Expected length and probability of winning a tennis game

CURTIS COOPER and ROBERT E. KENNEDY

1. *Introduction*

The game of tennis has provided mathematicians with many interesting problems. In [1], the problem of finding the probability that a certain player wins a tennis tournament was studied. Gale [2] determined the best serving strategy in tennis. First, we assume Alice and Bob play a game of tennis using the standard (or Deuce/Ad) scoring system, without a tiebreaker, and that Alice serves the game. We also assume that the probability that Alice wins any point she serves is p. Stewart [3] proved that the probability that Alice wins is

$$\frac{15p^4 - 34p^5 + 28p^6 - 8p^7}{1 - 2p + 2p^2}.$$

We will determine the number of points Alice and Bob can expect to play in this game. In answering this question, we will integrate the ideas of states, probability matrices, the Cayley-Hamilton theorem, generating functions, and infinite series. Hence the way we obtain our answer is as important as our answer to the question. Another example of the use of this technique can be found in [4].

Later, we will suppose that Alice and Bob play a 9-point tiebreaker. We will assume that the probability that Alice wins any point she serves in the tiebreaker to be p and that the probability that Bob wins any point he serves in the tiebreaker to be q. We will calculate the expected length of the tiebreaker and the probability that Alice wins the tiebreaker.

2. *The states during a game*

A game consists of a sequence of points played between two players, Alice, the server in this game, and Bob. The first person to win four points and win by two points wins the game. The scoring system can be represented by a state diagram. A score of no points is 0, one point is 15, two points is 30, and three points is 40. Thus, if three or fewer points have been played, we have possible scores of 0-0, 15-0, 0-15, 30-0, 15-15, 0-30, 40-0, 30-15, 15-30, 0-40; the server's score is always given first. Each of these scores represents a state of the game. As points are played, the game moves from state to state. Once four points have been played, we either have a winner (the tennis game has reached the Game state) or we have possible scores of 40-15, 15-40, and 30-30. The 40-15 score represents a state and so does the 15-40 score. From the 40-15 or 15-40 state, the next point will either determine a winner or a score of 40-30 or 30-40. The 40-30 score is said to have entered the Ad-in state, since then the server can win the game by winning just the next point. Similarly the 30-40 score has reached the

Ad-out state. A score of 30-30 is in the Deuce state. From the Ad-in or Ad-out state, the next point either takes us to the Game state or the Deuce state. From the Deuce state, the next point results in the Ad-in or Ad-out state. Thus a game consists of playing points and moving from state to state, until we reach the Game state. As a result, we have the state diagram (Figure 1) which shows all possible states and transitions in a game, starting at the state 0-0 and ultimately ending in the Game state.

