Forecasting in tennis

Franc J.G.M. Klaassen[1], Jan. R. Magnus[2]
[1]*Department of Economics, University of Amsterdam, The Netherlands*
[2]*CentER, Tilburg University, The Netherlands*

We show how statistics can help us forecast the winner of a tennis match, not only at the beginning of the match, but also (and in particular) during the match.

INTRODUCTION

TV broadcasts of tennis matches present a large number of statistics, such as the percentage of first serves in and the number of aces. In addition to the current score, these statistics provide insight in the question which player performs well and is more likely to win. However, a direct estimate of the probability that a player wins the match is not shown. This is remarkable, because this statistic is the one that viewers want to know above all.

In this paper we discuss how to estimate the probability of winning a tennis match, not only at the beginning of a match, but in particular while the match is in progress. This leads to a profile of probabilities, which unfolds during the match and can be plotted instantly in a graph. The profile should be of interest to TV viewers, commentators, and players. As a side product we also produce a plot of the importance of each point in the match. A more complete (and more technical) analysis is provided in Klaassen and Magnus (in press).

The basis of the approach is our computer program TENNISPROB, to be discussed in the next section. For a match between two players A and B, the program calculates the probability π_a that A wins the match (or, equivalently, $\pi_b = 1 - \pi_a$ that B wins).

The following ingredients are needed to compute π_a. Let p_a denote the probability that A wins a point on service, and p_b the probability that B wins a point on service. Then, under the assumption that points are independent and identically distributed (i.i.d.), the match probability π_a depends on the point probabilities p_a and p_b, the type of tournament (best-of-3-sets or best-of-5-sets, tiebreak in final set or not), the current score, and the current server. Our program calculates the probabilities exactly and fast.

The main assumption we make is therefore that points in tennis are i.i.d. The validity of this was investigated in Klaassen and Magnus (2001). They concluded that, although points are not i.i.d., the deviations from i.i.d. are small and hence the i.i.d. assumption is justified in many applications, such as forecasting.

In the third section we estimate π_a at the start of a match, that is, we estimate the first point of the profile. We use data on all Wimbledon singles matches, 1992-1995, and thus focus on profiles for matches at Wimbledon; to get profiles for other tournaments one could use data on those tournaments. It is also allowed for the user of the program (say, the commentator) to use his/her own view on π_a, for instance, to account for an injury problem or fear against this specific opponent, information that is not available to us. In the end, there is one starting point for the profile.

To estimate π_a during the match, that is, to get the complete profile, TENNISPROB requires estimates of the two unknown point probabilities p_a and p_b. These estimates cannot be obtained from match data. Thus, we use point-to-point data, which we have for a subset of the 1992-1995 singles matches. We only have to use such data to estimate $p_a + p_b$. After all, π_a at the start of the match is a function of p_a and p_b and hence of $p_a - p_b$ and $p_a + p_b$, and since we now have estimates of π_a (from match data) and of $p_a + p_b$ (from point data), we obtain an estimate of $p_a - p_b$ by inverting the program. This gives us estimates of both p_a and p_b, and the profiles can now be drawn.

In the fourth section we demonstrate the use of the theory and the program by drawing profiles of two famous Wimbledon finals, Sampras-Becker (1995) and Graf-Novotna (1993). Such profiles can be drawn for any match, not only when the match is completed, but also while the match is in progress. The fifth section summarizes.

INTRODUCING TENNISPROB

Consider one match between two players A and B. As motivated in the previous section, we assume that points are independent and identically distributed, depending only on who serves. Then, modelling a tennis match between A and B depends on only two parameters: the probability p_a that A wins a point on service, and the probability p_b that B wins a point on service. Given these two (fixed) probabilities, given the rules of the tournament, given the score and who serves the current point, one can calculate exactly the probability of winning the current game (or tiebreak), the current set, and the match.

The program TENNISPROB is an efficient (and very fast) computer program that calculates these probabilities. The probabilities are calculated exactly; they are not simulated. The program is flexible, because it allows the user to specify the score and to adjust to the particularities of the tournament, but also because it allows for rule changes. For example, the traditional scoring rule at deuce can be replaced by the alternative of playing one deciding point at deuce, or we can analyse what would happen if the tournament requires 4 games rather than 6 to be won in order to win a set (not currently allowed by the official rules).

