Dissertation to obtain the Bachelor of Mathematics (BMATH)

The evolution of tennis modelling and simulations

GINA HENDRICKX
Student number: 0201681649

Guided by: Prof. dr. Christophe Ley

2022-2023

Contents

Abstract

The objective of this bachelor thesis is to give an overview of the evolution of tennis modelling and simulations from 1971 to 2023.
Tennis dates back to the 12th century and with time became more and more popular to being one of the most played two-players sport in the world. Even though tennis knew a huge growth in the 1960s, research for the sport started a little later in comparison to others.
Questions were answered and theories were developed such as:
- Simple Probabilistic Models
- Are points in Tennis independent and identically distributed?
- Markov Chain Models
- Predictions of the outcome of a match
- The best strategies to win a match

Court surfaces were taken into consideration and it was studied how the "server-effect" influences the game. As years have gone by, technology got better and more in-depth data was available making for example sound and audio analysis possible.
All the results can be used from players and coaches to optimize their way of playing tennis and the predictions can be used to make a bet. New theories could be established using the older ones and developing those even further.

Tennis Vocabulary

In order to understand the language used in the bachelor thesis, I start with the tennis vocabulary I am going to be using.
We will concentrate on singles match, which is played between two players. However, a tennis match can also be played as doubles match, where there are two players on each side.
This sport is played on a court with a specific surface: carpet, grass, clay or hard surface, which has two identical sides with a net in-between.
A tennis match consists of 2-5 sets which are determined by 6-7 games won by at least 4 points each. The points are known as 15 (first one), 30 (second one), 40 (third one), the fourth is the winning one. The players win a game if they lead by two points. 40-40 is known as deuce.
The players try to serve in the service area, which is a small rectangle behind the net in-between the centre service line, the service line and the singles sideline, they have two serves to do so. A risky serve is faster or closer to the lines, further from the middle or with more spin than a safer one.
Then they try to force their opponent to play outside the singles sideline and the baseline or into the net.
The server is changed every game and the sides are switched every two games, except for the first game of the set.

1 Introduction

What is going to follow is a closer look into scientific articles about tennis. These are often published in academical journals, accessible to the public. These papers use as much data as possible, in order to provide a correct statistical analysis and the best interpretation of the result.
These scientific findings are important to gain new knowledge and to support people in their decision-making, for example how tennis players should serve or to distinguish the most important point in a game or to adapt the rules of the game.
There has been a great evolution of such research in tennis, where the newer discoveries are based on the older ones and could be exploited even more with the help of new technologies and more available data.
We are going to start by having a look at four of the earliest papers about tennis in great detail.
The first article is the work of D. Gale (1971) [1]. He developed a simple model about the optimal strategy for serving in tennis.
B.P. Hsi and D.M. Burych (1971) [2] presented several comments and suggestions for the rules of Two Players Sports using relative probability of a server winning a game.
S.L. George (1973) [3] developed a similar model than D. Gale. A simple probabilistic model is established to find the perfect serving strategy.
J.S. Croucher (1986) [4] used the work of B.P. Hsi and D.M. Burych (1971) [2] to establish conditional probabilities of a player winning a game from any scoreline.
As mentioned above, there is much more research that has been undertaken. The detail will not be shown in this bachelor thesis, but a short introduction to 22 reports dating from $2000-2022$ will be given:
S.R. Clarke and D. Dyte (2000) [5] simulated major tennis tournaments and used the official rankings to predict the outcome of each match, their predictions were very close to the actual real results of the encounters.
A stochastic model is proposed by Y. Liu (2001) [6] to illustrate the dependence between the point winning probabilities and the outcome of a match.
F. Klaassen and J. Magnus (2003) [7] used a computer program and statistical analysis to be able to forecast the winner during any moment of the tennis match.
P.K. Newton and G.H. Pollard (2004) [8] studied the probability of winning a set in tennis and proved that it is independent of which player serves first, as long as the points are treated as independent and identically distributed random variables.
T. Barnett and S.R. Clarke (2005) [9] utilized the standard statistics published by the ATP to predict the serving statistics and further match outcomes, such as the length of the match and chance of either player winning, these calculations can be updated as the match progresses.
An explicit proof that the probability of winning a set, and hence a match, is independent of which player serves first is given by P.K. Newton and J.B. Keller (2005) [10] as the probability that a player wins a tournament with 128 opponents.
T. Barnett, A. Brown and S. Clarke (2006) [11] developed a more accurate model to predict the outcome of a tennis match by not assuming that the probability of winning a point on service is constant.
G. Hunter, A. Shihab and K. Zienowicz (2008) [12] analysed tennis rallies using information from both video and audio signals.
Through a stochastic Markov chain model P.K. Newton and K. Aslam (2009) [13] were able to obtain the probability density function for a player to win a match.
A. Bedford, T. Barnett, Gr. Pollard and Ge. Pollard (2010) [14] highlighted that a player should provide a balanced approach when it comes to analysing match statistics. He/She

