Tristan Barnett, Denny Meyer and Graham Pollard
Med Sci Tennis 2008;13(2):24-27

Applying match statistics to increase serving performance

Abstract

A large database of tennis statistics is used to calculate player match statistics for each court surface. Analysis is carried out to determine serving strategies to increase performance. The results of the analysis could be used by players and coaches to possibly increase serving performance.

Introduction

There is a range of published material on the use of match statistics/probabilities to increase serving performance. Gale[1] used a simple mathematical model to determine an optimal strategy for serving in tennis. Norman[2] used dynamic programming to determine an optimal strategy of whether to use a slow or fast serve on the first and second serve. George[3] used a simple probabilistic model to determine a serving strategy in tennis and stated that the usual serving strategy may not be optimal. Professional tennis matches were used as examples to support the claim. Gillman[4] developed a similar analysis to serving strategies. Hannan[5] also analysed different serving strategies, with the added complexity of the opponent returning the serve in such a way that the server countered with a strong shot or was forced to hit a weak shot. Pollard[6] determined a method for finding the optimal risks that should be taken by the server on the first and second serves.

In the above papers the effect of the receiver's receiving capacity on the probability that the server wins a point on the first serve and on the probability that he wins a point on the second serve, is typically handled implicitly. Nowadays, with the availability of substantial data on each player's receiving capacity (as well as on each player's serving capacity), it is possible to address the effect of the strength of the receiver on the server's first and second serve probabilities explicitly rather than just implicitly. Thus, a player's serving probabilities against one player can be different from his/her serving probabilities against another. Correspondingly, a player's best serving strategy against one player can be different from the best serving strategy against another. Barnett and Clarke[7] showed how to predict serving and receiving player statistics/probabilities prior to the start of a match. Their model can be used to evaluate the optimal serving strategy for a player against a particular opponent. Barnett and Pollard[2] showed that players' performances are affected by the court surface for both men and women. Thus, a player's optimal serving strategy can vary from opponent to opponent and from surface to surface. In earlier studies, such variations were handled implicitly, and were not clearly identifiable to the reader.

As an example of a player's optimal serving strategy varying from one opponent to another, we consider the following. Consider a typical professional man (player A) whose optimal serving strategy is typically a hard first serve and a softer second serve with spin. If this player's opponent is equally as good at receiving a first serve as receiving a second serve, it is clear that player A may just as well serve a second serve as a first serve. On the other hand, if player A's opponent is very much better at receiving a second serve than receiving a first serve, it may be better for player A to serve two first serves than to serve a fast first serve and a slower second serve. In this paper we see how the merits of this potential strategy of two fast serves can be assessed statistically by a player about to play a specific opponent.

This paper uses a large database as provided by KAN-soft (www.oncourt.info) to calculate player match statistics for each court surface. The methods of the analysis could be used by players and coaches to increase potential serving performance.

OnCourt database

The OnCourt database provides some match statistics since the 2003 French Open. Not all the match statistics for the ATP and WTA events are given. However the number of matches and tournaments included in the database has increased in recent years. The database is taken from the 6th August, 2007. The surfaces are categorized as grass, hard, indoor hard, clay, carpet and acrylic. For simplicity hard and indoor hard are considered as one surface. It is noted that acrylic will be played at the Australian Open from 2008. At August 2007, there are no match statistics recorded by the OnCourt database for matches played on acrylic.

A program was written in SAS to calculate the average serving and receiving statistics for each player on each surface. The serving and receiving statistics averaged across all matches on each surface was also calculated. Tables 1 and 2 below give these overall averages for men and women. The results indicate that women serve a higher percentage of 1st serves in play compared to men for all four surfaces. However, the results indicate that men win a higher percentage of points on the first and second serve compared to the women for all four surfaces. Overall, the results indicate that men win a higher percentage of points on serve compared to women on all four surfaces. This agrees with the results of Barnett and Pollard.[8] The difference between the percentage of wins on first serve for men and for women is 8.8%. The difference between the percentage of wins on second serve for men and women is 5.5%. As the former value is greater than the latter value, there is a suggestion or possibility that a fast first and fast second serve strategy is more likely to be a reasonable one in a men's match than in a women's match. >>

