APPLYING TENNIS MATCH STATISTICS TO INCREASE SERVING PERFORMANCE DURING A MATCH IN PROGRESS

Tristan Barnett
Faculty of Business and Law, Victoria University, Australia

Introduction

Analysing risk-taking strategies in tennis is complicated. There has been a tendency to analyse risk-taking on the serve more often than other shots. This seems reasonable as the serve is the first shot to be played and therefore simplifies the analysis by not having to consider previous shots in the rally. Early research in tennis serving strategies includes Gale (1971), Norman (1985), George (1973), Gillman (1985) and Hannan (1976). More recently, Barnett et al. (2008) analysed the situation where players may choose to serve two fast serves by taking into account the type of court surface, and the serving and receiving capabilities of both players. Pollard and Pollard (2007) analysed the situation where there is a continuum amount of risk that players can take on their serve. They concluded that in most practical situations it was optimal for a player to serve a higher risk first serve and a lower risk second serve. Pollard (2008) analysed the typical situation in which a medium risk serve has a quadratic outcome rather than a linear one; one which gives greater weighting to the outcome of serving a high risk serve rather than the outcome of a low risk serve.

In this article, we analyse the situation where a player takes a high level of risk on the first serve, and either a high or low level of risk on the second serve. Similar analyses could be carried out for when a player takes a medium or low level of risk on either serve (athough the risk on the second serve should always be no greater than that taken on the first serve). A player may deviate from the strategy of taking a high level of risk on the first serve as a surprise strategy, but this problem is not considered in this paper. The model considered here allows serving strategies that are dependent on the type of court surface and the serving and receiving capabilities of any two top players. The model is extended to allow for the possibility of players changing serving strategies throughout the match in progress. Practical suggestions are given for the use of these methods by elite players in order to increase their performance. In particular, match statistics are displayed on the scoreboard for the players to view, and these statistics can assist the players as to whether to use two high risk serves or a typical high risk first serve and low risk second serve. Given that coaching intervention is allowed in team competition, this creates even more opportunities for the use of technology to assist players in determining serving strategies.

Serving strategies before a match

The following definitions are given for each type of serve:

A high risk serve is a typical first serve by each player
A low risk serve is a typical second serve by each player

Player match statistics are given by OnCourt (www.oncourt.info). Note that for the OnCourt statistics, the percentage of points won on the first serve are conditional on the first serve being in, whereas the percentage of points won on the second serve are not conditional on the second serve being in. Note that this difference leads to the definitions below. Estimates for the serving and receiving percentages for each player within a match can be obtained using the method in the paper by Barnett et al. (2008).

Let:
a_{his} = percentage of high risk serves in play for player i on surface s
a_{lis} = percentage of low risk serves in play for player i on surface s (note that this percentage is not needed)
b_{his} = percentage of points won on high risk serves (conditional on them being 'in') for player i on surface s
b_{lis} = percentage of points won on low risk serves (unconditional) for player i on surface s
c_{his} = percentage of points won on return of high risk serves (conditional on them being 'in') for player i on surface s
c_{lis} = percentage of points won on return of low risk serves (unconditional) for player i on surface s
d_{hijs} = percentage of points won on high risk serves (conditional on them being 'in') for player i, for when player i meets player j on surface s
d_{lijs} = percentage of points won on low risk serves (unconditional) for player i, for when player i meets player j on surface s
c_{has} = average percentage (for all players) of points won on return of high risk serves (conditional on them being 'in') for surface s
c_{las} = average percentage (for all players) of points won on return of low risk serves (unconditional) for surface s

The following assumptions are given:

1. $a_{his} < a_{lis}$
2. $b_{his} > b_{lis}$
3. $c_{his} < c_{lis}$
4. $d_{hijs} > d_{lijs}$
5. $d_{hijs} = b_{his} - c_{hjs} + c_{has}$
6. $d_{lijs} = b_{lis} - c_{ljs} + c_{las}$

Assumptions 1 - 4 are transitivity assumptions. For example, assumption 2 states that, for surface s, the percentage of points won by player i on a high risk serve when it went into play is assumed to be greater than the percentage of points won by player i on a low risk serve whether that serve went into court or not. Assumptions 5 and 6 are 'combining' formulas (Barnett and Clarke, 2005) to guarantee that the serve of player i and the return of player j sum to 100% of cases. As an example, we consider assumptions 5 and 6.