should focus on all areas of the game and not focus on certain data, in order to not neglect some parts of their game.
D. Paindaveine and Y. Swan (2011) [15] conducted a stochastic analysis of some two-person sports, including tennis, where they discovered that the number of games won by a player follows a Poisson distribution and the number of points won by a player in a set follows a normal distribution.
A complex network analysis of the history of professional tennis between 1968 and 2010 was performed by F. Radicchi (2011) [16] in order to identify the best player ever.
T. Barnett, D. O'Shaughnessy, A. Bedford's (2011) [17] article demonstrates how spreadsheets can generate the probability of winning a tennis match conditional on the state of the match.
A hierarchical Markov model is established by W.J. Knottenbelt, D. Spanias and A.M. Madurska (2012) [18] to analyse match statistics of previous encounters and furthermore give a pre-play estimate of the probability of each player to win a match.
C. Roure (2014) [19] performed a point-by-point analysis of sequences of play in order to determine the points that have the most influence on the chances of winning a match.
M. Bevc (2015) [20] was able to make match outcome predictions from a combination of in play and historical data, which turned out to be more effective than some previous findings.
C. Gray's (2015) [21] article provides a valuable overview of the use of statistics in tennis and highlights the importance of analyzing different types of data.
C. Cooper and R.E. Kennedy (2021) [22] determined the number of points two players can expect to play in a game, they were able to conduct a mathematical analysis to estimate the expected length and probability of winning a tennis game.
A more resource-effective alternative in the form of a combinatorial approach based on a binomial distribution is proposed by A. Sarcevic, M. Vranic and D. Pintar (2021) [23] to predict the outcome of a match.
J.C. Yue, E.P. Chou, M.-H. Hsieh, L.-C. Hsiao (2022) [24] presented a study using a statistical approach to forecast tennis matches via the Glicko model in addition to the exploratory data analysis.
S.A. Kovalchik, J. Albert (2022) [25] developed a statistical model to analyse the impact patterns of serve returns in professional tennis using a Gaussian mixture model.
Finite Markov chains are especially suited in the modelling of net games like tennis and allowed F. Rothe, M. Lames (2022) [26] to simulate tennis behaviour.

2 D. Gale (1971) - Optimal Strategy for Serving in Tennis

2.1 The main goal

The author wanted to know if the players should continue doing a risky first serve and a safer second one or if another combination: risky+risky or safe+risky or safe+safe would make more sense.

2.2 Calculations

Let's consider the set S with possible serves with $s \in S$ one of them.
We associate s with two probabilities $p(s)$ and $q(s)$:

- $p(s)$ = the probability that the serve is inside of the serving area, it is considered as "good"
- $q(s)$ = the probability that the server wins the point after a good serve

It is clear that the probability of winning the point as the server is: $p(s) \cdot q(s)$, denoted as $w(s)$. The two probabilities $p(s)$ and $q(s)$ go hand in hand.

In other words, if the serve is safer $p(s)$ is higher but $q(s)$ is lower as the serve is easier to return. But if the serve is trickier it is less likely to be inside of the serving area and more difficult to return. It goes without saying that a server should try to avoid serves which are risky to get inside the court and easy to return.

By knowing this, what is the optimal strategy for serving in tennis?
The server wants to win the point by using his first or second serve, the pair chosen from the set $S \times S = \{s_1 \in S \text{ and } s_2 \in S | s_1 \text{ the first serve and } s_2 \text{ the second serve}\}$.
It will either be won on first serve $w(s_1)$ or on second serve $(1 - p(s_1)) \cdot w(s_2)$, with $(1 - p(s_1))$ the probability that the first serve is not inside the serving area.
We have the probability for winning a point on first or second serve:

$$\begin{aligned} P(s_1, s_2) &= w(s_1) + (1 - p(s_1)) \cdot w(s_2) \\ &= w(s_1) + w(s_2) - p(s_1) \cdot w(s_2) \\ &= p(s_1) \cdot q(s_1) + (1 - p(s_1)) \cdot p(s_2) \cdot q(s_2) \end{aligned}$$

