Analyzing Wimbledon

Franc Klaassen (University of Amsterdam, The Netherlands)
& Jan R. Magnus (VU University Amsterdam, The Netherlands)

ITF Coaching and Sport Science Review 2014; 62 (22): 6 - 7

ABSTRACT

This paper highlights some of the topics discussed in our recently published book 'Analyzing Wimbledon', which provides a statistical analysis of many topics relevant for tennis players, coaches, commentators, and spectators.

Key words: commentators, prediction, service strategy, winning mood, statistics

Article received: 12 January 2014

Corresponding Author: f.klaassen@uva.nl

Article accepted: 22 February 2014

INTRODUCTION

This paper introduces and summarizes our recently published book, entitled Analyzing Wimbledon. In this book we study commentators' wisdoms, which are of interest to tennis fans, players, and statisticians. The book contains the results of more than fifteen years of research, using data on 100,000 points played in the men's and women's singles at Wimbledon and data of other grand slam tournaments. The book also discusses how the outcome of a match can be predicted (also while the match is in progress), which points are important and which are not, how to choose an optimal service strategy, and how tennis results can be used to better understand human behavior in general. The book uses tennis as a vehicle to illustrate the power and beauty of statistical reasoning.

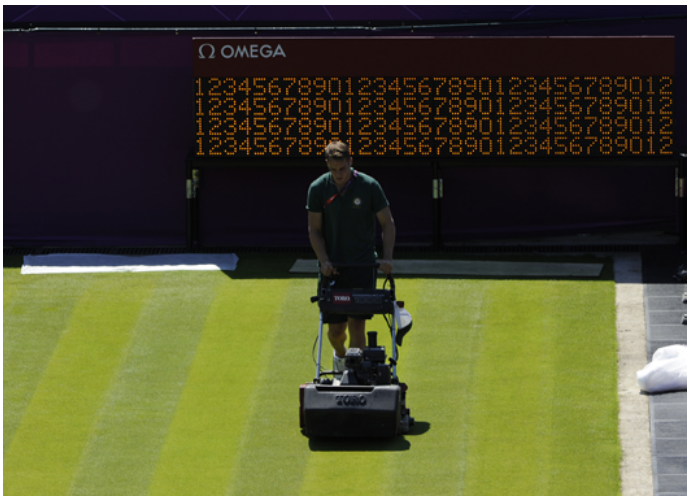
COMMENTATORS' WISDOMS

Suppose you are watching a tennis match between Novak Djokovic and Rafael Nadal. The commentator says: 'Djokovic serves first in the set, so he has an advantage'. Why would this be the case? Perhaps because he is then 'always' one game ahead, thus serving under less pressure. But does it actually influence him and, if so, how? Now we come to the seventh game, according to some the most important game of the set. But is it? Nadal serves an ace at breakpoint down (30-40). Of course! Real champions win the big points. But they win most points on service anyway, including the unimportant points. Do the real champions over-perform on big points or do weaker players underperform, so that it only seems that the champions over-perform? (The latter is the case.) Then Nadal wins three consecutive games. He is in a winning mood, the momentum is on his side. But does a 'winning mood' actually exist in tennis? (It does, but it is smaller than many expect.)

An Example

Let us consider some concrete examples. To study the serving-first advantage we first use data on more than one thousand sets played at Wimbledon and calculate how often the player who served first also won the set. This statistic shows that for the men there is a slight advantage in the first set, but no advantage in the other sets. On the contrary, in the other sets there is a disadvantage: the player who serves first in the set is more likely to lose the set than to win it. This is surprising. What could be the explanation? Perhaps it is different for the women? But no, the same pattern occurs in the women's singles. The explanation is that the player who serves first in a set (if it is not the first set) is usually the weaker player. This is so, because (a) the stronger player is more likely to win the previous set, and (b) the previous set is more likely won by serving the set out than by breaking serve. Therefore, the stronger player typically wins the previous set on service, so that the weaker player serves first in the next set. The weaker player is more likely to lose the current set as well, not because of a service (dis)advantage, but because he or she is the weaker player.

