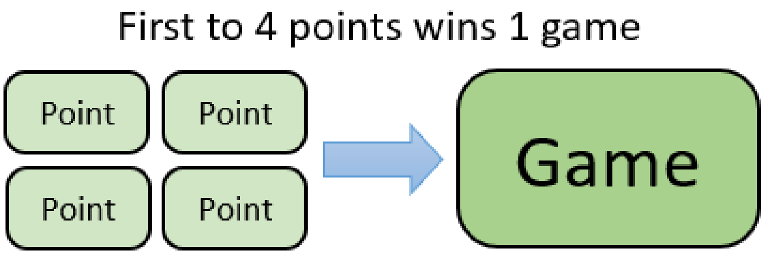
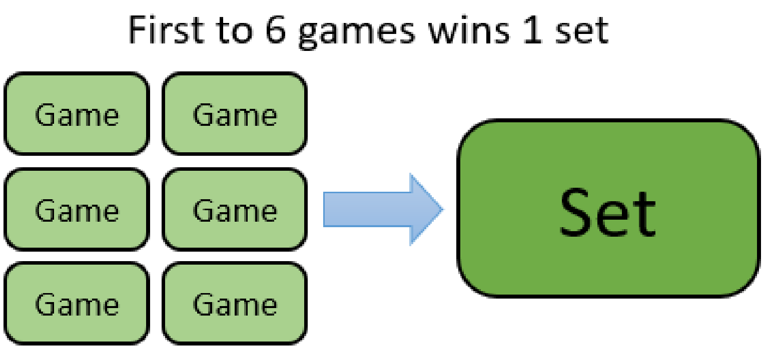
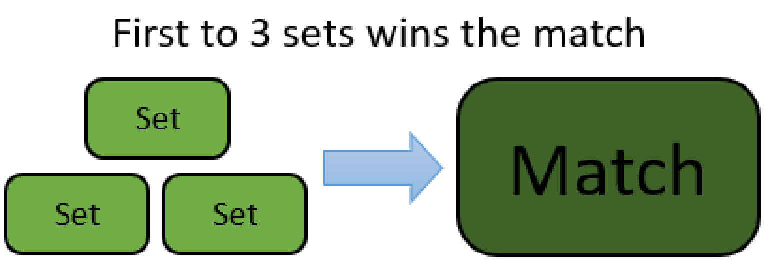
What is the probability of winning a tennis match based on the probability of winning a point?

1 Introduction

In tennis's scoring system, there are 'points' which a player obtains when winning an exchange, you need to be the first to win **four** 'points' to win a 'game', after being the first to win **six** 'games', you win a 'set', and finally you need **three** 'sets' to win the match. This is visualised below.







However to win a game, you need to win by at least a margin of two points. To win a set, you need to win six games by at least a margin of two games. And finally to win a match you need three points (best of five) without a need to win by a margin.

It is also important to mention that tennis has an unorthodox scoring system. In a game, points are counted in this sequence:

$$0\ points = love$$
$$1\ points = 15$$
$$2\ points = 30$$
$$3\ points = 40$$
$$4\ points = game$$

Tennis is a world-renowned sport loved. As someone whose family actively watches Grand Slam tournaments, I thought it would be very interesting to look at its mathematical aspect for this investigation. Tennis is a highly skilled sport which takes a lot of discipline of experience to master, but even then, there is an aspect of probability, even in high-level games. The aim of this investigation is to calculate the probability of someone winning a game based on the probability of them scoring a point. More specifically, to be 90% sure of winning a game, set or match what does the probability of winning a point have to be?

2 Investigation

In this investigation we will be looking at how tennis's score advancing rules affect winning a match.

In tennis, the top players have played a lot of matches, and have all played against each other at least once, this gives us a lot of data that records the probability of one winning a point against the other.

We can call this p.

$$P(point) = p$$

2.1 Important notes

For the simplicity of this investigation, we are negating factors such as fatigue, stamina, psychological pressure, injuries, terrain conditions or any other factors that could affect the probability throughout the match. We are assuming a consistent p for scoring a point throughout the match.