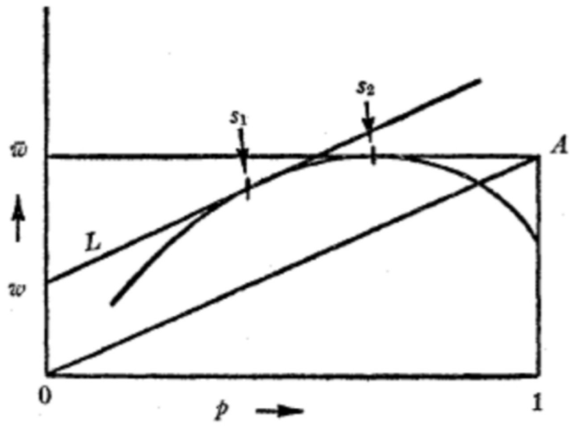
Optimal strategy for serving in tennis:

One optimal strategy for serving would be to find the highest possible value for the equalities above.
To do so, let us choose an s_2 that maximizes the function $w(s)$. Let us refer to this s_2 as b. Then we have that

$$b \in S, \text{ such that } w(b) = \max_{s \in S}\{w(s)\} := \overline{w}$$

where $\overline{w}$ is the maximum of the function $w(s)$.
Then we choose s_1 to maximize $P(s_1, b) = w(s_1) + \overline{w} - p(s_1) \cdot \overline{w}$.
Let's look at the illustration of this strategy.

Graphical illustration about how s_1 and s_2 are chosen in order to maximize the equalities above



D. Gale (1971) Optimal Strategy for Serving in Tennis. Mathematics Magazine Volume 44 - Issue 4, pp. 197-199 - figure on p.198

Explanation of the figure:
The horizontal axis represents $p(s)$, the probability that the serve is inside of the serving area, it is considered as "good", with values lying between $[0;1]$.
The vertical axis represents $w(s)$, the probability that the server wins the point after the first or second serve, where $\overline{w}$ is the maximum of the function w or in other words, $\overline{w} = \max_{s\in S}\{w(s)\}$.
We draw a parabola that represents the set of points $P = \{s \in S | (p(s), w(s))\}$, with S the set of possible serves. The figure suggests that the player has a variety of possible serves.
s_2 is clearly the highest point on the graph of possible serves as it maximizes $w(s_2)$, so $\overline{w}$.

Different possibilities for the optimal strategy:

- The player uses a first riskier serve and a second less risky serve which satisfy the equalities above, shown by the figure above.

- He/She can also serve such that $s_1 = s_2$. The point of intersection of L with the graph is then s_2.

- But it is never a good idea to make a riskier second serve than the first one, because if the first serve fails one still has the opportunity to do a second one and win the point. But if the second one fails, the point is lost. So it is better to do the risky serve first and then the safer one in order to have the highest probability of winning the point.

- Let's see what the player should do with his/her two available serves s_1 and s_2 with coordinates $(p(s_1), w(s_1))$ and $(p(s_2), w(s_2))$ while $p(s_1) < p(s_2)$.
In order to get an optimal strategy, we deduce from the equalities above:
$P(s_1, s_2) = w(s_1) + (1 - p(s_1)) \cdot w(s_2)$,
that the following inequalities need to hold:

$$w(s_2) \geq w(s_1) \geq w(s_2)(1 - (p(s_2) - p(s_1)))$$

$$\Rightarrow w(s_2) \geq w(s_1) \geq w(s_2) - w(s_2) \cdot p(s_2) + w(s_2) \cdot p(s_1)$$

Explanation of the inequalities:
We want to maximize $P(s_1, s_2)$.
We have $p(s_1) < p(s_2)$ and in order to maximize $(1 - p(s_1)) \cdot w(s_2)$:
$p(s_1)$ has to be as small as possible and $w(s_2)$ as big as possible.
Therefore, $w(s_2) \geq w(s_1)$.
While observing the second inequality, we get that $(p(s_2) - p(s_1))$ lies between $[0; 1]$ because $p(s_1) < p(s_2)$ and both probabilities lie between $[0; 1]$.
The subtraction $(1 - (p(s_2) - p(s_1))$ is also located in that domain, as the multiplication $w(s_2)(1 - (p(s_2) - p(s_1)))$.
To show $w(s_1) \geq w(s_2)(1 - (p(s_2) - p(s_1))$ it is easier to assume the case where $w(s_1) < w(s_2)(1 - (p(s_2) - p(s_1)))$, then

$$\begin{aligned}
P(s_1, s_2) &= w(s_1) + (1 - p(s_1))w(s_2) \\
&< w(s_2)(1 - (p(s_2) - p(s_1)) + (1 - p(s_1))w(s_2) \\
&= w(s_2)(2 - p(s_2)) \\
&= w(s_2) + w(s_2)(1 - p(s_2)) = P(s_2, s_2).
\end{aligned}$$