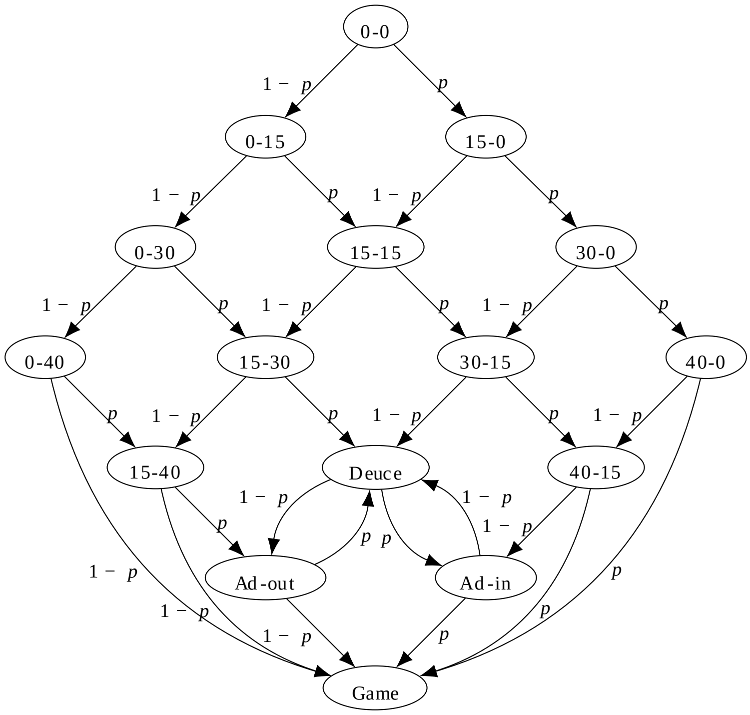

FIGURE 1: Transitions between states

The arrows between states denote the probability that Alice or Bob win a point in the process of playing the Deuce/Ad tennis game. We assume that the probability that Alice wins any point is p and the probability that Bob wins any point is $1 - p$.

3. *Expected length of a game*

Since we are interested in calculating the expected length of a game, we use the state diagram to obtain the probability matrix T; the entries of the matrix give the probability that we can go from one state to another state. The rows and columns of the probability matrix represent states in the game. We number the states in the diagram from top to bottom and from left to right. There are 16 states in the game diagram. State 0-0 is number 1, state 0-15 is number 2, state 15-0 is 3, state 0-30 is 4, state 15-15 is 5, state 30-0 is 6, etc. Finally, the Game state is number 16. The i, j entry in the matrix denotes the probability of moving from state number i to state number j in the tennis game. For example, since the probability that Alice wins the first point of the game is p, and state 0-0 is number 1 and state 15-0 is number 3, the 1, 3 entry in the T matrix is p. The probability that Bob wins the first point is $1 - p$. Since state 0-0 is 1 and state 0-15 is 2, the 1, 2 entry of the T matrix is $1 - p$. For another example, the probability that Bob wins the point in state 15-30 is $1 - p$, and state 15-30 is 8 and state 15-40 is 11. Therefore, the 8, 11 entry in the T matrix is $1 - p$. Putting all of this into matrix T, we have

$$T = \begin{matrix} 0 & 1-p & p & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-p & p & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p & p & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & p & 0 & 1-p \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1-p & p & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 1-p & p \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & p & 0 & 0 & 0 & 1-p \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1-p & 0 & 0 & 0 & p \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

We note that the probability that a game goes from one state to another state in n points can be found by computing T^n. Next, let c_n be the probability that the game starts at 0-0 and ends after the nth point in the Game state. Therefore c_n is the entry 1, 16 of T^n. To compute c_n, we will use the characteristic polynomial of T and the Cayley-Hamilton Theorem. Based on the structure of the T matrix, the characteristic polynomial of T is

$$\det(xI - T) = x^{16} + 2p(p - 1)x^{14}.$$

Thus, by the Cayley-Hamilton Theorem,

$$T^{16} + 2p(p - 1)T^{14} = O.$$

where, O is the 16×16 zero matrix. Thus

$$T^n + 2p(p - 1)T^{n-2} = O \text{ for } n \geqslant 16. \tag{1}$$

Using (1), we can compute the first few values of c_n.

n	c_n
1	0
2	0
3	0
4	$1 - 4p + 6p^2 - 4p^3 + 2p^4$
5	$4p - 16p^2 + 24p^3 - 12p^4$
6	$10p^2 - 40p^3 + 70p^4 - 60p^5 + 20p^6$
7	0
8	$20p^3 - 100p^4 + 220p^5 - 260p^6 + 160p^7 - 40p^8$
9	0
10	$40p^4 - 240p^5 + 640p^6 - 960p^7 + 840p^8 - 400p^9 + 80p^{10}$
11	0
12	$80p^5 - 560p^6 + 1760p^7 - 3200p^8 + 3600p^9 - 2480p^{10} + 960p^{11} - 160p^{12}$
13	0
14	$160p^6 - 1280p^7 + 4640p^8 - 9920p^9 + 13600p^{10} - 12160p^{11}$ $+6880p^{12} - 2240p^{13} + 320p^{14}$
15	0
16	$320p^7 - 2880p^8 + 11840p^9 - 29120p^{10} + 47040p^{11} - 51520p^{12}$ $+38080p^{13} - 18240p^{14} + 5120p^{15} - 640p^{16}$