Table 1. Match statistics for men separated by court surface

Statistic	Grass	Carpet	Hard	Clay
WinPercentage of 1st serves in play	61.9%	61.3%	60.0%	61.5%
Percentage of points won on first serve	74.1%	73.0%	71.0%	67.1%
Percentage of points won on second serve	51.8%	51.3%	50.9%	49.2%
Percentage of points won on serve	65.5%	64.6%	62.5%	60.2%
Percentage of points won on return on first serve	25.9%	27.0%	29.0%	32.9%
Percentage of points won on return on second serve	48.2%	48.7%	49.1%	50.8%
Percentage of points won on return of serve	34.5%	35.4%	37.5%	39.8%
Number of matches	928	304	4319	3331

Table 2. Match statistics for women separated by court surface

Statistic	Grass	Carpet	Hard	Clay
Percentage of 1st serves in play	63.1%	63.4%	62.1%	63.4%
Percentage of points won on first serve	65.4%	63.1%	62.0%	59.6%
Percentage of points won on second serve	46.1%	46.4%	45.3%	43.6%
Percentage of points won on serve	58.1%	57.0%	55.5%	53.5%
Percentage of points won on return on first serve	34.6%	36.9%	38.0%	40.4%
Percentage of points won on return on second serve	53.9%	53.6%	54.7%	56.4%
Percentage of points won on return of serve	41.9%	43.0%	44.5%	46.5%
Number of matches	881	199	3432	2293

Serving Strategies

The players' serving and receiving statistics are defined as:

a_{is} = percentage of first serves in play for player i on surface s,
b_{is} = percentage of points won on first serve given that first serve is in for player i on surface s,
c_{is} = percentage of points won on second serve for player i on surface s,
d_{is} = percentage of points won on return of first serve for player i on surface s,
e_{is} = percentage of points won on return of second serve for player i on surface s.

The surfaces are defined as: s=1 for grass, s=2 for carpet, s=3 for hard and s=4 for clay.

Combining player statistics is a common challenge in sport. While we would expect a good server to win a higher proportion of serves than average, this proportion would be reduced somewhat if his opponent is a good receiver. Using the method developed by Barnett and Clarke[7] we can calculate the combined percentage a player wins on his/her first and second serve for each surface. The equations are given as follows:

$$f_{ijs} = b_{is} - d_{js} + davs \quad (1)$$
$$g_{ijs} = c_{is} - e_{js} + eavs \quad (2)$$

where:
f_{ijs} = percentage of points won on first serve given that first serve is in when player i meets player j on surface s,
g_{ijs} = percentage of points won on second serve when player i meets player j on surface s,
davs represents the average percentage of points won on return of first serve on surface s,
eavs represents the average percentage of points won on return of second serve on surface s.

A simple analysis can now be used to compare two serving strategies. The first strategy is where a player serves a fast serve on the first serve and a slow serve on the second serve. It is assumed that for the data collected, players have always used a fast first serve and a second slower serve. Even if this is not always the case, it would appear to be a reasonable assumption for the following analysis. Using this strategy, the percentage of points won on serve is given by:

$$a_{is} * f_{ijs} + (1- a_{is}) * gijs$$

The second strategy is where a player serves a fast serve on both the first and second serve. Using this strategy, the percentage of points won on serve is given by:

$$a_{is} * f_{ijs} + (1- a_{is}) * a_{is} * f_{ijs}$$

Therefore, a player should use the second strategy if

$$(1-a_{is}) * a_{is} * f_{ijs} > (1-a_{is}) * g_{ijs}$$

which simplifies to

$$a_{is} * f_{ijs} > g_{ijs}$$

Example: Andy Roddick versus Rafael Nadal

Serving and receiving statistics for Andy Roddick and Rafael Nadal are given in Table 3. Equations 1 and 2 are used to calculate fijs and gijs, where davs and eavs are obtained from Table 1. The lack of matches played on carpet by both players is noted. The results from Table 3 indicate that Roddick might be encouraged to serve fast on both the first and second serve when playing Nadal on grass. However he should use a fast first serve and slower second serve when playing Nadal on both hard court and clay. Nadal on the other hand should use a fast first serve and slower second serve when playing Roddick on grass, hard court and clay. This example illustrates the fact that it can be important for a player to identify the particular surface statistics for himself and his opponent.

