Tennis Analytics: a(n incomplete) Review

Presentation · November 2017

1 author:

Francesco Lisi
University of Padova

Tennis Analytics: a(n incomplete) Review

Francesco Lisi

Department of Statistical Sciences, University of Padua

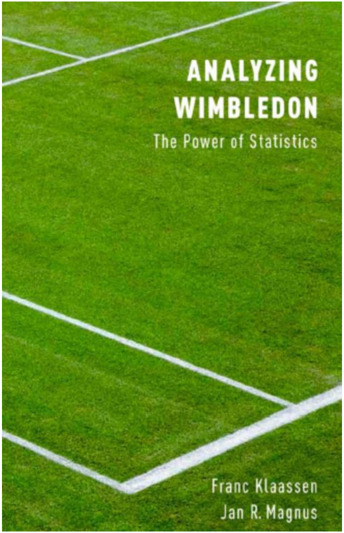
BDSports: Big Data Analytics in Sports
16 November, 2017, Brescia

Some topics in literature

1. Probabilistic modeling and simulations of tennis matches
2. Analysis of play and features of tennis matches
3. Predictive models for outcomes of tennis matches
4. Betting
5. Men-Women and/or Top-Medium players differences
6. Performance analysis
7. Efficient tournaments management
8. Ranking and scoring systems
9. Physiological aspects and injuries prevention
10. Equipment performance (rackets, balls, hawk-eye, etc..)
11. Frauds detection

2. Analysis of play

2.1 Are all points (games, sets) equally important?
2.2 Are points i.i.d?
2.3 Does the prob. of being broken back increase after a break?
2.4 Is it an advantage to serve first in a set?
2.5 Is it true that the 7th game in a set is the most important?
2.6 Is there a home advantage in tennis?
2.7 How "quality" of players can be defined?
2.8 It is true that a player is as good as his/her second service?
2.9 Do players have an efficient service strategy?
2.10 Is it true that players play safer at important points?
2.11 Do players take more risks when they are in a winning mood?
2.12 Are top players more stable than others?
2.13 Are new balls an advantage for the server?



Klaassen and Magnus (2014). Analyzing Wimbledon. The Power of Statistics, Oxford University Press

The tennis scoring system

- A tennis match is composed of sets, games and points.
- **match**: won by the player who first wins 2 sets in a "best of three" match (o 3 in a "best of five" match)
- **set**: a sequence of games played with service alternating between games. To win a set, a player must win at least 6 games, with a margin of at least 2 games over the opponent.
  - **advantage set**: at 6-6 the set continues until a player wins two consecutive games;
  - **tie-break set**: a seven-point game concludes the set.
- **game**: a sequence of points played with the same player at service. To win a game a player must win at least 4 points, with a margin of at least 2 points over the opponent.
- **points**: "15", "30", "40" and "game".
- The 40-40 score is called "**deuce**". After a deuce the game is won by the player who first wins two consecutive points.
- The first point after a deuce is called "**advantage**".

1. Probabilistic tennis

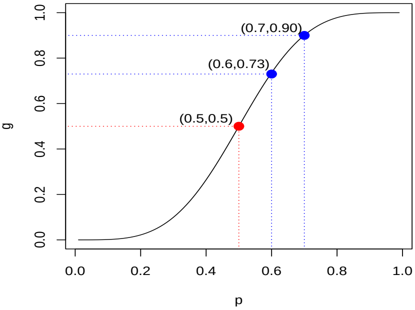
- Tennis scoring can be described using techniques based on counting paths, binomial theorem, Markov Chains theory, backward and forward recursions (Newton and Keller, 2005; Newton and Aslam, 2009; Barnett and Brown, 2012).
- Consider two players A and B.

$$p_i = Prob(\text{Player } i \text{ wins a point at service}) \qquad i = A, B$$

- Under the *i.i.d. assumption* the probability for a player of winning a game *at service* is

$$g_i = P(\text{Player wins a game}) = \frac{p_i^4(-8p_i^3 + 28p_i^2 - 34p_i + 15)}{p_i^2 + (1 - p_i)^2}$$

1. Probability of winning a game (g_i)



- $p = 0.5$ implies $g = 0.5$
- $p = 0.6$ implies $g = 0.73$, and $p = 0.7$ implies $g = 0.90$.
- Magnification effect: for $p > 0.5$, $g > p$: an advantage at point level leads to a bigger advantage at game level.
- S-shaped curve: the magnification effect is large when p is close to 0.5 and small when p is close to 1.