The program can also be used to calculate the importance of a point, defined by Morris (1977) as the probability that A wins the match if he/she wins the current point minus the probability that A wins the match if he/she loses the current point. TENNISPROB can tell us during the match what the important points are, and we will plot these just like the profiles.

FORECASTING THE WINNER OF A TENNIS MATCH

We consider a match between two players A and B and wish to estimate the probability π_a that A wins the match at each point in the match. This section describes the main reasoning behind the estimation; the complete argument can be found in Klaassen and Magnus (in press).

We first need to know how 'good' the two players are. As an indicator of this we use the rankings of the two players as determined by the lists published just before Wimbledon by the Association of Tennis Professionals (ATP) for the men, and the Women's Tennis Association (WTA) for the women. The ranking of player A is denoted RANK_a.

Direct use of the rankings is not satisfactory, because quality in tennis resembles a pyramid: the difference between the top two players (ranked 1 and 2) is generally larger than between two players ranked 101 and 102; see also Lebovic and Sigelman (2001). The pyramid can be based on the 'expected round', that is the round in which we expect the player to lose. For example, 3 for a player who is expected to lose in round 3, 7 for a player who is expected to reach the final (round 7) and lose, and 8 for the player expected to win the final.

A problem with 'expected round' is that it does not distinguish, for example, between players ranked 9-16 since all of them are expected to lose in round 4. Thus we propose a smoother measure of 'expected round' by transforming the ranking into a variable R:

$$R_a = 8 - \log_2(\text{RANK}_a)$$

For example, RANK = 3 implies R = 6.42, while RANK = 4 implies R = 6.00. Klaassen and Magnus (2001) provide further discussion and justification of this measure.

We shall always assume, obviously without loss of generality, that A is the 'better' player in the sense that $R_a > R_b$. The better player does not always win. At Wimbledon 1992-1995 the better player won 68% of the matches in the men's singles and 75% of the matches in the women's singles. So, upsets occur regularly, especially in the men's singles.

We now estimate π_a – the probability that the 'better' player A wins the match – at the start of the match. This will be the first point of our profile. We assume a simple logit model,

$$\pi_a = \exp(\lambda(R_a - R_b)) / [1 + \exp(\lambda(R_a - R_b))]$$

If $R_a = R_b$, then both players are equally strong. In that case $\pi_a = 0.5$, as one would expect.

Estimating π_a by maximum likelihood, based on our 1992-1995 Wimbledon data set of all matches, gives an estimate of 0.3986 (0.0461) in the men's singles, and 0.7150 (0.0683) in the women's singles, with the standard errors given in parentheses. For example, for $R_a = 8$ and $R_b = 4$ (that is, number 1 against number 16 in the official rankings), we find an estimated π_a of 0.8312 in the men's singles and 0.9458 in the women's singles. In Figure 1 we plot the estimated π_a as a function of $R_a - R_b$ for both men and women.

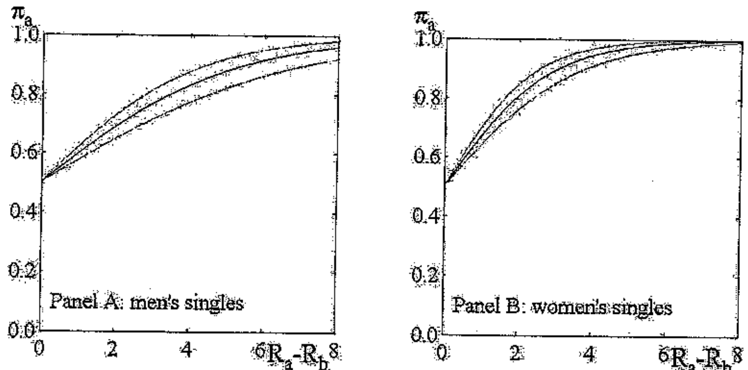

Figure 1. Estimated probability π_a that A wins match as a function of ranking difference.

For $R_a - R_b = 0$ we have $\pi_a = 0.5$, but when $R_a - R_b$ increases, π_a increases to 1. The increase is faster for the women than for the men, illustrating again that upsets are less likely in the women's singles than in the men's singles. Also plotted are the 95% confidence bounds, based on the uncertainty about λ.