To not overcomplicate, I'll avoid using tennis terms (such as Love or Deuce) to keep it clear and not create confusion.

2.2 The probability of winning a game

To win a game, we need to win 4 points with a margin of 2. To calculate the probability of winning a game, we're going to calculate every way a game can be won and add them together. This can be written as the following equation:

$$P(Game) = P(Game - 0) + P(Game - 15) + P(Game - 30) + P(Game - 40)$$

Or be visualised as a probability tree:

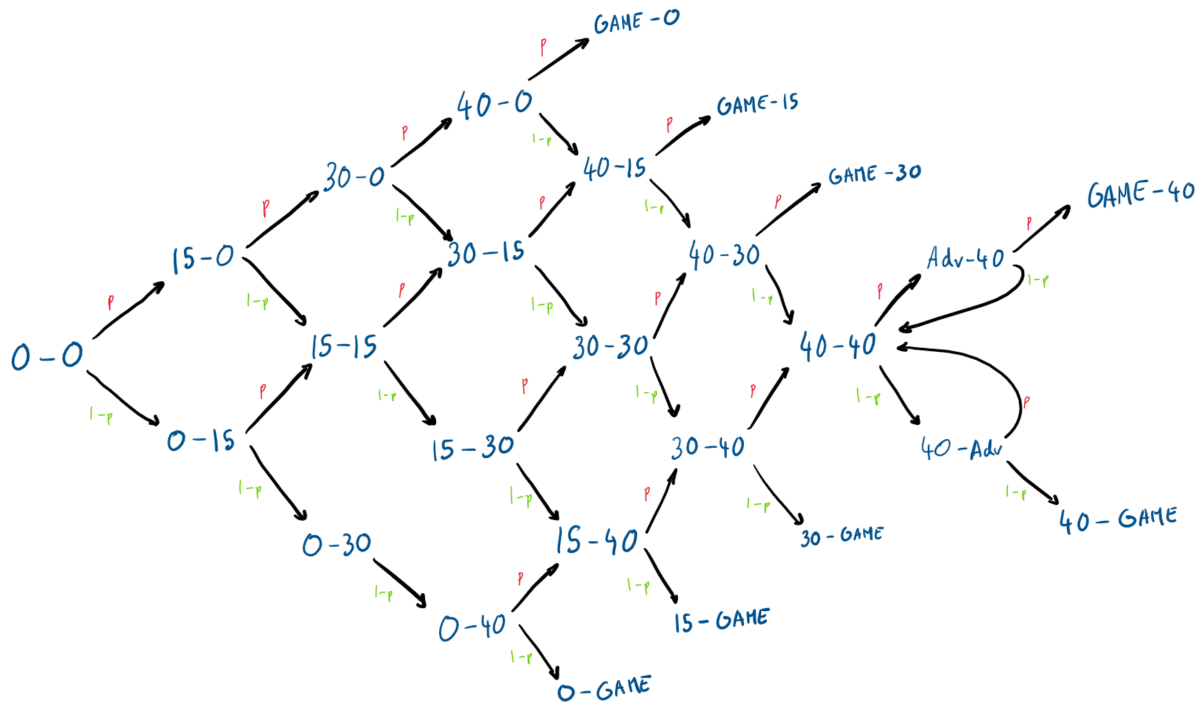


Figure 1 - Probability tree of winning a game

<u>Game – 0</u>

This is when a player wins every exchange, the probability of winning four points is $p \times p \times p \times p$.

$$P(Game - 0) = p^4$$

<u>Game – 15</u>

This means that there are five points won in total, we would need to win the last point and any three of the last four points (for a total of four), whilst our opponent only wins one. Using the probability tree, we can deduce that this outcome can happen $4C1$ ways.

$$P(Game - 15) = 4C1 \times p^4 \times (1 - p)^1$$

$$P(Game - 15) = 4p^4(1-p)$$

Game – 30

Like the one above, this means there are six points won in total, we would need to win the last point and any three of the last five points (for a total of four), whilst our opponent wins two points. This can happen in $5C2$ ways.

$$P(Game - 30) = 5C2 \times p^4 \times (1-p)^2$$

$$P(Game - 30) = 10p^4(1-p)^2$$

Game – 40

To find out the probability of winning Game – 40, we need to first find out the probability of winning from 40 – 40, otherwise known as Deuce. This is where tennis's score advancing rules come into play and things get more complicated.

