A new markovian model for tennis matches

Andrea Carrari(1), Marco Ferrante (1), Giovanni Fonseca (2)

(1) Dipartimento di Matematica "Tullio Levi-Civita", Università di Padova
(2) Dipartimento di Scienze Economiche e Statistiche, Università di Udine

Sevilla, 10 December 2016

DIPARTIMENTO MATEMATICA

UNIVERSITÀ DEGLI STUDI DI PADOVA

Advertising

Upcoming conference:

MathSport International 2017, Padua (Italy)
26-28 June 2017

URL: www.mathsportinternational2017.math.unipd.it

Advertising (2)

Deadlines:

- Abstract submission: 30 January 2017
- Latest notification of acceptance: 24 February 2017
- Submission of papers for Proceedings: 24 April 2017
- Early registration: 15 May 2017

Advertising

Keynote speakers:

- Michael Trick, Carnegie Mellon University (Sports scheduling)
- Fabrizio Renzi, IBM (Sports Analytics)
- Nicola Parolini, Politecnico di Milano (Numerics in Sports)
- ...

Outline

Tennis is a sport that can be nicely described with a simple mathematical model.

Tennis is a sport that can be nicely described with a simple mathematical model.
Assuming that the probability that a player on service 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** that forms a sort of Chinese Box (see e.g the recent book "Analyzing Wimbledon" by Klaassen and Magnus).

However, the assumption that the probability to win any point is **independent** of the previous points played and **identically distributed** along the match, has been criticised by some authors.

However, the assumption that the probability to win any point is **independent** of the previous points played and **identically distributed** along the match, has been criticised by some authors. To partially overcome this problem, we propose a simple modification of the model at the game's level.

However, the assumption that the probability to win any point is **independent** of the previous points played and **identically distributed** along the match, has been criticised by some authors.
To partially overcome this problem, we propose a simple modification of the model at the game's level.
We will assume that during any game there are two different situations:

- the initial points
- the, possible, additional points played after the (30:30), score that in our model coincide with the "Deuce".

This modelling is suggested by real data, where we see that the winning probability of the serving player decreases consistently after the (30,30) score, when he/she is just two points far from winning or loosing the game.

This modelling is suggested by real data, where we see that the winning probability of the serving player decreases consistently after the (30,30) score, when he/she is just two points far from winning or loosing the game.
In the present talk we will apply this model to evaluate the winning probabilities and the expected number of points played in a game and in a set.

This modelling is suggested by real data, where we see that the winning probability of the serving player decreases consistently after the (30,30) score, when he/she is just two points far from winning or loosing the game.

In the present talk we will apply this model to evaluate the winning probabilities and the expected number of points played in a game and in a set.

To perform this computation we will need some additional ingredients not considered before in the literature.

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).

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 length is motivated by the need to compute the expected number of **points** played in a set, where the exact length of the previous four types of games is needed.

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 length is motivated by the need to compute the expected number of **points** played in a set, where the exact length of the previous four types of games is needed.
All the previous results in the literature, concerning the duration of a tennis set, consider the (expected) number of games needed to complete a set, which is not enough to determine the exact (expected) number of points played.

We will assume that the probability to win any point by the player on service depend only on the present score.

We will assume that the probability to win any point by the player on service depend only on the present score.
In this way the score of a game can be described by a Markov Chain whose state space can be represented by

$$\{(0,0),(15,0),(0,15),(30,0),(15,15),(0,30),(40,0),(30,15),(15,30),(0,40),(40,15),(15,40), Deuce, Adv_A, Adv_B, Win_A, Win_B\}$$

but for simplicity we will define the states, in the same order, as

$$S = \{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17\}.$$

The Game

In order to determine the transition probabilities $p_{i,j}$, $i, j \in S$, the usual assumption is that the probability p to win any point by the player on service is independent of the score and constant during the game.

In order to determine the transition probabilities $p_{i,j}, i, j \in S$, the usual assumption is that the probability p to win any point by the player on service is independent of the score and constant during the game.
As empirical data confirm, the estimated winning probabilities before and after the "Deuce" are very different.

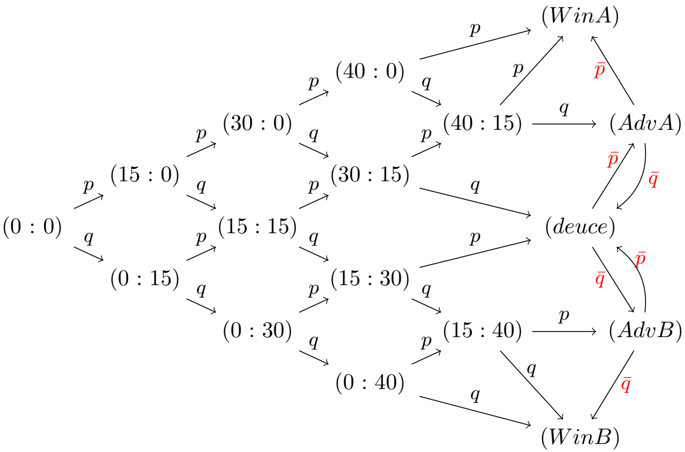
DIPARTIMENTO MATEMATICA
UNIVERSITÀ DEGLI STUDI DI PADOVA

In order to determine the transition probabilities $p_{i,j}, i, j \in S$, the usual assumption is that the probability p to win any point by the player on service is independent of the score and constant during the game.