Of course, the user of the profile (say the commentator) may be unhappy with our pre-match estimate that A will win. Very likely, the commentator will have information about the players in addition to their rankings, for example special ability on grass, fear against this specific opponent, and health/injury problems. The commentator can adjust our estimate of π_a taking his or her own knowledge into account.

We now have an estimate of the probability π_a that A wins the match, at the start of the match. In order to calculate the other points of the profile we need an estimate of $p_a + p_b$, where p_a denotes the probability that A wins a point on service against B. To estimate $p_a + p_b$, match data are not enough; we need point-to-point data. We have such data on a subset of matches; these data are fully described in Magnus and Klaassen (1999). As argued in Klaassen and Magnus (in press), we estimate the probability $p_a + p_b$ by

$$p_a + p_b = 2[\ \beta_0 + \beta_2\ (R_a + R_b)\ ],$$

where the estimated β_0 is 0.6276 (0.0044) in the men's singles and 0.5486 (0.0051) in the women's singles, and the estimated β_2 is 0.0036 (0.0009) in the men's singles and 0.0022 (0.0010) in the women's singles (standard errors in parentheses). It is clear that the estimated $p_a + p_b$ increases with $R_a + R_b$, since the estimate of β_2 is positive, and also that the increase is very slight, since the estimate is small.

Our forecast strategy is then as follows. Before the start of a given match, we know R_a and R_b. This gives us an estimate of π_a based on match data (Figure 1), possibly adjusted by the commentator. We also have an estimate of $p_a + p_b$ based on point data. For given $p_a + p_b$, π_a at the start of a match is a monotonic function of $p_a - p_b$. Hence, by inverting TENNISPROB, we obtain an estimate of $p_a - p_b$ as well. We thus find estimates of $p_a + p_b$ and $p_a - p_b$ and hence of p_a and p_b. With these estimates we can calculate the probability that A wins the match at each point in the match, using TENNISPROB.

PROFILES

To illustrate this theory we present profiles of two important Wimbledon finals. Each profile is a graph of the estimated probability that a given player wins the match at the beginning of each point. It is important to realize that the profile unfolds during the match, so that, for instance, after the first few games only a short line is visible. In this paper, however, we can present the complete profiles, because the matches under consideration have already been completed.

The first match is the 1995 men's final Sampras-Becker. Here Sampras (player A) was the favourite, having RANK = 2 and hence $R_a = 7$, while Becker (player B) had RANK = 4 and $R_b = 6$. Our pre-match estimates are that Sampras has a 59.8% chance of winning the championship (because the estimate of π_a is 0.5983), and that $p_a + p_b$ is estimated at 1.3487. As a consequence, TENNISPROB calculates an estimate of 0.0161 for $p_a - p_b$, and hence that the estimates of p_a and p_b are 0.6824 and 0.6663, respectively.

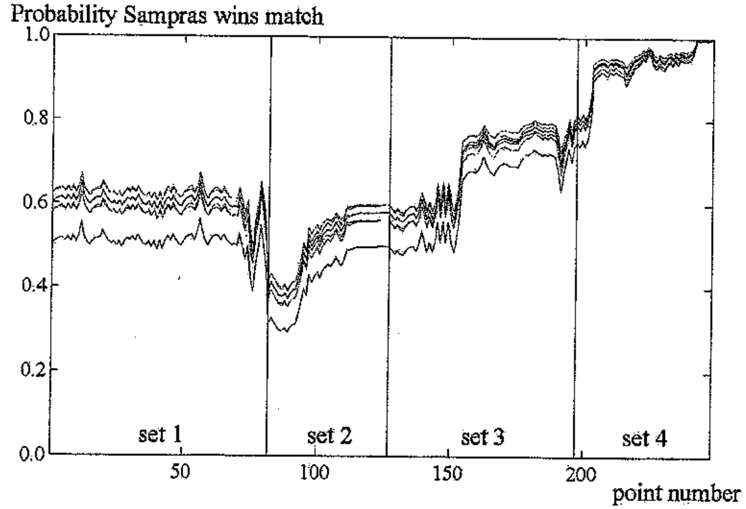

Figure 2a. Profile of Sampras-Becker 1995 final.