We first need to find the probability of getting to 40 – 40. To get to 40 – 40, there are six points won in total, three from us and three from our opponent, this can happen in $6C3$ ways.

$$P(40 - 40) = 6C3 \times p^3 \times (1-p)^3$$

$$P(40 - 40) = 20p^3(1-p)^3$$

Now we have to find out the probability of us winning from 40 – 40. And since we must win with a margin of

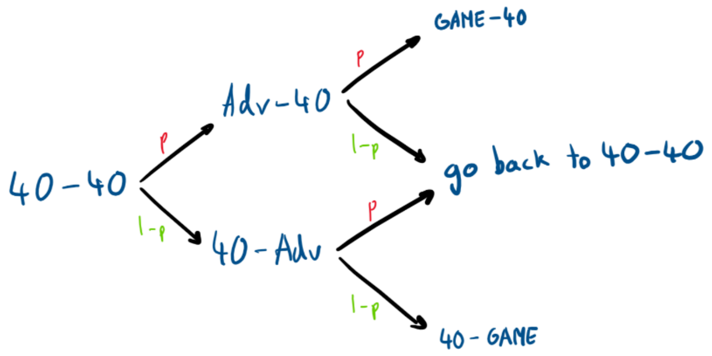


Figure 2 - Probability tree of winning from 40-40

two, this can happen in four ways, as shown below.

To find the probability of winning from 40 – 40, we need to find the probability of each outcome from the probability tree above.

$$P(Game - 40) = p^2$$

$$P(40 - Game) = (1-p)^2$$

$$P(go\ back\ to\ 40-40) = p(1-p) + (1-p)p$$

$$P(go\ back\ to\ 40-40) = 2p(1-p)$$

For us to win, we can either win by scoring twice in a row, or we can win by going back to 40 – 40 and then scoring twice in a row. And since we can go back to 40 – 40 as many times as we like, this situation could go on for a long time.

Thus, the probability of winning from 40 – 40 can be written as.

$$P(winning\ from\ 40-40) = p^2 + 2p(1-p) * P(winning\ from\ 40-40)$$

For the sake of simplicity, we're going to define $P(winning\ from\ 40-40)$ as w.

$$w = p^2 + 2p(1-p) * w$$

We then expand the brackets and solve for w.

$$w = p^2 + 2wp - 2wp^2$$

$$w - 2wp + 2wp^2 = p^2$$

$$w(1 - 2p + 2p^2) = p^2$$

$$w = \frac{p^2}{1 - 2p + 2p^2}$$

Putting $P(winning\ from\ 40-40)$ back into the equation and giving us.

$$P(winning\ from\ 40-40) = \frac{p^2}{1 - 2p + 2p^2}$$

Which means we can finally go back and find out the probability of winning with Game – 40.

$$P(Game-40) = P(40-40) * P(winning\ from\ 40-40)$$

$$P(Game-40) = 20p^3(1-p)^3 * \frac{p^2}{1 - 2p + 2p^2}$$

$$P(Game-40) = \frac{20p^5(1-p)^3}{1 - 2p + 2p^2}$$

Winning a game

Going back to the start, we can now plug these values into our original equation.

$$P(Game) = P(Game-0) + P(Game-15) + P(Game-30) + P(Game-40)$$

$$P(Game) = p^4 + 4p^4(1-p) + 10p^4(1-p)^2 + \frac{20p^5(1-p)^3}{1-2p+2p^2}$$

We can plot this equation to get the following:

