 4\mathbb{P}[5,0,3,1](5k^G_{aA}+3k^G_{bB}+k^G_{bA})\\ &+ 24\mathbb{P}[4,1,2,2](4k^G_{aA}+k^G_{aB}+2k^G_{bB}+2k^G_{bA})\\ &+ 24\mathbb{P}[3,2,1,3](3k^G_{aA}+2k^G_{aB}+k^G_{bB}+3k^G_{bA})\\ &+ 4\mathbb{P}[2,3,0,4](2k^G_{aA}+3k^G_{aB}+4k^G_{bA})\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}_a[D|(6,4)]\cdot\mathbb{P}_a[(A,B)=(6,4)] = {} & \mathbb{P}[5,0,4,1](5k^G_{aA}+4k^G_{bB}+k^G_{bA}) \\
& + 20\mathbb{P}[4,1,3,2](4k^G_{aA}+k^G_{aB}+3k^G_{bB}+2k^G_{bA}) \\
& + 60\mathbb{P}[3,2,2,3](3k^G_{aA}+2k^G_{aB}+2k^G_{bB}+3k^G_{bA}) \\
& + 40\mathbb{P}[2,3,1,4](2k^G_{aA}+3k^G_{aB}+k^G_{bB}+4k^G_{bA}) \\
& + 5\mathbb{P}[1,4,0,5](k^G_{aA}+4k^G_{aB}+5k^G_{bA})
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_a[D|(7,5)]\cdot\mathbb{P}_a[(A,B)=(7,5)] = {} & \mathbb{P}[6,0,5,1](6k^G_{aA}+5k^G_{bB}+k^G_{bA}) \\
& + 25\mathbb{P}[5,1,4,2](5k^G_{aA}+k^G_{aB}+4k^G_{bB}+2k^G_{bA}) \\
& + 100\mathbb{P}[4,2,3,3](4k^G_{aA}+2k^G_{aB}+3k^G_{bB}+3k^G_{bA}) \\
& + 100\mathbb{P}[3,3,2,4](3k^G_{aA}+3k^G_{aB}+2k^G_{bB}+4k^G_{bA}) \\
& + 25\mathbb{P}[2,4,1,5](2k^G_{aA}+4k^G_{aB}+k^G_{bB}+5k^G_{bA}) \\
& + \mathbb{P}[1,5,0,6](k^G_{aA}+5k^G_{aB}+6k^G_{bA})
\end{aligned}
$$

Note that $\mathbb{E}_a[D|(i,j)]$ can be obtained by the previous formulas dividing by $\mathbb{P}_a[(A,B)=(i,j)]$ and that we have indeed evaluated the conditional expected durations given all the possible values $(x_1,\ldots,x_4)$.

The case $(7,6)$ or $(6,7)$ is more delicate, since we are not able to evaluate the mean duration of a tiebreak won by A or by B, separately. We have

$$
\begin{aligned}
\big(\mathbb{E}_a[D|(7,6)]+\mathbb{E}_a[D|(6,7)]\big)\cdot\mathbb{P}_a[(A,B)=(6,6)] = {} & \\
= \mathbb{P}_a[6,0,6,0](6k^G_{aA}+6k^G_{bB}+k^T_a) & \\
+ 26\mathbb{P}_a[5,1,5,1](5k^G_{aA}+k^G_{aB}+5k^G_{bB}+k^G_{bA}+k^T_a) & \\
+ 125\mathbb{P}_a[4,2,4,2](4k^G_{aA}+2k^G_{aB}+4k^G_{bB}+2k^G_{bA}+k^T_a) & \\
+ 200\mathbb{P}_a[3,3,3,3](3k^G_{aA}+3k^G_{aB}+3k^G_{bB}+3k^G_{bA}+k^T_a) & \\
+ 125\mathbb{P}_a[2,4,2,4](2k^G_{aA}+4k^G_{aB}+2k^G_{bB}+4k^G_{bA}+k^T_a) & \\
+ 26\mathbb{P}_a[1,5,1,5](k^G_{aA}+5k^G_{aB}+k^G_{bB}+5k^G_{bA}+k^T_a) & \\
+ \mathbb{P}_a[0,6,0,6](6k^G_{aB}+6k^G_{bA}+k^T_a)\ , &
\end{aligned}
$$

where k^T_a is the expected length of a tiebreak when A starts serving. Note that in this case it is possible to derive the expected duration of a set ending with the scores $(0,6)$,$(1,6)$,…,$(5,7)$ from the previous calculation. Indeed, if

$$\mathbb{E}_a[A,B]\cdot\mathbb{P}_a[A,B] = H(P^G_{aA},P^G_{aB}.P^G_{bB},P^G_{bA},k^G_{aA},k^G_{aB},k^G_{bB},k^G_{bA})$$

we have

$$\mathbb{E}_a[B,A]\cdot\mathbb{P}_a[B,A] = H(P^G_{aB},P^G_{aA}.P^G_{bA},P^G_{bB},k^G_{aB},k^G_{aA},k^G_{bA},k^G_{bB}).$$

References

Barnett, T., Brown, A., and Pollard, G. (2006). Reducing the likelihood of long tennis matches. *Journal of Sports Science and Medicine*, 5(4):567–574.

Carter Jr, W. H. and Crews, S. L. (1974). An analysis of the game of tennis. *The American Statistician*, 28(4):130–134.

Ferrante, M. and Fonseca, G. (2014). On the winning probabilities and mean durations of volleyball. *Journal of Quantitative Analysis in Sports*, 10(2):91–98.

Kemeny, J. G. and Snell, J. L. (1976). Finite markov chains, undergraduate texts in mathematics.

Klaassen, F. and Magnus, J. R. (2014). *Analyzing Wimbledon: The power of statistics.* Oxford University Press, USA.

Klaassen, F. J. and Magnus, J. R. (2001). Are points in tennis independent and identically distributed? evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, 96(454):500–509.

Newton, P. K. and Keller, J. B. (2005). Probability of winning at tennis i. theory and data. *Studies in applied Mathematics*, 114(3):241–269.

Norris, J. R. (1998). *Markov chains.* Number 2. Cambridge university press.