There are many lines in figure 2a. We first discuss the central profile which, given the estimate of π_a, starts at 59.8%. In fact, at the start of the match this is the only visible point of the profile. The first set goes to a tiebreak. After losing the tiebreak, Sampras' probability of winning has decreased to 39.6%. The profile has developed through the first vertical line. In the second set, Sampras breaks Becker's service at 1-1 and again at 3-1, and wins the set. In the third and fourth sets, Becker's service is broken again three times. The last break (at 4-2 in the fourth set) increases Sampras's chances only marginally, since he is already almost certain to win. Eventually Sampras wins 6-7, 6-2, 6-4, 6-2 after 246 points.

The two fuzzy curves just above and below the central profile both consist of two curves. Each of these four additional curves is based on a different combination of π_a and $p_a + p_b$. The upper two, hardly distinguishable, curves are based on the upper 95% confidence bound for π_a (see figure 1) in combination with either the upper or the lower 95% confidence bound for $p_a + p_b$. The lower two curves are based on the lower 95% confidence bound for π_a with either the upper or the lower 95% confidence bound for $p_a + p_b$. What we see is that the level of the profile can shift a bit, but that the movement of the profile is not affected when the initial estimates of π_a and $p_a + p_b$ and thus p_a and p_b are somewhat biased. Even when we simply take $\pi_a = 0.5$ at the start of the match (also plotted), the movement of the profile is the same. We conclude that the level of the profile depends on the correct estimation of p_a and p_b, but that the movement of the profile is robust.

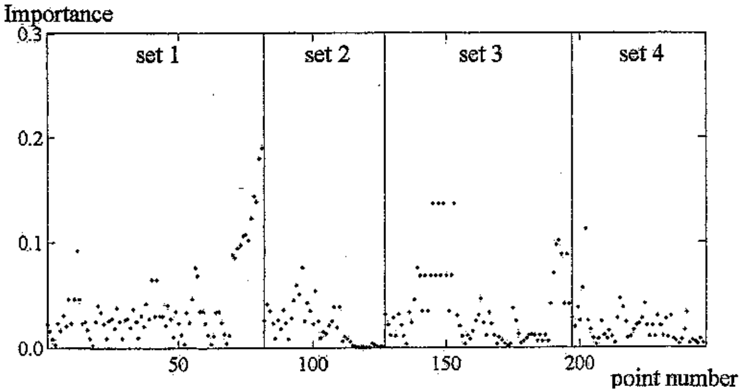

Figure 2b. Importance of each point in the Sampras-Becker 1995 final.

Figure 2b shows the importance of each point, as defined in the second section. One can clearly see the importance of the tiebreak at the end of the first set, and in particular the last point of the tiebreak (Becker's set point at 6-5), which is the most important point of the whole match. The importance is 0.19, meaning that the probability of winning the match for Sampras is 19 percentage points higher if Sampras wins this point than if he loses the point. Also important are the four breakpoints at 1-1 in the third set.

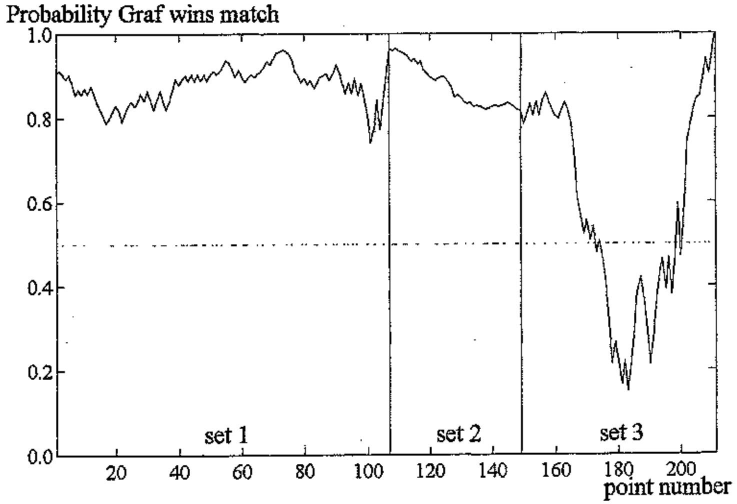

Figure 3a. Profile of Graf-Novotna 1993 final.