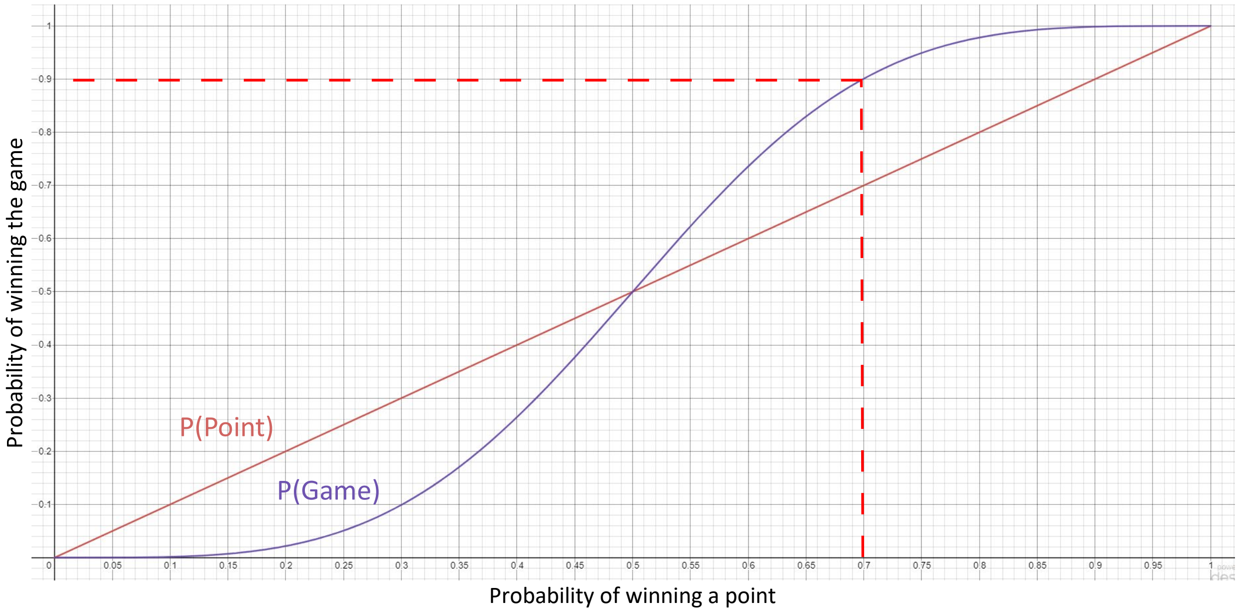


Figure 3 - Graphing the probability of winning a game

The purple line shows the probability of winning a game based on the probability of winning a point. The red line represents $P(Game) = p$ for reference. This graph shows that just being slightly more likely to score a point can have a big difference and make it easier to win the game, for example if we had a probability of 0.7, we would have a 90% chance of winning the game.

2.3 The probability of winning a set

To win a set, we need to have won 6 games with a margin of 2, however differently to how games are played, if the score reaches 6 – 6, then there is a tie break (explained later). Similarly to how we looked at the probability of winning a game, we'll calculate each way a set can be won and adding them all up, this equation can be written as:

$$P(Set) = P(6-0) + P(6-1) + P(6-2) + P(6-3) + P(6-4) + P(7-5) + P(7-6)$$

Or be visualised as a probability tree:

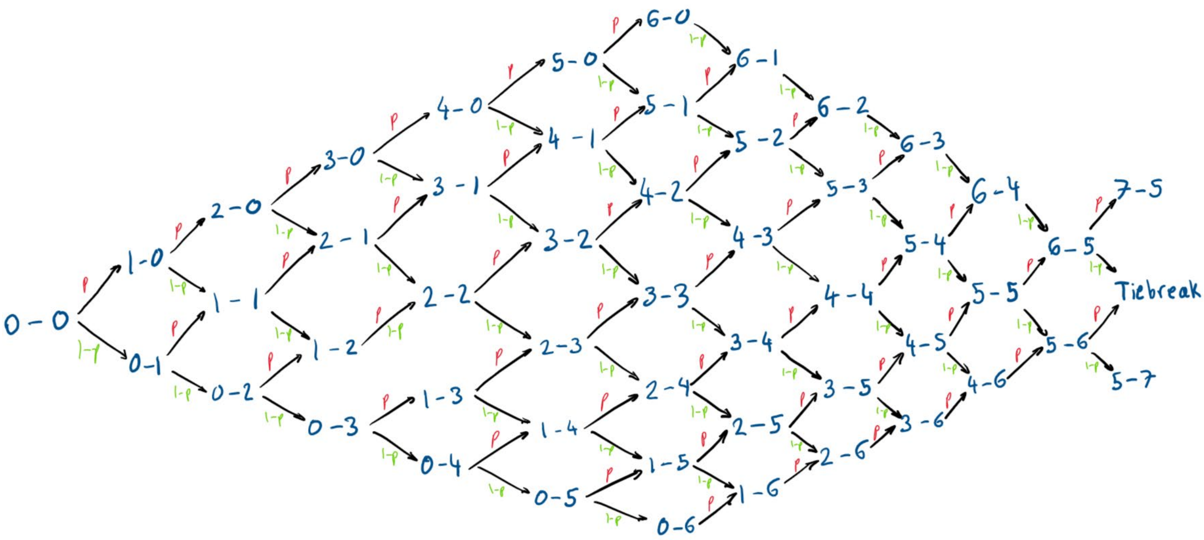


Figure 4 - Probability tree of winning a set

The probability of winning a game, that we calculated above, will be defined as g.

Score of 6 – 0

Again, like winning a game, we would have to win six games in a row to win a set. The probability of that is

$g \cdot g \cdot g \cdot g \cdot g \cdot g$

$$P(6-0) = g^6$$

Score of 6 – 1

This means that there are seven games won in total, we would need to win the last game and any five of the last six games (for a total of six), whilst our opponent only wins one. This can happen in $6C5$ different ways.

$$P(6-1) = 6C1 \times g^6 \times (1-g)^1$$

$$P(6-1) = 6g^6(1-g)$$

Score of 6 – 2

This means there are eight games won in total, we would need to win the last game and any five of the last seven games (for a total of six), whilst our opponent only wins two. This can happen in $7C5$ different ways.

$$P(6-2) = 7C5 \times g^6 \times (1-g)^2$$

$$P(6-2) = 21g^6(1-g)^2$$

Score of 6 – 3

This means there are nine games won in total, we would need to win the last game and any five of the last eight games (for a total of six), whilst our opponent only wins three. This can happen in $8C5$ different ways.

$$P(6-3) = 8C5 \times g^6 \times (1-g)^3$$

$$P(6-3) = 56g^6(1-g)^3$$

Score of 6 – 4

This means there are ten games won in total, we would need to win the last game and any five of the last nine games (for a total of six), whilst our opponent only wins four. This can happen in $9C5$ different ways.

$$P(6-4) = 9C5 \times g^6 \times (1-g)^4$$

$$P(6-4) = 126g^6(1-g)^4$$

Score of 7 – 5

Even though it is first to win six games to win the set, there has to be a margin of two points. This scenario can exclusively only happen if we get to a 5 – 5 score, then after that win the next two games consecutively. If we don't win both games, then the set turns into a tie-break (which will be looked at next).

We first have to find the probability of getting to a score of 5 – 5, which isn't difficult, there are ten games won in total, five from us and five from our opponent. This can happen in $10C5$ different ways.

$$P(5-5) = 10C5 \times g^5 \times (1-g)^5$$

$$P(5-5) = 252g^5(1-g)^5$$

To get to a score of 7 – 5, we then win the next two games consecutively.

$$P(7-5) = g^2 \times P(5-5)$$

$$P(7-5) = g^2(252g^5(1-g)^5)$$

$$P(7-5) = 252g^7(1-g)^5$$

Score of 7 – 6

If a match gets to 6 – 6, a tie-break rule is put in place. Players only play one more game that ultimately determines who takes the set. However, this is a special game as the points are counted using ordinary numbers, and the first player who gets to seven with a margin of two wins.