TABLE 1: Initial values of c_n

Now the expected length of the tennis game is the quantity

$$\sum_{n=1}^{\infty} n \times c_n. \tag{2}$$

To calculate (2), we first determine the generating function of c_n,

$$f(x) = \sum_{n=1}^{\infty} c_n \times x^n,$$

and then use the fact that

$$\sum_{n=1}^{\infty} n \times c_n = f'(1).$$

Noting that

$$2p(p-1)x^2 f(x) = \sum_{n=1}^{\infty} 2p(p-1)c_n x^{n+2},$$

we have

$$f(x)\left(1 + 2p(p-1)x^2\right) = \sum_{n=1}^{\infty} c_n\left(x^n + 2p(p-1)x^{n+2}\right).$$

By the recurrence relation given by (1), it follows that

$$f(x) = \frac{c_1 x + c_2 x + \sum_{n=3}^{15} (c_n + 2p(p-1)c_{n-2})x^n}{1 + 2p(p-1)x^2}.$$

Examining the terms involving c and simplifying, we obtain

$$f(x) = \frac{c_4 x^4 + c_5 x^5 + (c_6 + 2p(p-1)c_6)x^6 + 2p(p-1)c_5 x^7}{1 + 2p(p-1)x^2}.$$

and so

$$f'(1) = \frac{4 - 12p + 20p^2 + 8p^3 - 112p^4 + 264p^5 - 312p^6 + 192p^7 - 48p^8}{1 - 4p + 8p^2 - 8p^3 + 4p^4}$$

$$= 4\,\frac{1 - p + p^2 + 6p^3 - 18p^4 + 18p^5 - 6p^6}{1 - 2p + 2p^2}.$$

We define this expected length function as

$$F(p) = 4\,\frac{1 - p + p^2 + 6p^3 - 18p^4 + 18p^5 - 6p^6}{1 - 2p + 2p^2}$$

on the interval [0, 1]. Then $F(0) = F(1) = 4$, $F(0.5) = 6.75$ and $F(1-p) = F(p)$. In addition, the values of F at $p = 0$ and $p = 1$ are the absolute minimums and the value of F at $p = 0.5$ is the absolute maximum. For $p = 0.5$, we can interpret this result by saying that the expected game length between two evenly matched players is 6.75 points. The function is increasing on [0, 0.5] and decreasing on [0.5, 1]. A graph of the expected length function is shown in Figure 3.

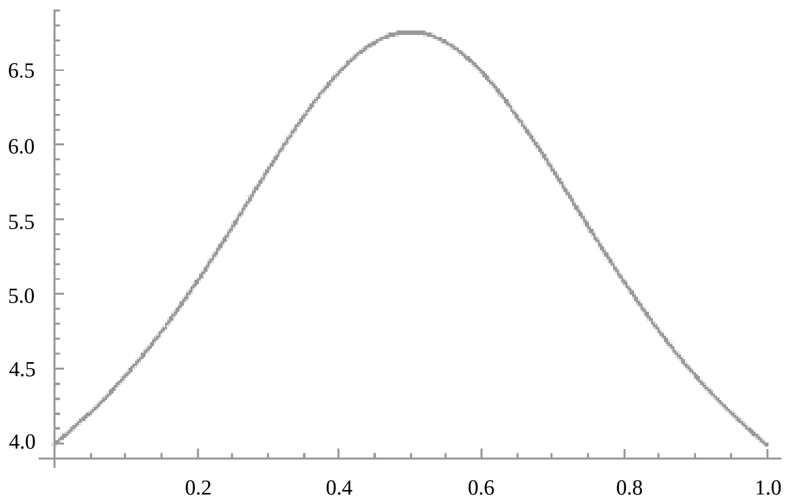

FIGURE 2: Expected length of a Deuce/Ad tennis game

4. *9-point tiebreaker*

A 9-point tiebreaker in tennis adds a new dimension to our state diagram, since there are only a finite number of points needed to determine a winner and both Alice and Bob get to serve. If Alice serves first, she serves the first 2 points, Bob serves the next 2 points, Alice serves the next 2 points, and Bob serves the final 3 points, assuming the tiebreaker lasts that long. The winner of the tiebreaker is the first player to win 5 points. So the 9-point tiebreaker could last as few as 5 points or last as long as 9 points; in the latter case, the penultimate score would be 4 points to 4 points and the winner would be the winner of the final point.