The above analysis indicates that Roddick might do slightly better when playing Nadal on grass by using two first serves rather than using a first serve and a slower second serve. The effect however is not statistically significant. Nevertheless, Roddick might do well to mix his first and second serve when serving a second serve to Nadal. He would appear to have little to gain or lose sta-

Table 3. Serving and receiving statistics for Andy Roddick and Rafael Nadal

	Andy Roddick				Rafael Nadal			
Statistic	**Grass**	**Carpet**	**Hard**	**Clay**	**Grass**	**Carpet**	**Hard**	**Clay**
a_{is}	0.67	0.69	0.66	0.57	0.69	0.70	0.66	0.71
b_{is}	0.82	0.78	0.82	0.73	0.76	0.63	0.75	0.72
c_{is}	0.56	0.43	0.59	0.55	0.57	0.53	0.59	0.58
d_{is}	0.28	0.23	0.29	0.28	0.28	0.26	0.31	0.42
e_{is}	0.47	0.48	0.51	0.48	0.53	0.50	0.53	0.60
f_{ijs}	0.799	0.790	0.800	0.639	0.739	0.670	0.750	0.769
g_{ijs}	0.512	0.417	0.551	0.458	0.582	0.537	0.571	0.608
$a_{is} * f_{ijs}$	0.535	0.545	0.528	0.364	0.510	0.469	0.495	0.546
Matches	37	3	99	17	24	4	72	72

tistically by such a strategy, but he might gain a moderate amount from the 'surprise' factor in such a strategy. Similar analyses to test for significance can be performed for the other surfaces. However, it is clear that two first serves will not be such a good strategy for Roddick in the case of clay surfaces.

Conclusions

The results indicate that separating player match statistics into different court surfaces can be useful (for some players against some other players on particular surfaces) in making decisions on serving strategies. An example where one player might benefit by serving two fast serves has been given. The method of analysis could be used by any player or coach to see whether serving performance might be enhanced during a forthcoming match. Thus, this approach could be quite a valuable tool for some players.

Player match statistics could be used in other applications. For example, match statistics separated by court surface could be used by television broadcasters as a guide to likely match outcomes and comments on strategies.

Acknowledgement

The authors would like to thank KAN-soft for providing the database for calculating serving and receiving statistics.

References

1. Gale D. Optimal strategy for serving in tennis. Mathematics Magazine 1971;44:197-9.
2. Norman JM. Dynamic programming in tennis-when to use a fast serve. J Oper Res Soc 1985;36:75-7.
3. George SL. Optimal strategy in tennis: a simple probabilistic model. Appl Stat 1973;22:97–104.
4. Gillman L. Missing more serves may win more points. Mathematics Magazine 1985;58(4):222-4.
5. Hannan EL. An analysis of different serving strategies in tennis. In: Management Science in Sports. Ladany SP, Machol RE, Morrison DG (eds). Amsterdam: North–Holland Publishing Company, 1976:125–36.
6. Pollard GH, Pollard GN. Optimal risk taking on first and second serves. In: Tennis Science & Technology 3. Miller S, Capel-Davies J (eds). London: International Tennis Federation, 2007:273-80.
7. Barnett T, Clarke SR. Combining player statistics to predict outcomes of tennis matches. IMA J Management Math 2005;16:113-20.
8. Barnett T, Pollard G. How the tennis court surface affects player performance and injuries. Med Sci Tennis 2007;12(1):34-7.

About the authors

Dr. Denny Meyer is a Senior Lecturer in Statistics at the Swinburne University of Technology. She has co-authored two books and has published upward of 50 articles in refereed journals. She is an applied statistician, working in areas such as sport statistics, management, tourism, mineral processing, advertising, agriculture and social research.

Dr. Tristan Barnett is an Adjunct Research Fellow in sports statistics at Swinburne University of Technology. He has written several published papers, given presentations at international conferences and has a PhD in tennis modelling. He has appeared on many occasions in the media including SEN sports radio and 3RRR sports segment "Run Like You Stole Something". Tristan currently works as a gaming mathematician for Sportsbet21 Pty Ltd.

Emeritus Professor Graham Pollard is a former professor of applied statistics and pro-vice-chancellor of the University of Canberra, Australia. He is a former first grade tennis player and state squash champion, has a PhD in statistics from the Australian National University, and has international research publications in tennis, squash, theoretical statistics, maths education and physics.

Address for correspondence:
Dr. Tristan Barnett, Faculty of Life and Social Sciences, Swinburne University of Technology, John St, Hawthorn, VIC 3122 Australia
Website: www.swin.edu.au/sport
Email: strategicgames@hotmail.com