1. Probability of winning a set (s_i)

- From g_i, using recursive formulas, we can obtain:

$$s_i = P(\text{Player } i \text{ wins a set}) \qquad i = A, B$$

- For s_i we need both p_A and p_B (we need g_A and g_B). Each couple of (p_A, p_B) leads to a different probability s_i .
- Two new parameters:

$$\delta = p_A - p_B, \qquad \gamma = p_A + p_B$$

- p_A depends on how well A serves ($serv_A$) and on how well B returns (rec_B)

$$p_A = serv_A - rec_B, \qquad p_B = serv_B - rec_A$$

and this implies

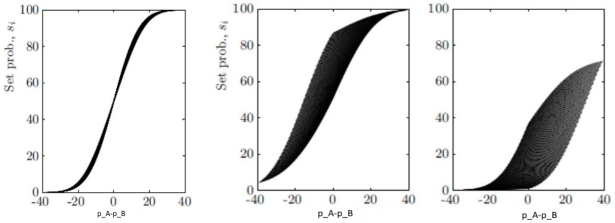
$$\begin{aligned} \delta = p_A - p_B &= (serv_A + rec_A) - (serv_B + rec_B) \\ \gamma = p_A + p_B &= (serv_A - rec_A) - (serv_B - rec_B) \end{aligned}$$

1. Probability of winning a set (s_i)

- $\delta = p_A - p_B =$ quality difference between the two players
- $\gamma = p_A + p_B =$ sum of serve-return difference for both players
- Two main advantages:
  - p_A and p_B are related to each other; δ and γ are much less related.
  - s_A (which is what interests us here) depends primarily on δ.

1. Probability of winning a set (s_i)

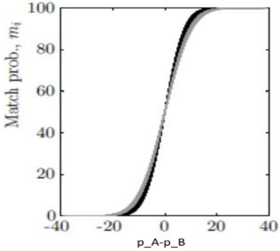
- Consider s_i as a function of $\delta = p_A - p_B$ and $\gamma = p_A + p_B$ (figures from Klaassen and Magnus, 2014)



- First figure shows s_i at the beginning of the set as a function of δ for different values of γ , (from 0.8 to 1.6).
  The collection of all curves for $0.8 < \gamma < 1.6$ gives the thickness of the curve.
- Is this also true during a set? At equal scores, the answer is yes. At unequal scores, however, the dependence on γ is stronger, and the two most extreme cases occur at 5-4 and 4-5.

1. Probability of winning a match (m_i)

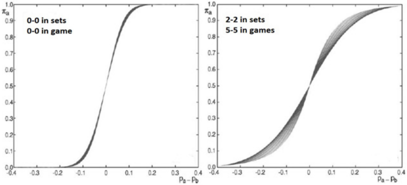
- From s_i, by recursive formulas, we can find m_i, the probability that player i wins the match ($i = A, B$).
- Probability of winning a match best-of-five (dark) versus best-of-three (light), *at the beginning of the match*



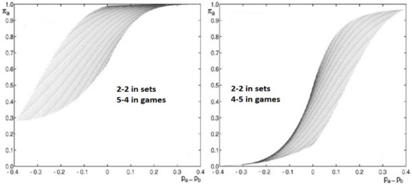
- The best-of-five curve is always above the best-of-three curve when $p_A > p_B$ (and below when $p_A < p_B$). It is another example of the magnification effect.

1. Probability of winning a match (m_i)

- At *equal scores*:
  m_i depends almost entirely on δ and only marginally on γ.
  Dependence on γ increases towards the end of the match.



- At *unequal scores*:
  dependence on γ is much stronger.



1. Probability of winning a match (m_i)

- Summarizing, the probability of winning a match is:

$$m_i = w(p_A, p_B, \text{tournament}, \text{current score}, \text{current server})$$

- m_i changes at different scores (also for p_A and p_B constant)

1. Probabilities estimation

- Usually, p_i is estimated as the relative frequency of winning service points in a given sample.
- g_i, s_i and m_i $(i = A, B)$ can be estimated:
  - as the relative frequency, with respect to games, sets and matches or
  - starting from $\hat{p}_A$ and $\hat{p}_B$ and using previous formulas. In this case particular attention should be payed to the magnification effect.
- m_i can also be estimated using bookmakers odds after suitable transformations.
- Estimation of p_i requires point-to-point data. Estimation of g_i, s_i and m_i can be based on match data.