We now draw the state diagram for the 9-point tiebreaker (see Figure 3). We will assume Alice serves first and the probability that Alice wins the points she serves is p and the probability that Bob wins the points he serves is q. In the state diagram, Alice's score is given first. For example, suppose the score is 2-3. In state 2-3, Alice has 2 points and Bob has 3 points and Alice would be serving. The arrows coming out of state 2-3 would have labels p and $1 - p$, for the probabilities of Alice or Bob winning the 6th point, respectively. After the 6th point, the score would be either 3-3 or 2-4, depending on whether Alice or Bob won the 6th point. For another example, suppose the score is 4-3. In state 4-3, Alice has 4 points and Bob has 3

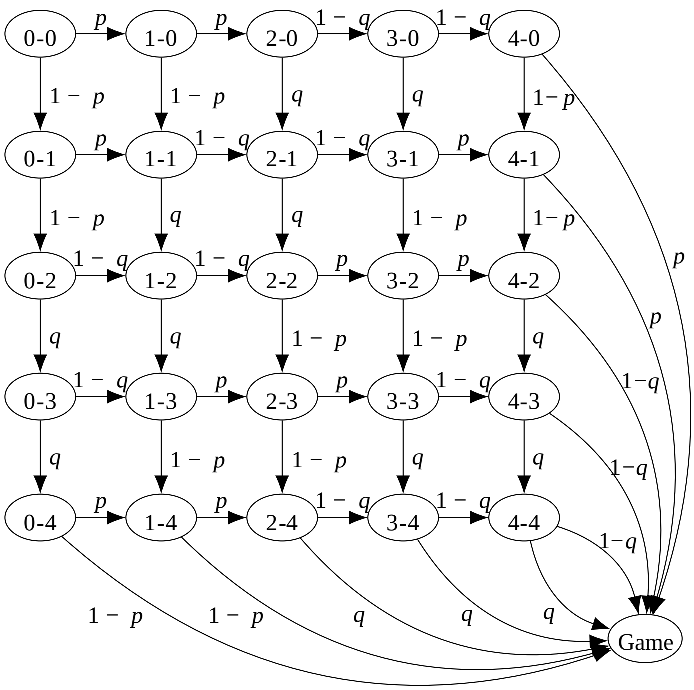

FIGURE 3: 9-point tiebreaker

points and Bob would be serving. The arrows coming out of state 4-3 would have labels q and $1 - q$, for the probabilities of the 8th point being won by Bob or Alice, respectively. After the 8th point, the score would be either 4-4 or Game, depending on whether Bob or Alice won the 8th point. The initial state of the 9-point tiebreaker is 0-0 and the final state of the 9-point tiebreaker is Game.

Next we give the probability matrix for the 9-point tiebreaker. The rows and columns of the probability matrix represent states in the 9-point tiebreaker. There are 26 states in the 9-point tiebreaker diagram. We number the states in the following order: state 0-0 is 1, state 1-0 is 2, state 0-1 is 3, state 2-0 is 4, state 1-1 is 5, state 0-2 is 6, etc. State Game is 26. The i, j entry in the matrix denotes the probability of moving from state number i to state number j in the tennis game. For example, if the score is 0-1, then Alice is serving. The probability that Alice wins this point is p. And state 0-1 is number 3 and state 1-1 is number 5. So the 3, 5 entry in the T matrix is p. For another example, if the score is 2-0, then Bob is serving. The probability that Bob wins this point is q. And state 2-0 is 4 and state 2-1 is 8. So the 4, 8 entry in the T matrix is q. For a final example, if the score is 2-4, then Bob is serving. The probability that Alice wins the point in the state 2-4 is $1 - q$, and state 2-4 is 22 and state 3-4 is 24. Therefore, the 22, 24 entry in the T matrix is $1 - q$. Putting all of this into matrix T, we have

$$T = \begin{matrix}
0 & p & 1-p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & p & 1-p & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p & 1-p & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1-q & q & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-q & q & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-q & q & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & q & 1-q \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1-q & q \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0
\end{matrix}$$