Suppose player i meets player j on a particular surface s
Suppose the match statistics for player i are $a_{his} = 0.67$, $b_{his} = 0.82$, $b_{lis} = 0.56$, $c_{his} = 0.29$ and $c_{lis} = 0.47$
Suppose the match statistics for player j are $a_{hjs} = 0.69$, $b_{hjs} = 0.76$, $b_{ljs} = 0.57$, $c_{hjs} = 0.28$ and $c_{ljs} = 0.53$
Suppose the averages are $c_{has} = 0.26$ and $c_{las} = 0.48$. Thus, for these numerical values, player j receives high risk serves better than the average player and player i receives high risk serves even better than player j. However, for low risk serves, player i receives low risk serves worse than average whilst player j receives low risk serves better than average.
In this example, $d_{hijs} = b_{his} - c_{hjs} + c_{has} = 0.82 - 0.28 + 0.26 = 0.80$, and $d_{hjis} = b_{hjs} - c_{his} + c_{has} = 0.76 - 0.29 + 0.26 = 0.73$.
Also, $d_{lijs} = b_{lis} - c_{ljs} + c_{las} = 0.56 - 0.53 + 0.48 = 0.51$ and $d_{ljis} = b_{ljs} - c_{lis} + c_{las} = 0.57 - 0.47 + 0.48 = 0.58$

The following two serving strategies are defined:
Strategy 1 – high risk serve followed by a high risk serve
Strategy 2 – high risk serve followed by a low risk serve
The percentage of points won on serve by player i by using each strategy is:
Strategy 1 – $a_{his} * d_{hijs} + (1- a_{his}) * a_{his} * d_{hijs}$
Strategy 2 – $a_{his} * d_{hijs} + (1- a_{his}) * d_{lijs}$
Thus, player i should use Strategy 1 (two high risk serves) rather than Strategy 2 if
$a_{his} * d_{hijs} + (1- a_{his}) * a_{his} * d_{hijs} > a_{his} * d_{hijs} + (1- a_{his}) * d_{lijs}$

and this inequality simplifies to $a_{his} * d_{hijs} > d_{lijs}$

An example of such a case is given in Barnett et al. (2008) between Andy Roddick (recognised as a 'strong' server) and Rafael Nadal (recognised as a 'strong' receiver). More detailed information on that example is now given. Serving and receiving statistics for Andy Roddick and Rafael Nadal are given in Table 1. The lack of matches played on carpet by both players is noted. The results from Table 1 indicate that Roddick might be encouraged to serve high risk on both the first and second serve when playing Nadal on grass. However he should use a high risk first serve and low risk second serve when playing Nadal on both hard court and clay. Nadal on the other hand should use a high risk first serve and low risk 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 high risk serves rather than using a high risk first serve and a low risk 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 statistically 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 high risk serves will not be such a good strategy for Roddick in the case of clay surfaces.

	Andy Roddick				Rafael Nadal			
Statistic	Grass	Carpet	Hard	Clay	Grass	Carpet	Hard	Clay
a_{his}	0.67	0.69	0.66	0.57	0.69	0.70	0.66	0.71
b_{his}	0.82	0.78	0.82	0.73	0.76	0.63	0.75	0.72
b_{lis}	0.56	0.43	0.59	0.55	0.57	0.53	0.59	0.58
c_{his}	0.28	0.23	0.29	0.28	0.28	0.26	0.31	0.42
c_{lis}	0.47	0.48	0.51	0.48	0.53	0.50	0.53	0.60
d_{hijs}	0.799	0.790	0.800	0.639	0.739	0.670	0.750	0.769
d_{lijs}	0.512	0.417	0.551	0.458	0.582	0.537	0.571	0.608
$a_{his} * d_{hijs}$	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

Table 1: Serving and receiving statistics for Andy Roddick and Rafael Nadal

Serving strategies during a match

A player's best strategy based on the available data when the match begins does not necessarily remain his/her best strategy throughout the entire match. The initial optimal strategy may be sensitive to variations in the serving and receiving statistics for the two players during the match.

It is assumed that after the first set a player's serving and receiving statistics (as given by the match statistics) give the best estimates of their serving and receiving parameters for the second set. After the second set a player's serving and receiving statistics give the best estimates for the serving and receiving parameters for the third set, and this process continues for each remaining set of the match.
Let the superscript ^ represent the actual match percentage statistics at a position in the match. Then player i should use Strategy 1 (two high risk serves) at the start of each set if

$a^\wedge_{his} * b^\wedge_{his} + (1- a^\wedge_{his}) * a^\wedge_{his} * b^\wedge_{his} > a^\wedge_{his} * b^\wedge_{his} + (1- a^\wedge_{his}) * b^\wedge_{lis}$
and this simplifies to $a^\wedge_{his} * b^\wedge_{his} > b^\wedge_{lis}$ (1)

Note that $b^\wedge_{his} + c^\wedge_{hjs} = 1$ and $b^\wedge_{lis} + c^\wedge_{ljs} = 1$ to ensure that the serve of player i and the return of player j sum to 100%.

Practical implementations

The International Tennis Federation (ITF) is the governing body of the game of tennis and its duties and responsibilities include the determination of the Rules of Tennis. Section 30 of those Rules states:

Coaching is considered to be communication, advice or instruction of any kind, audible or visible, to a player. In team events where there is a team captain sitting on-court, the team captain may coach the player(s) during a set break and when the players change ends at the end of a game, but not when the players change ends after the first game of each set and not during a tie-break game. In all other matches, coaching is not allowed.
Case 1: Is a player allowed to be coached, if the coaching is given by signals in a discreet way?
Decision: No.
Case 2: Is a player allowed to receive coaching when play is suspended?
Decision: Yes.

