DYNAMIC EVALUATION OF CONDITIONAL PROBABILITIES OF WINNING A TENNIS MATCH

Shaun Clowes, Graeme Cohen and Ljiljana Tomljanovic*
Faculty of Information Technology
University of Technology, Sydney
PO Box 123, Broadway
NSW 2007, Australia
shaman@progsoc.org, graeme.cohen@uts.edu.au, ltomljan@socs.uts.edu.au

Abstract

Let p and q be the probabilities of players A and B, respectively, winning a point in a game of tennis $(p+q=1)$. We describe a program which has been devised and implemented to give the probability of A or B winning the match ("best of three tiebreaker sets") from any stage of the match.

1 Introduction

It is possible, based on a study of earlier matches between two players under similar match conditions, to give each a probability of winning a point in any subsequent match between them. The probabilities may be subjectively altered, as it is felt warranted, and this may occur as the match is in progress. What are the respective chances of the two players winning the match, from any particular stage of the match?

A corresponding question was considered by Croucher [2], but for a fixed probability of winning a point and with regard to a single game.

Our players, A and B, will be assigned respective probabilities p and q of winning any point $(p+q=1)$. In the first place, these may be determined as the proportion of points won by each in all matches played previously under similar conditions. A more detailed analysis, such as that of Carter and Crews [1], might consider separately the proportions p_A and p_B of points won by A and B, respectively, in service games, so that, for example, $q_A = 1-p_A$ is the probability of B winning a point against A's service. This level of detail is not required for our purposes.

It is crucial to our work that "the outcome of any point, game or set is independent of the outcome of any other point, game or set" (Pollard [4]). Some doubt has recently been cast on this statement by Klaassen and Magnus [3]. They concluded that winning a point in tennis has a positive effect on winning the following point and that at important points the server has a disadvantage. However, they also conclude that divergence from independence is small, particularly for strong players, and "in many practical applications concerning tennis—such as predicting the winner of the match while the match is in progress—the iid hypothesis will still provide a good approximation *if we correct for the quality of the players*" [their emphasis]. We explicitly allow this correction at any stage of the match, and so are in fact reinforced in our application of independence.

*This project began as an assignment carried out by the first and third authors in a subject taught by the second author.

Our aim is to describe the calculations in the implementation of a publicly-available program which gives the probability of A or B winning a "best of three tiebreaker sets" match, from any stage of the match, as well as the probability of winning the current game or set. This would be of interest to media commentators and to those who might be having a bet on the result. Since it assumes an overall probability of winning a point, whether serving or receiving, results produced by the program should be quoted only after an even number of games have been played.

For example, if player B is initially given a probability of only 0.44 of winning a point (based on past experience), but has just won the first four games of the first set, would you bet on B to win that set? Would you bet on B to win the match? What odds would you accept? Our calculations show that, if you still believe B's previous form to apply, then there is a 75% chance of winning that set, but only a 20% chance of winning the match. If you feel it warranted to raise B's probability of winning a point to a modest 0.5, then there is better than a 90% chance that B will win this set and better than a 70% chance of winning the match.

2 Conditional probabilities of winning a game

Let A and B have probability p, q, respectively, of winning a point ($p + q = 1$). We take the point of view always of A winning a game, set or match. If A leads 30–15, for example, then A wins the game by winning the next two points (probability p^2), or two of the next three including the third (probability $2p^2q$), or by reaching deuce (probability $3pq^2$) and winning from there. The probability d of winning from deuce satisfies

$$d = p^2 + 2pqd,$$

so that

$$d = \frac{p^2}{p^2 + q^2}. \tag{1}$$

Hence the probability of A winning the game when ahead 30–15 is given by

$$p^2 + 2p^2q + 3pq^2\frac{p^2}{p^2 + q^2}.$$

Such calculations have previously been given by Croucher [2], for example.

We need to determine the corresponding probability from any stage in the game in such a way that it may be readily programmed.