In the 9-point tiebreaker system, we can compute the expected length of a 9-point tiebreaker and Alice's chance of winning a 9-point tiebreaker by a direct, all-be-it large, calculation using our state diagram and T matrix. To tackle the expected length question, we see that the expected length of the 9-point tiebreaker is the quantity

$$5 \times (T^5)_{1,26} + 6 \times (T^6)_{1,26} + 7 \times (T^7)_{1,26} + 8 \times (T^8)_{1,26} + 9 \times (T^9)_{1,26}.$$

The probability that Alice wins the 9-point tiebreaker is the quantity

$$p \times (T^4)_{1,11} + p \times (T^5)_{1,16} + (1-q) \times (T^6)_{1,20} + (1-q) \times (T^7)_{1,23} + (1-q) \times (T^8)_{1,25}.$$

Using *Mathematica* on a Raspberry Pi, we can compute all of these quantities. We start with the table of entries for certain powers of T.

i	j	n	$(T^n)_{i,j}$
1	26	5	$p^3 - 2p^3q + q^2 - 3pq^2 + 3p^2q^2$
1	26	6	$3p^3 - 3p^4 + 2q - 8pq + 12p^2q - 14p^3q + 10p^4q - 2q^2 + 11pq^2 - 21p^2q^2 + 20p^3q^2 - 10p^4q^2$
1	26	7	$6p^2 - 12p^3 + 6p^4 + q - 4pq - 12p^2q + 40p^3q - 25p^4q - 2q^2 + 16pq^2 - 18p^2q^2 - 20p^3q^2 + 25p^4q^2 + q^3 - 12pq^3 + 30p^2q^3 - 20p^3q^3$
1	26	8	$4p - 12p^2 + 12p^3 - 4p^4 + q - 20pq + 72p^2q - 88p^2q + 35p^4q - 3q^2 + 48pq^2 - 180p^2q^2 + 240p^3q^2 - 105p^4q^2 + 3q^3 - 52pq^3 + 210p^2q^3 - 300p^3q^3 + 140p^4q^3 - q^4 + 20pq^4 - 90p^2q^4 + 140p^3q^4 - 70p^4q^4$
1	26	9	$1 - 4p + 6p^2 - 4p^3 + p^4 - 4q + 32pq - 72p^2q + 64p^3q - 20p^4q + 6q^2 - 72pq^2 + 216p^2q^2 - 240p^3q^2 + 90p^4q^2 - 4q^3 + 64pq^3 - 240p^2q^3 + 320p^3q^3 - 140p^4q^3 + q^4 - 20pq^4 + 90p^2q^4 - 140p^3q^4 + 70p^4q^4$
1	11	4	$p^2 - 2p^2q + p^2q^2$
1	16	5	$3p^2 - 3p^3 - 6p^2q + 8p^3q + 3p^2q^2 - 5p^3q^2$
1	20	6	$6p^2 - 12p^3 + 6p^4 - 12p^2q + 32p^3q - 20p^4q + 6p^2q^2 - 20p^3q^2 + 15p^4q^2$
1	23	7	$4p - 12p^2 + 12p^3 - 4p^4 - 12pq + 54p^2q - 72p^3q + 30p^4q + 12pq^2 - 72p^2q^2 + 120p^3q^2 - 60p^4q^2 - 4pq^3 + 30p^2q^3 - 60p^3q^3 + 35p^4q^3$
1	25	8	$1 - 4p + 6p^2 - 4p^3 + p^4 - 4q + 32pq - 72p^2q + 64p^3q - 20p^4q + 6q^2 - 72pq^2 + 216p^2q^2 - 240p^3q^2 + 90p^4q^2 - 4q^3 + 64pq^3 - 240p^2q^3 + 320p^3q^3 - 140p^4q^3 + q^4 - 20pq^4 + 90p^2q^4 - 140p^3q^4 + 70p^4q^4$