To do this we first need to calculate the probably of getting a score of 6 – 6 from a score of 5 – 5. Which can happen if we win our first game and lose the second, or vice versa.

$$P(6-6) = 2C1 \times g \times (1-g) * P(5-5)$$

$$P(6-6) = 504g^6(1-g)^6$$

We can define the probability of winning the tie-break as t, which represents all winning outcomes.

$$t = \sum all\ winning\ outcomes$$

$$P(7-6) = 504g^6(1-g)^6 t$$

The probability of winning a tie-break game up to the score of 7 – 5 are as follows.

$$P(7-0) = p^7$$

$$P(7-1) = 7p^7(1-p)$$

$$P(7-2) = 28p^7(1-p)^2$$

$$P(7-3) = 84p^7(1-p)^3$$

$$P(7-4) = 210p^7(1-p)^4$$

$$P(7-5) = 462p^7(1-p)^5$$

The problem is when the tie break score is 6 – 6, to win the game you need a margin of two points. Although we already solved this when we calculated the probability of winning from a game score of 40 – 40, otherwise known as a deuce.

$$P(n+2-n) = P(6-6) \times w$$

$$P(n+2-n) = 924p^6(1-p)^6 \times \frac{p^2}{1-2p+2p^2}$$

$$P(n+2-n) = \frac{924p^8(1-p)^6}{1-2p+2p^2}$$

Then all this sums up to:

$$t = p^7 + 7p^7(1-p) + 28p^7(1-p)^2 + 84p^7(1-p)^3 + 210p^7(1-p)^4 + 462p^7(1-p)^5 + \frac{924p^8(1-p)^6}{1-2p+2p^2}$$

Then putting $P(7-6)$ in terms of p.

$$P(7-6) = 504g^6(1-g)^6 t$$

$$P(7-6) = 504\left(p^4 + 4p^4(1-p) + 10p^4(1-p)^2 + \frac{20p^5(1-p)^3}{1-2p+2p^2}\right)^6 \left(1 - (p^4 + 4p^4(1-p) + 10p^4(1-p)^2 + \frac{20p^5(1-p)^3}{1-2p+2p^2})\right)^6 (p^7 + 7p^7(1-p) + 28p^7(1-p)^2 + 84p^7(1-p)^3 + 210p^7(1-p)^4 + 462p^7(1-p)^5 + \frac{924p^8(1-p)^6}{1-2p+2p^2})$$

Winning a set

Going back to the start, we can now plug these values into our original equation.

$$P(Set) = P(6-0) + P(6-1) + P(6-2) + P(6-3) + P(7-5) + P(7-6)$$

$$P(Set) = g^6 + 6g^6(1-g) + 21g^6(1-g)^2 + 56g^6(1-g)^3 + 126g^6(1-g)^4 + 252g^7(1-g)^5 + 504g^6(1-g)^6 t$$

Where:

$$g = P(Game) = p^4 + 4p^4(1-p) + 10p^4(1-p)^2 + \frac{20p^5(1-p)^3}{1-2p+2p^2}$$

$$t = p^7 + 7p^7(1-p) + 28p^7(1-p)^2 + 84p^7(1-p)^3 + 210p^7(1-p)^4 + 462p^7(1-p)^5 + \frac{924p^8(1-p)^6}{1-2p+2p^2}$$

We can plot this equation to get the following:

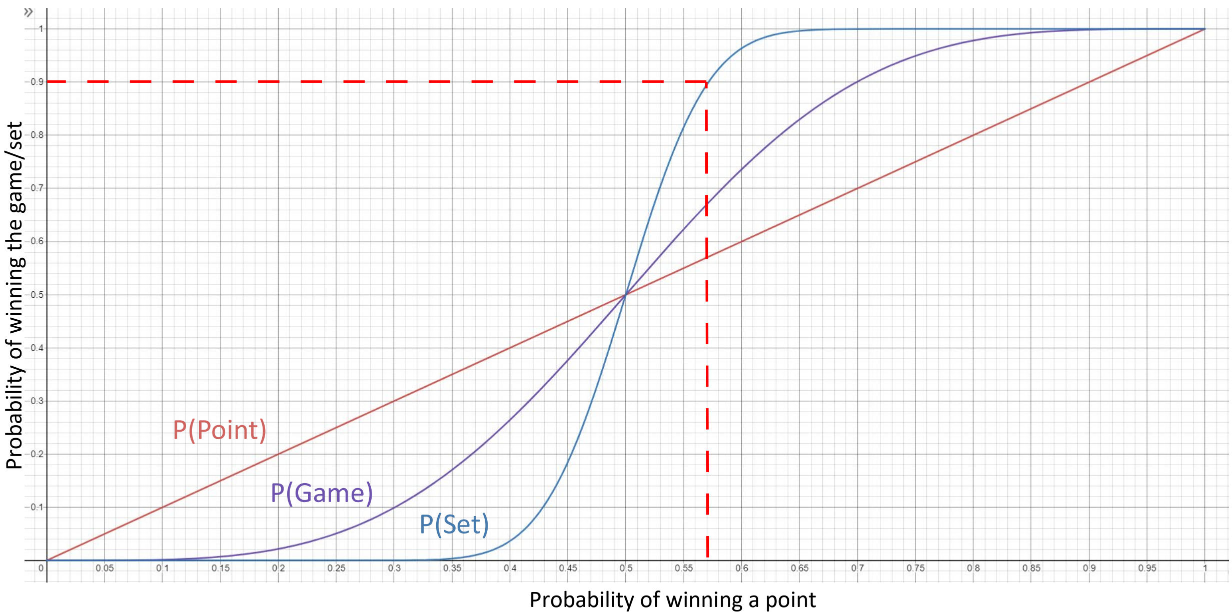


Figure 5 - Graphing the probability of winning a set

Comparing the probability of winning a game (purple) to the probability of winning a set (blue), we can see that the blue line is a lot steeper towards the middle and has significantly more leverage the higher the probability of scoring a point. Here to get a 90% chance of winning, we only need to have the probability of p to be above 0.57.

2.4 The probability of winning a match

Now we can use our calculations to find the probability of winning a match. In some tournaments they use a best-of-three sets system, but in my investigation, I'll be looking at the best-of-five system, where the first player to win three sets wins the tennis match.

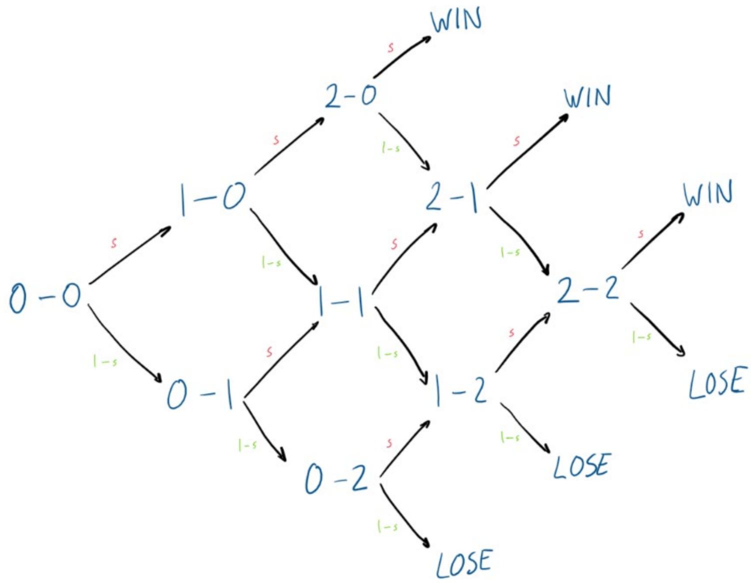


Figure 6 - Probability tree of winning a match

There are three different ways to win a match, we can either win the first three sets consecutively (3 sets total), win the last set and two of the last three (4 sets total) or the last set and two of the last four (5 sets total).

$$P(winning\ the\ match) = P(3\ sets\ \ total) + P(4\ sets\ total) + P(5\ sets\ total)$$

Using the same methods as before, we can calculate the probability of each outcome.

$$P(3\ sets\ total) = s^3$$

$$P(4\ sets\ total) = 3s^3(1-s)$$

$$P(5\ sets\ total) = 6s^3(1-s)^2$$

Hence, we can deduce our final solution.