Therefore $w(s_1) \geq w(s_2)(1 - (p(s_2) - p(s_1)))$, because, if we decide to serve differently on the first serve than on the second serve (with s_1 and s_2), we want it to be more beneficial than serving identically on both serves (s_2 and s_2).

In a real tennis match the optimal strategy might not always be possible and the first or the second inequality might fail. Let's look at the two different cases.
The first inequality fails:
If $w(s_1) \geq w(s_2)$ and $p(s_1) < p(s_2)$ then we have $q(s_1) \geq q(s_2)$ as $w(s) = p(s) \cdot q(s)$.
We defined $w(s)$ as the probability of winning the point as the server. Now in this case the probability of winning the point is higher with the risky first serve as seen above: $w(s_1) \geq w(s_2)$ and $p(s_1) < p(s_2)$. The player should therefore use his riskier serve twice to get the best chances of winning the point.
The second inequality fails:
If $w(s_1) \leq (w(s_2) - w(s_2)p(s_2) + w(s_2)p(s_1))$ and $p(s_1) < p(s_2)$ then we have $q(s_1) \leq q(s_2)$.
We have $w(s_1) \leq w(s_2)$, so the probability of winning the point is higher with the safe second serve. Therefore the server should use his safer serve twice to get the best chances of winning the point.

2.3 Results and Conclusions

By looking at the calculations above one can see that the strategy risky+safe is not the only option that a player can go for, but also the combination risky+risky and safe+safe are good solutions for the appropriate situation.
For every player, the best strategy is different as it depends on the probabilities $w(s_1) = p(s_1)q(s_1)$ and $w(s_2) = p(s_2)q(s_2)$. These differ from athlete to athlete and their skills for serving. Customizing the best strategy for oneself is the optimal approach to winning the own serve.
One can understand though why players would go for the traditional way of serving: risky+safe. When doing a risky serve the probability that this one hits the ground in the right area: $p(s)$, is lower than for a safer serve. But the probability that the server wins: $q(s)$, is higher for a riskier serve. As a result, a lot of athletes do a risky first serve because if it doesn't land in the service area, they can just do a second safe serve, which allows them to still win the point. If they use a second risky serve, the probability is high that it doesn't land in the service zone and they lose the point immediately.

3 B.P.Hsi and D.M.Burych (1971) - Games of two Players

3.1 The main goal

Creating a probability model that includes the player's skill, his/her mental strength, his/her physical abilities, the surface played on and if one is serving or not was the goal of the mathematicians.
The "server-effect" can be explained as follows: the server is in a favourable position as he/she can direct the ball wherever he/she wants and places the returner in a difficult situation to return the ball which is again convenient for him/her. This effect fades as the rally goes on.
As serving first is often an advantage, the authors developed the relative probability of a server winning a game in a match. They were then able to establish the probability that a server wins a set and deduced the probability to win a match.

3.2 Calculations

The game of tennis is played point by point, one either wins an exchange with probability p or loses with probability $1-p$.
Let's look at a specific case: The player A serves during a game with p the chance of winning a point. In order for him/her to win a game, he/she needs a score of $4-k$ and $k \in \{0;1;2\}$, as there needs to be a difference of two points in order for A to win a game (the score 3-3 will be discussed further below).
The probability that player A wins a game with a score of $4-k$ and $k \in \{0;1;2\}$ is:

$$\boxed{\binom{3+k}{k} \cdot p^4 \cdot (1-p)^k}$$

Justification:
We will proceed with a proof by exhaustion.
The probability that player A wins all 4 points with $\underline{k=0}$ losses is

$$p^4 = 1 \cdot p^4(1-p)^0 = \binom{3+k}{k} p^4(1-p)^k \text{ for } k=0$$

The probability that player A wins 4 points with $\underline{k=1}$ loss is

$$4 \cdot p^4(1-p) = \binom{3+k}{k} p^4(1-p)^k \text{ for } k=1$$

where we multiply by 4 due to the 4 ways in which he/she can lose a point. He/She can lose the first, second to fourth exchange, but not the fifth because player A wins after 5 exchanges and therefore must win the fifth exchange.