TABLE 2: Entries of certain powers of T

Using the table of entries of powers of T, we find that the expected length of a 9-point tiebreaker between Alice (serving first with probability p of winning her service points) and Bob (serving second with probability q of winning his service points) is

$$9 - 4p - p^3 + p^4 - 9q + 52pq - 84p^2q + 58p^3q - 15p^4q + 9q^2$$
$$-101pq^2 + 267p^2q^2 - 260p^3q^2 + 85p^4q^2 - 5q^3 + 76pq^3 - 270p^2q^3$$
$$+340p^3q^3 - 140p^4q^3 + q^4 - 20pq^4 + 90p^2q^4 - 140p^3q^4 + 70p^4q^4.$$

In addition, the probability that Alice wins the 9-point tiebreaker is

$$1 - 5q + 20pq - 30p^2q + 20p^3q - 5p^4q + 10q^2 - 80pq^2 + 180p^2q^2 - 160p^3q^2$$
$$+ 50p^4q^2 - 10q^3 + 120pq^3 - 360p^2q^3 + 400p^3q^3 - 150p^4q^3 + 5q^4 - 80pq^4$$
$$+ 300p^2q^4 - 400p^3q^4 + 175p^4q^4 - q^5 + 20pq^5 - 90p^2q^5 + 140p^3q^5 - 70p^4q^5.$$

If Alice and Bob are evenly matched (i.e. $p = q$), then the expected length of the 9-point tiebreaker is

$$9 - 13p + 61p^2 - 191p^3 + 403p^4 - 565p^5 + 515p^6 - 280p^7 + 70p^8.$$

We define the expected length of a 9-point tiebreaker between two evenly matched players as

$$G(p) = 9 - 13p + 61p^2 - 191p^3 + 403p^4 - 565p^5 + 515p^5 - 280p^7 + 70p^8$$

on the interval [0, 1]. Then $G(0) = G(1) = 9$, $G(0.5) = 965/128$ and $G(1 - p) = G(p)$. In addition, the values of G at $p = 0$ and $p = 1$ are the absolute maximums and the value of G at $p = 0.5$ is the absolute minimum. For $p = 0.5$, we can interpret this result by saying that the expected length of a 9-point tiebreaker between two evenly matched players is 7.5390625. The function is decreasing on [0, 0.5] and increasing on [0.5, 1]. The graph of this function between $0 \leqslant p \leqslant 1$ is shown in Figure 4.

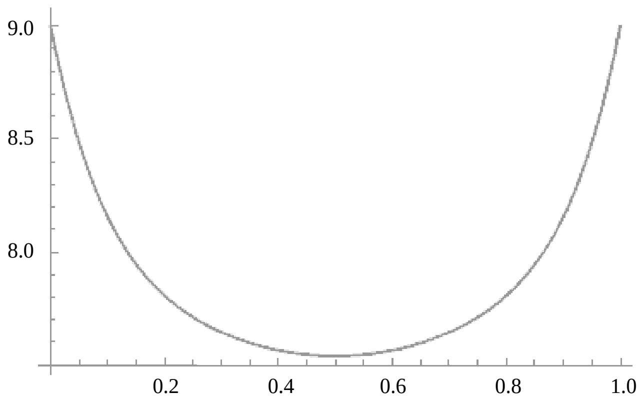

FIGURE 4: Length of 9-point tiebreaker

And, if Alice and Bob are evenly matched ($p = q$) and Alice serves first in the 9-point tiebreaker, then the probability that Alice wins the 9-point tiebreaker is

$$1 - 5p + 30p^2 - 120p^3 + 325p^4 - 606p^5 + 770p^6 - 640p^7 + 315p^8 - 70p^9.$$