$$P(winning\ the\ match) = s^3 + 3s^3(1-s) + 6s^3(1-s)^2$$

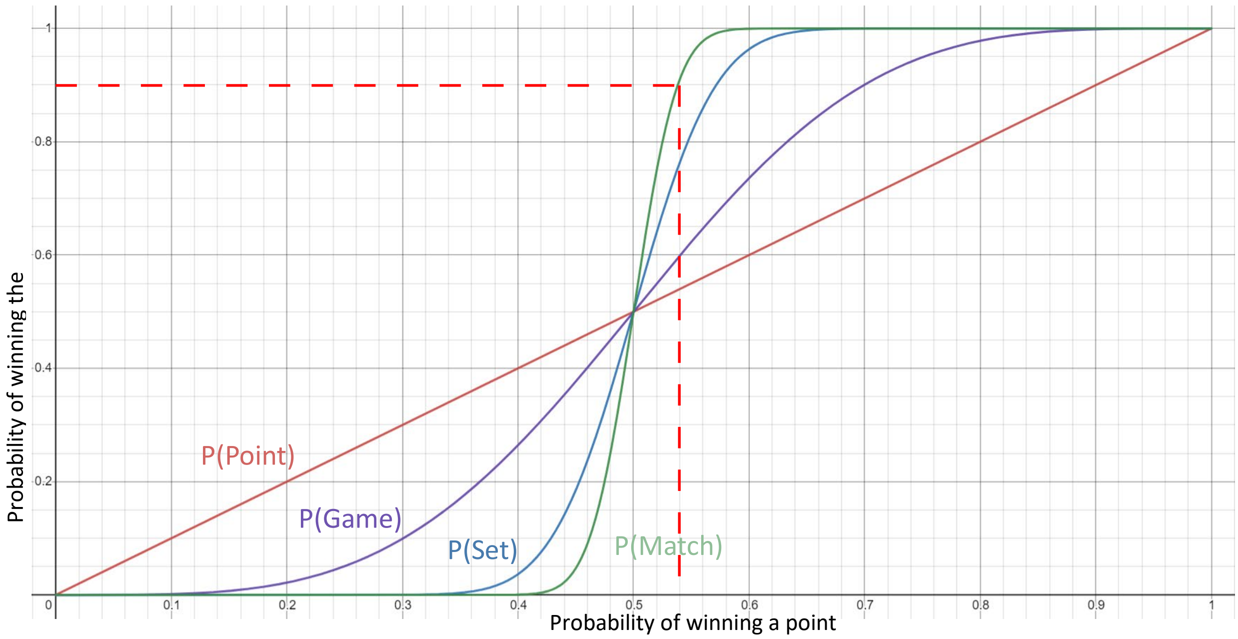
Where:

$$s = P(Set) = g^6 + 6g^6(1-g) + 21g^6(1-g)^2 + 56g^6(1-g)^3 + 126g^6(1-g)^4 + 252g^7(1-g)^5 + 504g^6(1-g)^6 t$$

$$g = P(Game) = p^4 + 4p^4(1-p) + 10p^4(1-p)^2 + \frac{20p^5(1-p)^3}{1-2p+2p^2}$$

$$t = p^7 + 7p^7(1-p) + 28p^7(1-p)^2 + 84p^7(1-p)^3 + 210p^7(1-p)^4 + 462p^7(1-p)^5 + \frac{924p^8(1-p)^6}{1-2p+2p^2}$$

This can be plotted into a graph.



Here we can use that the graph is more easily affected, you only need a 0.54 probability of scoring a point to have a 90% chance of winning the entire match. If you have even a tiny advantage, let's say a probability of 0.51, the chance of you winning a match is ~63%.

3 Conclusion

All in all, I found this topic very interesting, it demonstrated that having even slightly more chance of scoring a point can have a massive impact on the outcome of the game. It's pretty fascinating how steep the curve is for the probability of winning a match based on the probability of winning a point. You'll struggle to win against someone who may be just a little more skilled than you.

However, it has to be understood that external factors such as fatigue, who is serving, type of terrain etc. aren't taken into consideration and could ultimately manipulate the final outcome, there is more to a tennis match than just one's probability of scoring a point. In this investigation we are also assuming a consistent p for scoring a single point, which could change as the match plays on.