2. Analysis of play

2.1 Are all points (games, sets) equally important?

2.2 Are points i.i.d?

2.3 Does the prob. of being broken back increase after a break?

2.4 Is it an advantage to serve first in a set?

2.5 Is it true that the 7th game in a set is the most important?

2.6 Is there a home advantage in tennis?

2.7 How “quality” of players can be defined?

2.8 It is true that a player is as good as his/her second service?

2.9 Do players have an efficient service strategy?

2.10 Is it true that players play safer at important points?

2.11 Do players take more risks when they are in a winning mood?

2.12 Are top players more stable than others?

2.13 Are new balls an advantage for the server?

2.1. Importance of points in tennis

- Intuitively, a point is important if winning or losing that point has a large impact on winning the game, the set, the match.
- In a game, for players A and B consider the score (a, b). Scores a and b can be 0, 15, 30, 40, and Advantage (Ad) but, for convenience, we count them as 0, 1, 2, 3, 4(=Adv). Define

$$g_i(a, b) = P\,(i \text{ wins the game from point (a,b) at service})$$

- Morris (1977) defines the **importance** of point (a, b) in a game as

$$imp_{pg}(a, b) = g_i(a+1, b) - g_i(a, b+1)$$

  This is Prob(server (i) wins the game | wins the point) - Prob(server (i) wins the game | loses the point).
- It can be shown that the importance of point (a, b) is the same for A and B.

2.1. Important points in a game ($imp_{pg}(a, b)$)

- To quantify the importance, we need p_i the probability that player i wins the point at service.
- We use $p_A = p_B = 64\%$ which is the average value in the 2010 grand slams for men (Klaassen and Magnus, 2014).

		Point score B (receiver)				
		0	15	30	40	Adv
	0	23	33	40	31	–
	15	17	30	45	49	–
Point score A	30	10	21	43	76	–
(server)	40	3	9	24	43	76
	Adv	–	–	–	24	–

- For $p > 0.5$, $30 - 40$ (break-point) is more important than $40 - 30$ (game-point).

2.1. Important games in a set ($imp_{gs}(a, b)$)

- The importance of the game(a,b) is defined as

$$imp_{gs}(a, b) = s_i(a + 1, b) - s_i(a, b + 1) \tag{1}$$

- Assuming $p_A = p_B = 64\%$ implies $g_A = g_B = 81.3\%$
- For a tie-break set we have:

		Game score B						
		0	1	2	3	4	5	6
	0	30	30	18	14	3	1	–
	1	30	33	33	17	12	1	–
Game score A	2	30	33	37	37	14	8	–
	3	14	32	37	42	42	9	–
	4	10	12	36	42	50	50	–
	5	1	6	8	41	50	50	50
	6	–	–	–	–	–	50	100

2.1. Important sets in a match ($imp_{sm}(a, b)$)

- To complete the three-step analysis (point-game, game-set, set-match), we consider the importance of a set in a match.
- The importance of the set(a,b) is defined as

$$imp_{sm}(a, b) = m_i(a + 1, b) - m_i(a, b + 1) \tag{2}$$

- Assume A and B have equal strength, so that $m_i = 50\%$ for both. In a best-of-five match the importance imp_{sm} is

		Set score B		
		0	1	2
	0	38	38	25
Set score A	1	38	50	50
	2	25	50	100

2.1. Important points in a match ($imp_{pm}(a, b)$)

- Previous definitions (imp_{pg}, imp_{gs} and imp_{sm}), allow us to write the importance of a point in a match (imp_{pm}) as

$$imp = imp_{pm} = imp_{pg} \times imp_{gs} \times imp_{sm}$$

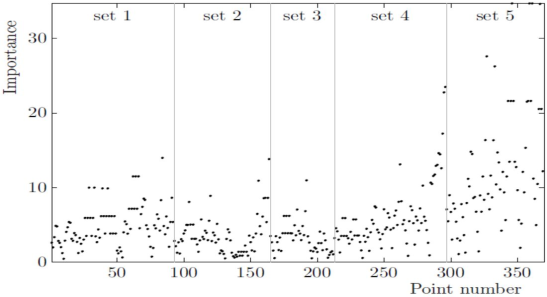


Figura: Importance profile of the match: Djokovic-Nadal, Australian Open 2012. Final score: 5-7, 6-4, 6-2, 6-7, 7-5. (from Klaassen and Magnus, 2014)