As empirical data confirm, the estimated winning probabilities before and after the "Deuce" are very different.

For this reason, we will consider a second parameter $\bar{p}$, that will be the probability to win a point after the Deuce by the player on service and, to avoid trivial cases, we will assume that both p and $\bar{p}$ belong to $(0, 1)$.

Graph of a game



Winning probabilities

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 as the minimal, non negative solution of

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

Winning probabilities

DIPARTIMENTO MATEMATICA
UNIVERSITÀ DEGLI STUDI DI PADOVA

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 as the minimal, non negative solution of

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

The solution can be easily calculated and we obtain that

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

Remark

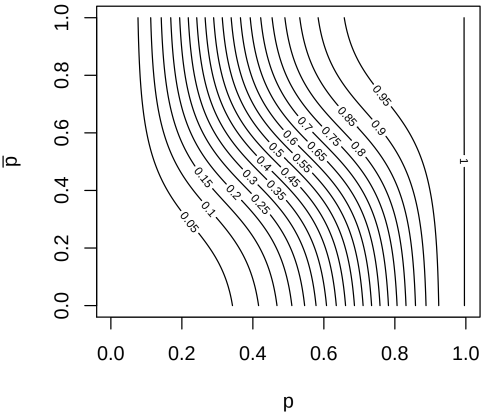
Denoting by $G(p, \bar{p}) := h_1$, by A and B the two players, and by P^G_{xY} the probability that the player Y wins a game when X serves, we obtain that:

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

Table: 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

Winning probabilities



Expected length

Since in the next section we will need to know the expected length of a game won by the player on service 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.

Expected length

Since in the next section we will need to know the expected length of a game won by the player on service 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.
The computation can be easily performed defining the **conditioned Markov chain** given to the event "The player on service wins the game".

Since in the next section we will need to know the expected length of a game won by the player on service 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.
The computation can be easily performed defining the **conditioned Markov chain** given to the event "The player on service wins the game".
Its transition matrix P' on the state space $\{1, \ldots, 16\}$ is given by:

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

Expected length

In order to compute this (conditional) expected duration, it will be sufficient to solve the linear system:

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

In order to compute this (conditional) expected duration, it will be sufficient to solve the linear system:

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

A simple computation gives

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

DIPARTIMENTO MATEMATICA

UNIVERSITÀ DEGLI STUDI DI PADOVA

Remark

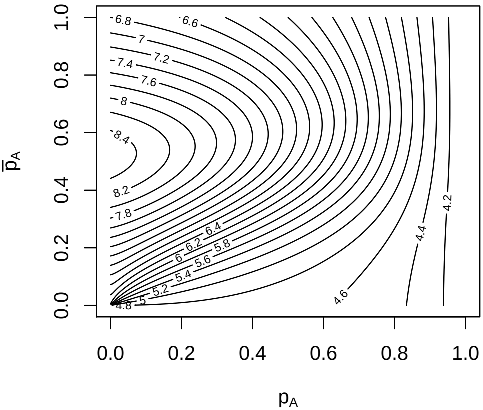
Defining $F(p,\bar{p}) := k_1$ and using the same notation as before, we get the expected length of the four types of outcomes of a game

$$k_{aA}^{G} = F(p_A, \bar{p}_A) \quad , \quad k_{aB}^{G} = F(1-p_A, 1-\bar{p}_A)$$
$$k_{bB}^{G} = F(p_B, \bar{p}_B) \quad , \quad k_{bA}^{G} = F(1-p_B, 1-\bar{p}_B).$$

Table: 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	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

Expected length



Remark

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.

Real data

We consider the matches between Rafael Nadal, Novak Djokovic and Roger Federer in the period 2009–2014.

Table: 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

Data are obtained from *www.tennis.earth.com*

First, we estimate the probabilities of winning a point after serving for each of the three players, considering also the possible opponent ability. Therefore, 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. Remember also that $P^G_{aB} = 1 - P^G_{aA}$.

First, we estimate the probabilities of winning a point after serving for each of the three players, considering also the possible opponent ability. Therefore, 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. Remember also that $P^G_{aB} = 1 - P^G_{aA}$.

Table: 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 model (k_1) and compare it with the mean duration of a game calculated on the played games recorded in the data ($\hat{k}_1$). The two parameter model results are generally closer to the empirical evidence, even if this is not true all the times.

Similarly, we estimate the expected length of the game using the proposed model (k_1) and compare it with the mean duration of a game calculated on the played games recorded in the data ($\hat{k}_1$). The two parameter model results are generally closer to the empirical evidence, even if this is not true all the times.