Let the current scoreline in an uncompleted game be specified as x points won by A ($x \geq 0$) and y points won by B ($y \geq 0$), including points beyond deuce. The value of x or y is incremented by 1 as each point is played, until the game is won. Notice that: if $y \leq 2$ then $x \leq 3$; if $y = 3$ then $x \leq 4$; if $y \geq 4$ then $x = y$ (the score is deuce), or $x = y + 1$ (advantage A) or $x = y - 1$ (advantage B). Let $G(x, y)$ denote the probability that A wins the game from this scoreline, and put $G_0 = G(0, 0)$.

If, initially, $y \leq 2$ and A wins the game without reaching deuce, then A must win a further $3 - x$ points as well as the final point of the game, with B winning at most a further $2 - y$ points. The probability of this is

$$G_1(x, y) = \sum_{k=0}^{2-y} \binom{3 - x + k}{k} p^{4-x} q^k.$$

If, still for $y \leq 2$, the game first reaches deuce then A wins a further $3 - x$ points and B wins a further $3 - y$ points. The probability of this happening is

$$G_2(x, y) = \binom{6 - x - y}{3 - y} p^{3-x} q^{3-y},$$

and the probability that A then wins the game is $G_2(x, y)d$.

Values of x and y	Value of $G(x,y)$
$y \le 2$ (so $x \le 3$)	$G_1(x,y) + G_2(x,y)d$
$y = 3$ and $x \le 3$	$G_2(x,3)d$
$y \ge 3$ and $x = y+1$	$p + qd$
$y \ge 4$ and $x = y$	d
$y \ge 4$ and $x = y-1$	pd

Table 1: The probability $G(x,y)$ that A wins a game, given the number of points x won by A and the number y won by B.

Take $x = 2$ and $y = 1$ to reconstruct the example above in which A is leading 30–15. Take $x = y = 0$ to determine G_0.

The calculations are easier when $y \ge 3$. The complete situation is summarised in Table 1.

When the games in a set reach six all, a (twelve-point) tiebreaker game is played. For us, the only effective differences between a tiebreaker game and an ordinary game are that the tiebreaker is played to seven points, rather than four, with "deuce" reached after six points all, rather than three points all. Let $B(x,y)$ be the probability that A wins a tiebreaker game, given that A has already won x points ($x \ge 0$) and B has won y points ($y \ge 0$). Put $B_0 = B(0,0)$ and

$$B_1(x,y) = \sum_{k=0}^{5-y} \binom{6-x+k}{k} p^{7-x} q^k,$$

$$B_2(x,y) = \binom{12-x-y}{6-y} p^{6-x} q^{6-y}.$$

Then, in the same way as we obtained Table 1, the values for $B(x,y)$ are given in Table 2. The expression for d in equation (1) is used again; it is the probability that A wins the tiebreaker game from six points all.

Values of x and y	Value of $B(x,y)$
$y \le 5$ (so $x \le 6$)	$B_1(x,y) + B_2(x,y)d$
$y = 6$ and $x \le 6$	$B_2(x,6)d$
$y \ge 6$ and $x = y+1$	$p + qd$
$y \ge 7$ and $x = y$	d
$y \ge 7$ and $x = y-1$	pd

Table 2: The probability $B(x,y)$ that A wins a tiebreaker game, given the number of points x won by A and the number y won by B.

3 Conditional probabilities of winning a set

Suppose, in an uncompleted set, that A has won g games ($g \ge 0$), B has won h games ($h \ge 0$), and, in the current game of that set, A has won x points and B has won y points. Let $S(g,h;x,y)$ be the probability that A wins the set. If $g = h = 6$, then the current game is a tiebreaker.

In the manner of an odometer, the point counters x and y are put back to zero with each new game and the game counters g and h are incremented by one, according as who wins the game, A or B, respectively.

Notice that: if $h \le 4$ then $g \le 5$; if $h = 5$ then $g \le 6$; if $h = 6$ then $g = 5$ or 6.

Recall that the probability that A wins a game (not a tiebreaker) from love all is G_0. Put $H_0 = 1 - G_0$; this is the probability that B wins a game. The probability that A wins a tiebreaker game is B_0.