2.2. Are points in tennis i.i.d. distributed?

- **Independence**: p_i does not depend on what has happened previously (previous points).
- **Identical distribution**: p_i is not affected by the score, by the "state" of the match or by other conditions.
- Only few works consider specifically the i.i.d issue (Klaassen and Magnus, 2001; Pollard and Pollard, 2012; Matteazzi and Lisi, 2017)
- Other related issues, i.e the effect of:
  break points (Knight and O'Donoghue, 2012);
  tie-break points (Pollard, 1983);
  other "special" points in a match (Pollard, 2004; Klaassen and Magnus, 2003; O'Donoghue and Ingram, 2001; Morris, 1977)
  momentum (Jackson and Mosurski, 1997)
  winning/losing mood (Terry and Munro, 2008; Klaassen and Magnus, 2014).

2.2. Are points in tennis i.i.d. distributed?

- Klaassen and Magnus (2001) test the iid hypothesis by means of a dynamic model for binary panel data with random effects.
- Let $y_i = 1$ if player A wins the i-th service point (against player B) and 0 otherwise. The K-M model is

$$y_i = Q + D_{i-1} + \epsilon_i \qquad (3)$$

  where:
  - Q is a "quality" component, expressing the difference between A and B;
  - D_i is a dynamic term depending on all match information (of both players) up to point $i-1$;
  - ϵ_i is a random binary term.
- If points were iid, then such match information would be useless in predicting y_i .

2.2. Are points in tennis i.i.d. distributed?

- K-M estimate the model (by FGLS) on a sample of $86,000$ points ($57,000$ for men$+29,000$ for women) played at Wimbledon, and find that:
  - winning the previous point has a positive effect on winning the current point;
  - it is more difficult for the player at service to win "important" points than less important points;
  - the weaker a player is, the stronger are these effects.
- K-M conclude that points are not i.i.d., but they also observe that deviations from i.i.d are small and hence the i.i.d. hypothesis provides a good approximation in many cases.

2.2. Are points in tennis i.i.d. distributed?

- Pollard and Pollard (2012) use several measures to check if (real) points at service are consistent with the i.i.d. hypothesis.
- P-P analyse the probability that a player wins a point at service:
  (a) if he is ahead, equal or behind in the game;
  (b) if he has won or lost the previous point;
  (c) given various combinations of (a) and (b);
  (d) given the importance of the point.
- They also consider:
  (e) the number of games won at service;
  (f) the duration (number of points) of a game;
  (g) the number of wins and losses.

2.2. Are points in tennis i.i.d. distributed?

- P-P test if the observed relative frequencies are consistent with those expected under the hypothesis of independent points.
- Tests are based on Chi-Squared with suitable df for (a)-(f) and on Wald-Wolfowitz two samples run test for (g)
- Tests are applied to 1172 points played by Nadal against 11 top players in 2011.
- P-P find that most tests accept the independence assumption, but they also find that Nadal can lift (i) when returning and being ahead in the game and (ii) when serving and having lost the previous point.
- P-P conclude that there are evidences of "non independence".

2.2. Are points in tennis i.i.d. distributed?

- Matteazzi and Lisi (2017) identify several "states" of the match where deviations from i.i.d can occur; they are:
  (a) game-points (break-, set- and match-points);
  (b) tie-break points;
  (c) various combinations of (a) and (b), i.e break-points and set-points;
  (d) the previous point.
- M-L test the i.i.d hypothesis vs specific deviations from i.i.d., comparing actual relative frequencies with those expected under the iid hypothesis by means of:
  - Logistic regression models (parametric)
  - Exact binomial tests (parametric)
  - Proportion tests (nonparametric)
  - Monte Carlo tests (nonparametric)

2.2. Are points in tennis i.i.d. distributed?

- Unlike previous works, analyses are run both for men's single and women's single matches and for different kinds of surfaces and championships.
- Men: 15 head-to-head match sequences, involving 150 matches, 21 players (high- and medium-ranked) and 26,196 points.
- Women: 8 head-to-head match sequences, involving 64 matches, 13 players (high- and medium-ranked) and 8,902 points.
- M-L find that i.i.d. is occasionally rejected but there is no clear evidence of significant deviations from the i.i.d. hypothesis.