For the second match (figure 3a) we only show the central profile (and the 50% line: at points above the 50% line we expect A to win, at points below the line we expect B to win). This is the plot that one may want to show to a television audience, updated after every few games. This profile concerns the famous 1993 women's singles final Graf-Novotna. Graf (player A) was the

favourite, having RANK = 1 and hence $R_a = 8$, while Novotna had RANK = 9 and hence $R_b = 4.83$. Our pre-match estimates are that Graf has a 90.6% chance of winning (as the estimate of π_a is 0.9060), and that $p_a + p_b$ is estimated at 1.1538. As a consequence, we calculate that $p_a - p_b =$ 0.0992, and hence that the estimates of p_a and p_b are 0.6265 and 0.5273, respectively.

The first set goes to a tiebreak. At the beginning of the tiebreak (point 93), Graf's probability of winning has decreased a little to 85.9%. After winning the tiebreak, the probability jumps to 96.5% (point 107). Novotna wins the second set easily. At the beginning of the third set, Graf's probability of winning is still 81.7% (point 149). At 1-1 in the third set Graf's service is broken, and at 3-1 again. When Novotna serves at 4-1, 40-30 (point 183), Graf's probability of winning has dropped to 14.9%. Then Graf breaks back, and holds service (after two breakpoints). When Novotna serves at 4-3, 40-40, the match is in the balance. This is the most important game of the match and the two breakpoints in this game are the most important points of the match (see the importance plot in figure 3b). Both have importance equal to 0.27, so that if Graf breaks, her probability of winning the match will be 27 percentage points higher than if she does not break. Novotna loses the second breakpoint, the next two games, and the match. Graf wins 7-6, 1-6, 6-4 after 210 points.

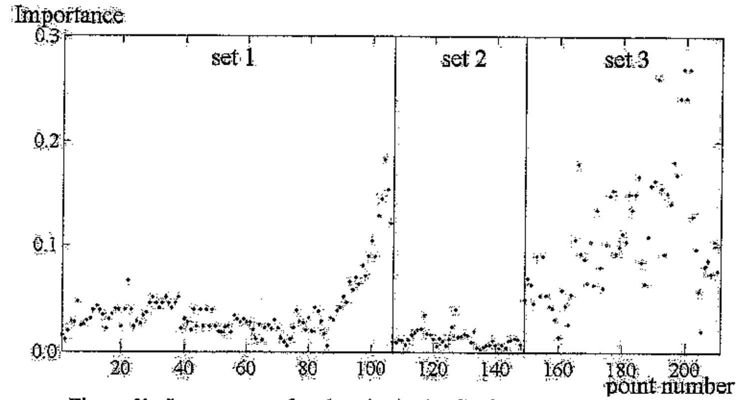

Figure 3b. Importance of each point in the Graf-Novotna 1993 final.

CONCLUSION

In this paper we have described a method of forecasting the outcome of a tennis match. More precisely, we have estimated the probability that one of the two players wins the match, not only at the beginning of the match but also as the match unfolds. The calculations are based on a flexible computer program TENNISPROB and on estimates using Wimbledon singles data 1992-1995, both at match level and at point level. Such profiles can be made for any match, not only matches at Wimbledon. The profiles can be used by commentators in assessing the 'turning points' in a match and, in addition, indicate which points of the match are the most important.

ACKNOWLEDGEMENTS

The authors are grateful to IBM UK and The All England Lawn Tennis and Croquet Club at Wimbledon for their kindness in providing the data, and to Dmitry Danilov for help in drawing the pictures.

REFERENCES

Alefeld, B (1984). Statistische Grundlagen des Tennisspiels. *Leistungssport*, **3**, 43-46.

Klaassen, F.J.G.M. and Magnus, J.R. (2001). Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, **96**, 500-509.

Klaassen, F.J.G.M. and Magnus, J.R. (in press). Forecasting the winner of a tennis match. *European Journal of Operational Research.*

Lebovic, J.H. and Sigelman, L. (2001). The forecasting accuracy and determinants of football rankings. *International Journal of Forecasting*, **17**, 105-120.

Magnus, J.R. and Klaassen, F.J.G.M. (1999). On the advantage of serving first in a tennis set: four years at Wimbledon. *The Statistician (Journal of the Royal Statistical Society, Series D)*, **48**, 247-256.

Morris, C. (1977). The most important points in tennis. In *Optimal Strategies in Sport* (edited by S.P. Ladany and R.E. Machol), pp. 131-140. Amsterdam: North-Holland Publishing Company.