Suppose initially that $x = y = 0$.

If $h \leq 4$ and A wins the set other than 7–5 or 7–6, then A must win a further $5 - g$ games as well as the final game of the set, with B winning at most a further $4 - h$ games. The probability of this is

$$S_1(g,h) = \sum_{k=0}^{4-h} \binom{5-g+k}{k} G_0^{6-g} H_0^k.$$

If A wins the set 7–5, then A must win a further $5 - g$ games and B a further $5 - h$ games, so that the score reaches 5–5, and then A must win the next two games. The probability of this is

$$S_2(g,h) = \binom{10-g-h}{5-h} G_0^{7-g} H_0^{5-h}.$$

If A wins the set 7–6, then the score must first reach 5–5, then A and B must win one game each (there are two ways to do this), and then A must win the tiebreaker game. The probability of this is

$$S_3(g,h) = 2\binom{10-g-h}{5-h} G_0^{6-g} H_0^{6-h} B_0.$$

When $h \leq 4$, we therefore have

$$S(g,h;0,0) = S_1(g,h) + S_2(g,h) + S_3(g,h).$$

It is easy now to complete Table 3, giving the probability that A wins a set after any number of completed games in the set.

Values of g and h	Value of $S(g,h;0,0)$
$h \leq 4$ (so $g \leq 5$)	$S_1(g,h) + S_2(g,h) + S_3(g,h)$
$h = 5$ and $g \leq 5$	$S_2(g,5) + S_3(g,5)$
$h = 5$ and $g = 6$	$G_0 + H_0 B_0$
$h = 6$ and $g = 5$	$G_0 B_0$
$h = 6$ and $g = 6$	B_0

Table 3: The probability $S(g,h;0,0)$ that A wins a set, given the number of games g won by A and the number h won by B, with no points played of the subsequent game.

Now consider the more general situation in which $x + y > 0$, and put $H(x,y) = 1 - G(x,y)$. We must determine all the variations of the following basic result, which is easily seen to be true. When $h \leq 4$ and $g \leq 4$, then

$$S(g,h;x,y) = G(x,y)S(g+1,h;0,0) + H(x,y)S(g,h+1;0,0).$$

One such variation, for example, is if $h \leq 4$ and $g = 5$. In that case,

$$S(g,h;x,y) = G(x,y) + H(x,y)S(g,h+1;0,0),$$

since A wins the set if B loses the current game.

All the possibilities are given in Table 4.

4 Conditional probabilities of winning the match

Let $M(s,t;g,h;x,y)$ be the probability that A wins the "best of three tiebreaker sets" match, when A has won s sets and B t sets ($s = 0$ or 1, $t = 0$ or 1), with g, h, x and y having their previous significance

Values of g and h	Value of $S(g,h;x,y)$
$h \le 4$ and $g \le 4$	$G(x,y)S(g+1,h;0,0)+H(x,y)S(g,h+1;0,0)$
$h \le 4$ and $g = 5$	$G(x,y)+H(x,y)S(5,h+1;0,0)$
$h = 5$ and $g \le 4$	$G(x,y)(S_2(g+1,5)+S_3(g+1,5))$
$h = 5$ and $g = 5$	$G(x,y)(G_0+H_0B_0)+H(x,y)G_0B_0$
$h = 5$ and $g = 6$	$G(x,y)+H(x,y)B_0$
$h = 6$ and $g = 5$	$G(x,y)B_0$
$h = 6$ and $g = 6$	$B(x,y)$

Table 4: The probability $S(g,h;x,y)$ that A wins a set, given the number g of games already won by A and the number h already won by B, and the number x of points won by A and the number y won by B in the current game $(x+y>0)$.