2.2. Are points in tennis i.i.d. distributed?

- Overall, results are mixed: sometimes authors find evidence of non i.i.d. points, sometimes the evidence is very weak
- Since it is not always easy to retrieve exhaustive datasets, results may depend on the specific data considered (only a player or a couple of players, only a surface, only a tournament, etc...)
- However, we can conclude that, even if there is some evidence that points are not i.i.d., among professional players this assumption still provides a good approximation.

2.3. Breaks and re-breaks

- A break occurs when the server does not win his/her own service game.
  Some commentators believe that there is a higher probability of being broken back after a break. Is this statement supported by data?
- If the hypothesis is true, then the p_i decreases after a break in the previous game (in the same set).
- Percentage of won points at service after winning or losing the previous game for a sample of matches played at Wimbledon:

	Men	Women
All points	64.4	56.1
After break	65.8	57.9
After no-break	64.1	54.9

- Translating these % from points to games, the probability of winning a service game, when a break occurred in the previous game, increases of 3.3% -points for men d 5.7% for women.

2.3. Breaks and re-breaks

- It seems that the hypothesis is wrong.
- Can we believe it?
- The 'after break' winning probability is dominated by good players, who are able to react; while the 'after no-break' winning probability, is dominated by weaker players.
  The positive difference can be explained by quality heterogeneity, even when a break in the previous game does not matter at all. This is an example of sample selection bias.
- When a suitable correction for quality is considered (Klaassen and Magnus ...) no effect is found for men. Instead such an effect has been found for women
- Thus the hypothesis is rejected for men but accepted for women.

3. Predictive models for outcomes of tennis matches

- Prediction of the outcome of the match between players A and B.
- Two types of predictions:
  3.1 before the match (using statistical models or odds from bookmakers)
  3.2 during the match (using statistical models)
- Good reviews are Kovalchik (2016), in the first case, and Klaassen and Magnus (2014) in the second case.

3.1. Before-match tennis outcomes forecasting

Most models fall into four categories (Kovalchik, 2016):

- **Regression-based models**: Boulier and Stekler (1999), Clarke and Dyte (2000), Klaassen and Magnus (2003), Glisdorf and Sukhatme (2008), del Corral and Prieto-Rodríguez (2010), Lisi and Zanella (2016)
- **Point-based models**: (Newton and Keller, 2005), Barnett and Clarke (2005), Knottenbelt et al. (2012), Spanias and Knottenbelt (2012)
- **Paired comparison models**: McHale and Morton (2011), FiveThirtyEight (2015)
- **Bookmakers-based methods**: Leitner et al. (2009)

3.1. Before-match tennis outcomes forecasting

- m_{AB} = probability that player A wins against player B, with A the player with the higher ATP rank.
- **Regression-based models** directly model the probability that player A wins the match.
- Most of the times they use:
  - probit models, where: $m_{AB} = \Phi(\mathbf{x}'\beta)$
    with Φ c.d.f of a $N(0, 1)$ and $\mathbf{x}$ a vector of predictors;
  - logit (logistic) models, where: $log\left(\frac{m_{AB}}{1-m_{AB}}\right) = \mathbf{x}'\beta$
- Models usually differ for the different predictors.

3.1 Forecasting before-match tennis outcomes

- **Point-based models**: first, they specify p_A and p_B. Then, usually under the IID assumption, use algebraic formulas to obtain m_{AB}.
- The difference among works consists in the way the basic probability of winning a point or a game is computed.
- Newton and Keller (2005) consider an adjustment for opponent ability and compute the frequency of points won at service and returning in the previous 12 months.
- Barnett and Clarke (2005) propose a tournament-specific adjustment

3.1 Forecasting before-match tennis outcomes

- **Paired comparison models**: directly compare of two players based on their features and past history (head-to-head matches).
- McHale and Morton (2011) use the Bradley-Terry model and assume that each player has a latent match win ability, α_i. In the B-T formulation, the odds that player A beats player B, is α_A/α_B and the corresponding m_{AB} is:

$$m_{AB} = \frac{\alpha_A}{\alpha_A + \alpha_B}$$

- McHale and Morton estimate the abilities α_i maximizing the (log-)likelihood of games won by player A (g_A) and by player B (g_B), with an exponential decay function to give more weight to the more recent matches.