We define the probability that Alice serves first and wins a 9-point tiebreaker between two evenly matched players as

$$H(p) = 1 - 5p + 30p^2 - 120p^3 + 325p^4 - 606p^5 + 770p^6 - 640p^7 + 315p^8 - 70p^9.$$

on the interval [0, 1]. Then $H(0) = 1$, $H(1) = 0$, $H(0.5) = 0.5$ and $H(1 - p) = 1 - H(p)$. In addition, the value of H at $p = 0$ is an absolute maximum and the value of H at $p = 1$ is an absolute minimum. An interesting observation in this situation ($p = q$) is that Alice's chances of winning the 9-point tiebreaker are better than Bob's if $0 \leqslant p < 0.5$ and Bob's chances of winning the 9-point tiebreaker are better than Alice's if $0.5 < p \leqslant 1$. In addition, the function is decreasing on [0, 1]. The graph of this function for $0 \leqslant p \leqslant 1$ is shown in Figure 5.

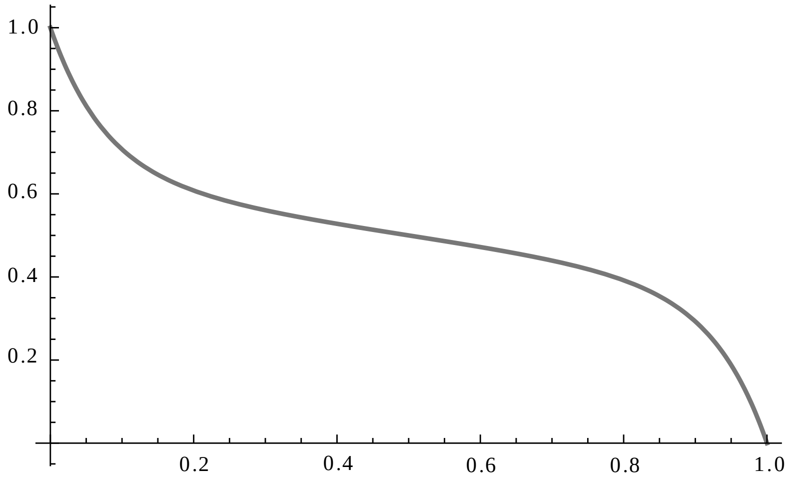

FIGURE 5: Probability Alice wins 9-point tiebreaker

5. *Open questions*

In tennis, a shorthand way of describing the standard scoring system in tennis is to say that the winning player is the first one to win $i \geqslant 4$ points and to win by 2 or more points. To generalise this scoring system, we can say that the winner of a game is the first one to win $i \geqslant n$ points and to win by k or more points. The winner of a 9-point tiebreaker must win $i = 5$ points and win by at least 1 point. There is an alternative scoring system where $n = 4$ and $k = 1$. The 12-point tiebreaker in tennis is the system where $n = 7$ and $k = 2$, and, the 10-point tiebreaker is the system where $n = 10$ and $k = 2$.

It might be interesting to study the expected length of these different scoring systems and the probability that a certain player wins in that scoring system.

Acknowledgement

The authors thank the referee for helpful suggestions which improved the presentation and readability of the paper.

References

1. C. Cooper, A knockout tournament problem, *Crux Mathematicorum*, **8** (1982) pp. 93-96.
2. D. Gale, Optimal strategy for serving in tennis, *Mathematics Magazine*, **44** (1971) pp. 197-199.
3. I. Stewart, *Game, set and math*, Dover (2007).
4. C. Cooper and R. E. Kennedy, A generating function for the distribution of the scores of all possible bowling games, *Mathematics Magazine*, **63** (1990) pp. 239-243.

CURTIS COOPER
School of Computer Science and Mathematics
University of Central Missouri
Warrensburg, MO 64093 USA
e-mail: *cooper@ucmo.edu*
ROBERT E. KENNEDY
School of Computer Science and Mathematics
University of Central Missouri
Warrensburg, MO 64093 USA
e-mail: *kennedy233@gmail.com*

10.1017/mag.2021.117