Values of s and t	Value of $M(s,t;g,h;x,y)$
$t = s = 0$	$S(g,h;x,y)(S_0+T_0S_0)+T(g,h;x,y)S_0^2$
$t = 0$ and $s = 1$	$S(g,h;x,y)+T(g,h;x,y)S_0$
$t = 1$ and $s = 0$	$S(g,h;x,y)S_0$
$t = s = 1$	$S(g,h;x,y)$

Table 5: The probability $M(s,t;g,h;x,y)$ that A wins the match, given the number s of sets already won by A, and the number t won by B, the number g of games won by A and the number h won by B in the current set, and the number x of points won by A in the current game of the current set and the number y won by B.

for the current set. The latter are all put back to zero with each new set, and s or t is incremented by one, depending on who won the previous set, A or B, respectively.

Put $T(g,h;x,y) = 1 - S(g,h;x,y)$, the probability that B wins the current set, and put $S_0 = S(0,0;0,0)$ and $T_0 = 1 - S_0$.

It is easy to arrive at the results in Table 5. This is our final table, giving, by reference back to the earlier tables, the probability of winning a match from any stage of the match.

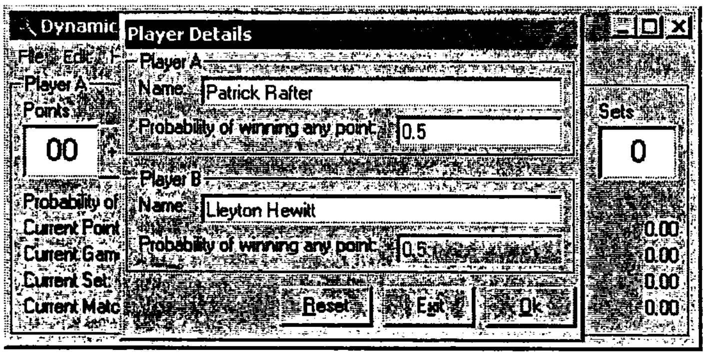

Figure 1: Player details entry screen.

5 Computer implementation

A computer application has been written in Visual Basic to accompany this paper. The program calculates the current game, set and match probabilities of winning for each player in a game of tennis ("best

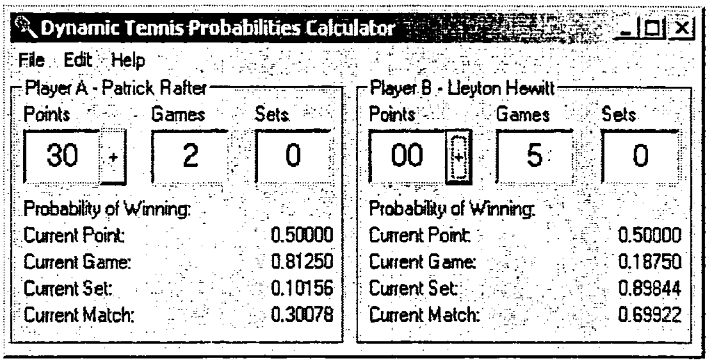

Figure 2: Main screen.

of three tiebreaker sets"), and recalculates them every time the score changes. The program may be freely downloaded from the following web site: http://www.progsoc.org/~shaman/dynamictennis.html.

When the program starts, it asks for the details of the players: the user must enter each player's name and the first player's probability of winning a point. The details screen is shown in Figure 1.

The user is then presented with the main program screen, shown in Figure 2. This shows the current score of each player, as well as each player's chance of winning the game, set and match. The user clicks on the "+" buttons to change the score, incrementing that player's point score by 1 (in the terms of the above analysis). Each time the score is altered the probabilities for each player, other than that for the current point, are recalculated.

If, at any stage of the match, the user wishes to modify the estimate of a player's chance of winning each point, the user can click on the "Edit" menu and choose "Probabilities". The "Edit" menu also provides the ability to "Undo" the last entry. This option can also be accessed using the standard Ctrl-Z keyboard shortcut.

For other forms of experimentation (or if you arrived late), it is possible for users to set the points, games and sets won by both players to values of their choosing using the "Scoreline" option in the "Edit" menu. An example of this screen is shown in Figure 3: any of the windows may be altered. If an impossible scoreline is entered, such as a games score of 6–3 for a set in progress, then an invalid score message of some form is received. There are restrictions on the use of this feature for scorelines that involve tiebreaker games, and in certain other cases.