3.1 Forecasting before-match tennis outcomes

- **Bookmakers-based methods**: they use bookmakers' odds. Odds can be regarded as expert opinions about the probability that a player wins against a specific opponent.
- Leitner et al. (2009) propose to aggregate these expectations into an overall prediction for match outcomes. They call this model bookmakers consensus model (BCM).
- For a given match, let $o_{AB,j}$ (j=1,...,N) be N bookmakers odds. These can be easily translated into winning probabilities. Leitner et al. estimate the logit probability that the favorite player wins as

$$logit(\hat{\pi}_{AB}) = N^{-1} \sum_{j=1}^{N} logit(\hat{\pi}_{AB,j}). \tag{4}$$

3.1 Regressors

- Difference of rank (possibly transformed)
- Difference of ATP (WTA) points
- Age (or functions of the ages)
- Surface (clay, grass, hard, carpet)
- Home factor
- Bookmakers odds (possibly transformed)
- Previous results (head-to-head)
- Career (best ranking or top-ten)
- Momentum
- Tournament (Grand Slams, Masters 1000, ATP 500,...)
- Prizes

3.2 Forecasting within-match tennis outcomes

- In the within-match modeling the prediction has to be updated on a point-by-point basis throughout the match.
- Only few papers account for this kind of prediction (Klaassen and Magnus, 2003; Easton and Uylangco, 2010), (Paulden, 2016)
- We have seen that, for the probability of winning a match we have:

$$m_i = w(p_A, p_B, \text{tournament}, \text{current score}, \text{current server})$$

- Given these 'ingredients', m_i is computed recursively.
- Note that m_i is different at different scores even if p_A and p_B are constant

3.2 Forecasting within-match tennis outcomes

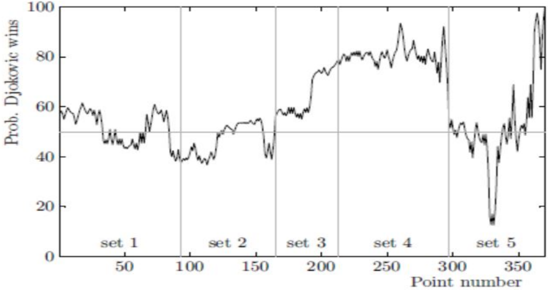
- In the within-match modeling m_i is updated on a point-by-point basis during the match.
- Estimation of p_A and p_B: due to the magnification effect, $\hat{p}_A$ and $\hat{p}_B$ differing just slightly from the true p_A and p_B can lead to a completely different m_i.
- Klaassen and Magnus (2003) suggest estimating p_A and p_B indirectly starting from the match probability m_i at the beginning of the match
- Estimation of m_i: using the rank of the two players or (better) using betting odds.

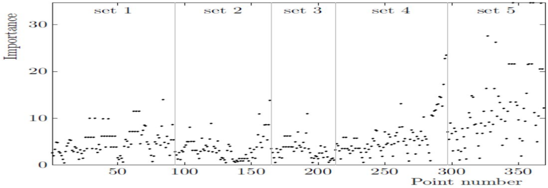
Within-match tennis outcomes forecasting

- At the beginning of the match m_i depends only marginally on $p_A + p_B$. Thus, the accuracy of its estimation is less crucial.
- Estimation of $p_A + p_B$: using $\widehat{p_A + p_B} = 2\bar{p}$, with $\bar{p}$ tournament average (over the players) probability of winning a point in a chosen year.
- Given (estimates of) m_i and $p_A + p_B$, 'inverting' function h we obtain $p_A - p_B$
- Once we have $p_A - p_B$ and $p_A + p_B$, we also have p_A and p_B and we can compute m_i at each score.

Within-match tennis outcomes forecasting

Djokovic-Nadal, Australian Open 2012 (from Klaassen and Magnus, 2014)





Tennis Data

- **ATP website** (www.atpworldtour.com)
- **WTA website** (www.wtatennis.com)
- **OnCourt** (www.oncourt.info/): a program and a tennis database containing results on tennis match played since 1990. It includes both match data and point-to-point data (in the paid version) and (some) odds.
- **Tennis-Data** (tennis-data.co.uk): a miner of match data and odds
- **Jeff Sackmann database** (https://github.com/JeffSackmann/): a large repository of free match data and point-to-point tennis data

References I

Barnett, T. and Brown, A. (2012). *The mathematics of tennis*. Strategic Games.