The ITF Chief of Officiating notes "*Players are not allowed access to any kind of electronic equipment during a match. This would include computers, mobile phones, iPods etc. Equally players are not able to access the statistics via the chair umpire - any statistics gathered are either taken from the side of the court by another operator or else generated by the chair umpires PDA (but he/she would not have access to the accumulated statistics anyway)*".

From the above, players can only access match statistics from the scoreboard, and they can use this information to make decisions as to the strategies they may use in the immediately forthcoming match play. For example, if only the first serve percentage was available, then a player would use Strategy 1 in the next set if
$\hat{a}_{his} * d_{hijs} + (1- \hat{a}_{his}) * \hat{a}_{his} * d_{hijs} > \hat{a}_{his} * d_{hijs} + (1- \hat{a}_{his}) * d_{lijs}$

and this simplifies to $\hat{a}_{his} > d_{lijs} / d_{hijs}$ (2)

Each side of this inequality could be calculated by the player on a sheet of paper that is taken with him/her into the match. For example if $d_{lijs} = 0.5$ and $d_{hijs} = 0.8$, then player i would use Strategy 1 in the second set only if $\hat{a}_{his}$ was greater than 0.5 / 0.8 = 0.625 for the first set.

Other inequalities include the percentage of points won on the first and second serve during the match. If only the percentage of points won on the first serve was available, it can be shown that a player would use Strategy 1 in the next set if

$\hat{b}_{his} > d_{lijs} / a_{his}$ (3)

If only the percentage of points won on the second serve was available, it can be shown that a player would use Strategy 1 in the next set if

$\hat{b}_{lis} < a_{his} * d_{hijs}$ (4)

Inequalities (2), (3) and (4) provide practical and easy-to-use strategies for a player during a match and are summarized in Table 2.

Serving Statistic	Inequality
First Serve Percentage	$\hat{a}_{his} > d_{lijs} / d_{hijs}$
Percentage of Points won on First Serve	$\hat{b}_{his} > d_{lijs} / a_{his}$
Percentage of Points won on Second Serve	$\hat{b}_{lis} < a_{his} * d_{hijs}$

Table 2: Inequalities for determining serving strategies during a match in progress

The ITF Chief of Officiating notes "*This changes in team competitions where there is a captain on court with them. Here there are usually team support personnel sitting right behind the captain so outside help could possibly make its way to the player via the captain. I would not allow a captain to be using computers/mobile phones etc but realistically if this information is wanted (and I wouldn't personally overestimate the value of such information) then that message has potential to reach the player"*.

Thus, there is potential to allow inequality (1) to be implemented during a match to determine serving strategies.

There are limitations to the above serving strategies. Firstly, the serving match statistics may not be available on the scoreboard at the end of each set, or when required throughout the match. Further, a player may be continually observing the scoreboard for this information and if the relevant statistic becomes available during an uncompleted game, a player may be required to remember the statistic for analysis at the next change of ends. Finally, the pressure of the occasion may cause a player to be below expectation in serving a high risk second serve. For this reason, the strategies should only be interpreted as a guide as to taking more risk on the second serve.

Conclusions

The use of technology can create opportunities for players to determine serving strategies during a match in progress. Match statistics data are entered by on-court statisticians throughout a match. This data is then sent in real-time to a central computer which is made available to the public through the internet. On occasions, these match statistics are displayed on the scoreboard for the players to view, and these statistics can assist the players as to whether to use two high risk serves or a typical high risk first serve and low risk second serve. Given that coaching intervention is allowed in team competition, this creates even more opportunities for the use of technology to assist players in determining serving strategies. This has been demonstrated in this article.

Acknowledgement

The author would like to acknowledge Andrew Jarrett from the ITF for discussions on the Rules of Tennis.

References

Barnett T and Clarke SR (2005). Combining player statistics to predict outcomes of tennis matches, *IMA Journal of Management Mathematics*. 16(2), 113–120.

Barnett T, Meyer D and Pollard G (2008). Applying match statistics to increase serving performance, *Medicine and Science in Tennis*. 13(2), 24-27.

Gale D (1971). Optimal strategy for serving in tennis, *Mathematics Magazine*. 44(4), 197–199.

George SL (1973). Optimal strategy in tennis: a simple probabilistic model, *Applied Statistics*. 22, 97–104

Gillman L (1985). Missing more serves may win more points, *Mathematics Magazine* 58(4), 222–224.

Hannan EL (1976). An analysis of different serving strategies in tennis, In *Management Science in Sports*, S.P. Ladany, R.E. Machol and D.G. Morrison eds., Amsterdam: North–Holland Publishing Company, 125–136.

Norman JM (1985). Dynamic programming in tennis-when to use a fast serve, *Journal of the Operational Research Society*. 36(1), 75–77.

Pollard GN and Pollard GH (2007). Optimal risk taking on first and second serves, In proceedings of *Tennis Science & Technology 3*, S. Miller and J. Capel-Davies eds., London: International Tennis Federation, 273-280.

Pollard GN (2008). What is the best serving strategy? *Medicine and Science in Tennis*. 13(2), 34-38.