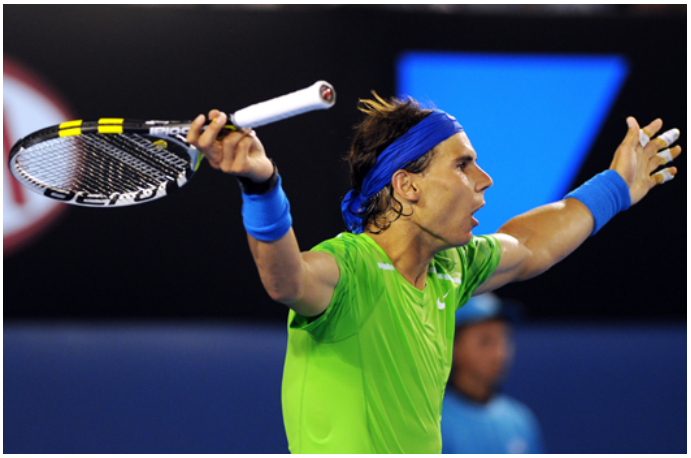
This example shows that we must be careful when we try to draw conclusions based on simple statistics. In this case, the fact that the player who serves first in the second and subsequent sets often loses the set is true, but this concerns weaker players while the hypothesis concerns all players. If we wish to answer the question of whether serving first causes a (dis)advantage, we have to control for quality differences. If we do this correctly, then we find that there is no advantage or disadvantage for the player who serves first in a set; in other words, it does not matter who serves first in the second or subsequent sets. But in the first set it does matter (the book shows why), so it is wise to elect to serve after winning the toss.



HUMAN BEHAVIOR

Studying tennis is not only of interest to those interested in tennis. There is a second (some would say a first) interest, namely the study of human behavior. In professional tennis the players' objectives are clear: they want to win. The incentives to win are strong and the players are highly trained. Tennis data are clean — there are few errors in the data — and each match generates a lot of data: many points, many services, and so on. And a good first impression of the quality of a tennis player is obtained from the world ranking. Such circumstances are rare in psychology, economics, and related disciplines, so that analyzing tennis can help.

One example is the question of whether people become more cautious when pressure mounts. In tennis, some points are more important than others. Do players behave differently in the key points? They do: they play safer at important points. This teaches us something about human behavior, and may have implications outside tennis, for example in economics. If salaries of agents working in the financial sector contain not only a bonus, but also a substantial component, then the consequences of their activities matter in both directions (like winning or losing a tennis match).



FORECASTING

Can we forecast the winner of a tennis match? Let us consider the Australian Open 2012 final between Novak Djokovic and Rafael Nadal. Of course, by now we know the result: Djokovic won by 5-7, 6-4, 6-2, 6-7, 7-5 in the longest grand slam final ever: five hours and fifty-three minutes. The question, however, is what the probability of Djokovic winning before the end was? At the start of the match betting odds revealed a winning probability of 57.4% for Djokovic (and thus 42.6% for Nadal.)

CONCLUSION

The book describes how we update this probability after each point played. The actual calculation is done by our computer program Richard, which is freely available from our websites in a user-friendly format. Richard delivers the updated winning probability within a second and the resulting graph of probabilities provides a quick overview of the match developments so far and a direct forecast of who will win the match.

The figure below shows a graph at 5-5 in the final set, from the point of view of Djokovic. The start of the graph is at 57.4%, so Djokovic was expected to win. At the beginning of the tiebreak in the fourth set, Djokovic had a winning probability of 78%, and at 5-3 in the tiebreak (two points from victory) even 92.2%. But Nadal won the tiebreak 7-5 and the probability dropped sharply to 54.2%. In the final set Nadal broke Djokovic in the sixth game. At that point, 4-2 for Nadal in the final set, the probability had dropped to 16.1%, and at 30-15 for Nadal even to 12.5%. This was the point in the match with the highest probability for Nadal to win: 87.5%. Then, Djokovic broke Nadal's service and so it became 5-5, where the probability was 53.2% and Djokovic was the favorite again. Such swings are not visible in the score, or from summary statistics typically presented on television (percentage of first services in, number of aces, and so on). But the swings are visible in the probability graph. Thus the graph should prove to be a powerful tool for commentators and viewers.

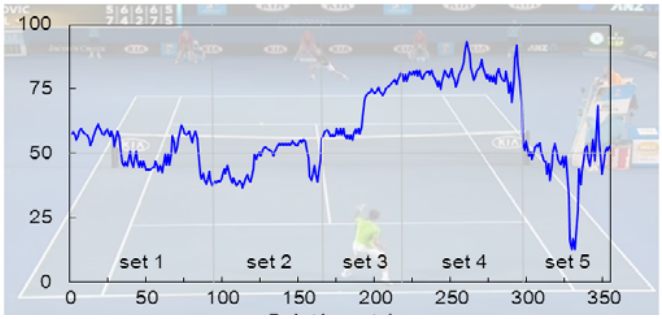


Figure 1. Probability of Djokovic winning the match.

REFERENCES

Klaassen, F. & J.R. Magnus (2014). Analyzing Wimbledon: The Power of Statistics. New York: Oxford University Press.

RECOMMENDED ITF TENNIS ICOACH CONTENT (CLICK BELOW)

TennisiCoach