Barnett, T. and Clarke, S. R. (2005). Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics*, 16:113–120.

Boulier, B. and Stekler, H. O. (1999). Are sports seedings good predictors? an evaluation. *International Journal of Forecasting*, 15:83–91.

Clarke, S. and Dyte, D. (2000). Using offcial ratings to simulate major tennis tournaments. *International Transaction in Operational Research*, 7:585–594.

del Corral, J. and Prieto-Rodríguez, J. (2010). Are differences in ranks good predictors for grand slam tennis matches? *International Journal of Forecasting*, 26:551–563.

References II

Easton, S. and Uylangco, K. (2010). Forecasting outcoms in tennis matches using within-match betting markets. *International Journal of Forecasting*, 26:554–575.

FiveThirtyEight (2015). *Serena Williams and the difference between All-Time Great and the Greatest of All Time*. http://fivethirtyeight. com/features/serena-williams-and-the-difference-between-alltime-great-and-greatest-of-all-time/.

Glisdorf, K. and Sukhatme, V. (2008). Tennis betting: can statistics beat bookmakers? *Journal of Sport Economics*, 9:287–303.

Jackson, D. and Mosurski, K. (1997). Heavy defeats in tennis: Psychological momentum of random effect. *Chance*, 10:27–34.

References III

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

Klaassen, F. and Magnus, J. (2003). Forecasting the winner of a tennis match. *European Journal of Operational Research*, 148:257–267.

Klaassen, F. and Magnus, J. (2014). *Analyzing Wimbledon. The Power of Statistics*. Oxford University Press.

Knight, G. and O'Donoghue, P. (2012). The probability of winnning break points in grand slam men's singles tennis. *European Journal of Sport Science*, 12:462–468.

Knottenbelt, W. J., Spanias, D., and Madurska, A. M. (2012). A common- opponent stochastic model for predicting the outcome of professional tennis matches. *Computers & Mathematics with Applications*, 64:3820–3827.

References IV

Kovalchik, S. A. (2016). Searching for the goat of tennis win prediction. *Journal of Quantitative Analysis in Sports*, 12:127–138.

Leitner, C. A., Zeileis, A., and Hornik, K. (2009). Is Federer stronger in a tournament without Nadal? an evaluation of odds and seedings for Wimbledon 2009. *Research Report Series/Department of Statistics and Mathematics 94*.

Lisi, F. and Zanella, G. (2016). Tennis betting: can statistics beat bookmakers? *Electronic Journal of Applied Statistics*, Forthcoming.

Matteazzi, F. and Lisi, F. (2017). A follow-up study on the issue of i.i.d. points in tennis. In *Proceedings of the MathSport International 2017 Conference*.

McHale, I. and Morton, A. (2011). A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting*, 27:619–230.

References V

Morris, C. (1977). The most important points in tennis. *Optimal Strategies in Sports*, 5:131–140.

Newton, P. and Aslam, K. (2009). Monte carlo tennis: A stochastic markov chain model. *Journal of Quantitative Analysis in Sports*, 3.

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

O'Donoghue, P. and Ingram, B. (2001). A notational analysis of elite tennis strategy. *Journal of Sport Sciences*, 19:107–115.

Pollard, G. (1983). An analysis of classical and tie-break tennis. *Australian Journal of Statistics*, 25:496–505.

Pollard, G. (2004). Can a tennis player increase te probability of winning a point when it is important? In *Proceedings of the 7th Australasian Conference on Mathematics and Computers in Sport*, pages 253–256.

References VI

Pollard, G. and Pollard, G. (2012). Applying statistical tests for the independence of points in tennis. In *Proceedings of the 11th Australasian Conference on Maths and Computers in Sport*, pages 113–121. Mathsport,ANZIAM.

Spanias, D. and Knottenbelt, W. J. (2012). Predicting the outcomes of tennis matches using a low-level point model. *IMA Journal of Management Mathematics*, 24:311–320.

Terry, P. and Munro, A. (2008). Mood and anxiety scores predict winning and losing performances in tennis. In *43rd Australian Psychological Society Annual Conference (APS 2008)*, pages 23–27.