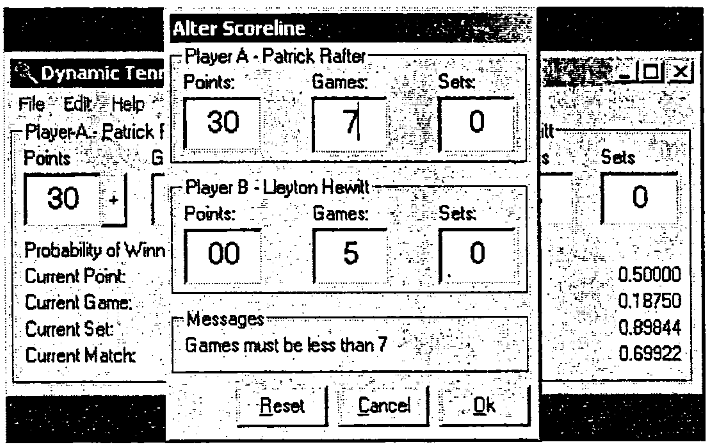

Figure 3: Alter scoreline screen.

References

[1] W. H. Carter, Jr and S. L. Crews, "An analysis of the game of tennis", *Amer. Statist.*, **28** (1974), 130–134.

[2] J. S. Croucher, "The conditional probability of winning games of tennis", *Res. Quart. for Exercise and Sport*, **57** (1986), 23–26.

[3] F. J. G. M. Klaassen and J. R. Magnus, "Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model", *J. Amer. Statist. Assoc.*, **96** (2001), 500–509.

[4] G. H. Pollard, "An analysis of classical and tie-breaker tennis", *Austral. J. Statist.*, **25** (1983), 496–505.

Proceedings of the Sixth Australian Conference on MATHEMATICS AND COMPUTERS IN SPORT

Bond University
Queensland

Edited by
Graeme Cohen and Tim Langtry
Department of Mathematical Sciences
Faculty of Science
University of Technology, Sydney

6M&CS

1 - 3 July 2002

Typeset by $\LaTeX\,2_\varepsilon$

Proceedings of the Sixth Australian Conference on Mathematics and Computers in Sport, 1–3 July 2002, at Bond University, Gold Coast, Queensland 4229, Australia.

ISBN 1 86365 532 8

Printed by UTS Printing Services

Contents

Abstract

Conference Director's Report

Welcome to the Sixth Australian Conference on Mathematics and Computers in Sport. This year we return to Bond University after the successful Fifth Conference in 2000 held in Sydney because of the Olympic Year in that city. It is a pleasure to renew acquaintances with one of our principal speakers, Ray Stefani who was at the First Conference in 1992, and has been collaborating with Steve Clarke from Swinburne for many years now. Our second principal speaker, Steve Gray from Queensland, will add a new dimension to the Sixth Conference with his interest in the economical aspects of sport. Graeme Cohen (now retired from UTS) is our third principal speaker, but he and Tim Langtry have once again taken over the responsibilities of producing the printed *Proceedings*. I thank Graeme and Tim for relieving me of this major task.

The conference has once again attracted academics from New Zealand, the United Kingdom, the United States and Canada. I welcome them all, including many familiar faces. I hope that you all find the conference rewarding in many aspects, including the content and presentation of the talks, the many discussions that are generated, and the close social contact with like-minded academics. We now have a website www.anziam.org.au/mathsport due to Elliot Tonkes, who is the webmaster. This site contains information about this and all previous conferences.

All the papers in these *Proceedings* have undergone a detailed refereeing process. I am indebted to the referees for their time and comments to improve the quality of all papers. The *Proceedings* begin with the papers of our principal speakers, followed by the contributed papers in alphabetical order of author, or first author. The *Proceedings* conclude with an abstract — the paper was accepted for presentation but was received too late to be submitted to the full refereeing process.

Neville de Mestre
June 2002