Table: 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

Remark

At the end, the assumption of a changing point probability depending on the score seems to be a good description of what is happening in a real tennis match: the probability of winning a game determined by the model is almost all the times closer to the empirical estimates and the average length estimate is most of the times closer than the one calculated with a constant point probability.

The Set

Let us now consider a set of tennis which may end also with a tiebreak. In this case we can 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 derive the results for the remaining cases easily.

The Set

Let us now consider a set of tennis which may end also with a tiebreak. In this case we can 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 derive the results for the remaining cases easily.
To simplify the notation, 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.

The Set

Let us now consider a set of tennis which may end also with a tiebreak. In this case we can 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 derive the results for the remaining cases easily.

To simplify the notation, 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.

Let us start with the Tiebreak.

The 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 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.

The 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, with **53** states.

The Tiebreak

In Newton and Keller (2005), the winning probabilities of the tiebreak, that we do not report here, are calculated by a recursive approach.

DIPARTIMENTO MATEMATICA
UNIVERSITÀ DEGLI STUDI DI PADOVA

In Newton and Keller (2005), 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. The expected number of points played to complete the tiebreak is equal to

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

Winning probability

Let us now consider the whole set. The computation of the winning probabilities is a little bit tedious, but with a compact notation it will be possible at least to write down the formulas in an efficient way.

Winning probability

Let us now consider the whole set. The computation of the winning probabilities is a little bit tedious, but with a compact notation it will be possible at least to write down the formulas in an efficient way.

The final score of the set will described by a five dimensional random vector X, where $X = (x_0, x_1, x_2, x_3, x_4)$ summarizes who starts serving (0 for A and 1 for B) and the number of aA, aB, bB and bA games, respectively.

Winning probability

Let us now consider the whole set. The computation of the winning probabilities is a little bit tedious, but with a compact notation it will be possible at least to write down the formulas in an efficient way.

The final score of the set will described by a five dimensional random vector X, where $X = (x_0, x_1, x_2, x_3, x_4)$ 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.

Winning probability

We get

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

and with a more compact notation

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

and so on for all the other possible scores.

Expected length

Expected length

Let D be the random length 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.

Expected length

Let D be the random length 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.
Denoting $\mathbb{E}_a[D|(i,j)] = \mathbb{E}_a[D|(A,B)=(i,j)]$, we have:

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

where:

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

and so on.

Expected length

Table: Expected (conditional) duration of a set that ends (6,4)

p_B \ p_A	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7
0.3	58.48	60.22	61.77	63.01	63.84	64.21	64.10	63.55	62.63
0.35	60.02	61.79	63.35	64.58	65.36	65.64	65.41	64.71	63.64
0.4	61.36	63.14	64.69	65.86	66.55	66.70	66.34	65.50	64.28
0.45	62.38	64.15	65.64	66.72	67.29	67.31	66.80	65.82	64.47
0.5	62.98	64.70	66.10	67.06	67.50	67.39	66.74	65.63	64.17
0.55	63.12	64.75	66.03	66.86	67.16	66.91	66.14	64.92	63.36
0.6	62.79	64.30	65.45	66.13	66.29	65.92	65.03	63.72	62.11
0.65	62.03	63.40	64.40	64.94	64.97	64.49	63.51	62.14	60.50
0.7	60.90	62.12	62.97	63.38	63.31	62.73	61.69	60.30	58.68

Djokovic vs. Federer

DIPARTIMENTO MATEMATICA
Università degli Studi di Padova

	$score_A$	$score_B$	freq	$k_{empirical}$	$k_{(p,p)}$	$k_{(p,\bar{p})}$
1	6	0	1	35.00	38.85	38.69
2	6	1	2	43.50	44.76	44.60
3	6	2	1	40.00	50.71	50.58
4	6	3	4	53.75	56.71	56.61
5	6	4	6	65.67	62.82	62.81
6	7	5	4	73.50	75.45	75.50
7	0	6	3	37.00	38.65	39.53
8	1	6	1	36.00	44.59	45.29
9	2	6	7	48.43	50.57	51.13
10	3	6	9	54.81	56.60	57.03
11	4	6	2	64.61	62.76	63.09
12	5	7	4	75.30	75.38	75.78
13	6	6	7	82.92	88.14	88.22

Table: Expected length: Federer(A) - Djokovic(B)

Djokovic vs. Federer 2015

Wimbledon Final on 12 July 2015

Wimbledon Final on 12 July 2015

- First set: (7,6) with 74 points played (estimate 88.22);
- Second set (6,7) with 102 points played (estimate 88.22);
- Third set (6,4) with 55 points played (estimate 63.09);
- Fourth set (6-3) with 55 points played (estimate 57.03).

Wimbledon Final on 12 July 2015

- First set: (7,6) with 74 points played (estimate 88.22);
- Second set (6,7) with 102 points played (estimate 88.22);
- Third set (6,4) with 55 points played (estimate 63.09);
- Fourth set (6-3) with 55 points played (estimate 57.03).

Note that

$$\frac{74 + 102}{2} = 88$$

Thank you!