The probability that player A wins 4 points with $\underline{k=2}$ losses is

$$\binom{5}{2} p^4(1-p)^2 = \binom{3+k}{k} p^4(1-p)^k \text{ for } k=2$$

where we multiply by $\binom{5}{2}$ to account for the ways in which player A can lose 2 of the 5 first exchanges. He/She wins the sixth exchange.

Now if the score is <u>3-3</u> it is called games going into deuces. In order for A to win the game and avoid further deuces he/she has to win two consecutive points.
The probability that A wins the game after a deuce:

$$\boxed{\binom{6}{3} \cdot p^5 \cdot (1-p)^3 \sum_{j=0}^{\infty} \{2p(1-p)\}^j = \left[\binom{6}{3} \cdot p^5 \cdot (1-p)^3\right] / (1 - 2p + 2p^2)}$$

<u>Justification</u>:
Let's start by splitting the formula to be able to clearly see what's going on:

$$\begin{aligned} &\binom{6}{3} \cdot p^5 \cdot (1-p)^3 \textstyle\sum_{j=0}^{\infty} \{2p(1-p)\}^j \\ = &\binom{6}{3} \cdot p^3 \cdot (1-p)^3 \cdot p^2 \textstyle\sum_{j=0}^{\infty} \{2p(1-p)\}^j \end{aligned}$$

The first part of the formula is the probability that the game goes into deuces:

$$\binom{6}{3} \cdot p^3 \cdot (1-p)^3$$

Player A and player B win each 3 points. We multiply by $\binom{6}{3}$ to account for the ways in which player A can lose 3 of the 6 exchanges.

Now let's turn to the second part which is the probability that player A wins from deuce:

$$p^2 \textstyle\sum_{j=0}^{\infty} \{2p(1-p)\}^j$$

p^2 is there as player A needs to win two consecutive points after going into deuces, but it is possible that player A doesn't immediately succeed and players A and B alternately get a point, which means that they are going into a new deuce, that is given by:

$$\textstyle\sum_{j=0}^{\infty} \{2p(1-p)\}^j$$

- <u>$j = 0$</u> Player A immediately wins two consecutive points after going into deuces:

  $$p^2 \cdot \{2p(1-p)\}^0 = p^2$$

- <u>$j = 1$</u> Player A wins two consecutive points after losing 1 point and winning 1 point (new deuce) after going into deuces:
  $p^2 \cdot \{2p(1-p)\}^1 = p^2 \cdot 2p(1-p)$ with 2 being the possibilities of either losing or winning the first point after the original deuce: $lw = d$ or $wl = d$
  with $d =$ deuce and $l =$ losing and $w =$ winning

- <u>$j = 2$</u> Player A wins two consecutive points after losing 2 points and winning 2 points alternately (2 new deuces) after going into deuces:
  $p^2 \cdot \{2p(1-p)\}^2 = p^2 \cdot 4p^2(1-p)^2$ with 4 the possibilities of the outcome of the four points after the deuce:
  $lw = d$ and $lw = d$ or $lw = d$ and $wl = d$
  or $wl = d$ and $lw = d$ or $wl = d$ and $wl = d$
  with $d =$ deuce and $l =$ losing and $w =$ winning

And so on...

- $\underline{j = n,\ n \in \mathbb{N}}$ Player A wins two consecutive points after losing n points and winning n points alternately (n new deuces) after the original deuce:
  $p^2 \cdot \{2p(1-p)\}^n = p^2 \cdot 2^n p^n (1-p)^n$ with 2^n the possibilities of the outcome of the $2n$ points after the deuce.
  There are 2 ways for player A to return to a deuce after a deuce. Winning first, then losing, or losing first and then winning. By writing these two possibilities in a set, we get $\{wl, lw\}$. The ways in which player A can therefore return to a deuce after a deuce n times is given by the set $\{wl, lw\}^n$, where its cardinality is 2^n.

Thus, the probability that player A wins after less than or equal to $n+1$ deuces is given by

$$\binom{6}{3} \cdot p^5 \cdot (1-p)^3 \textstyle\sum_{j=0}^{n} \{2p(1-p)\}^j$$

Because there is in theory no limit to how many deuces a game can have, the probability that player A wins the game is given by the limit of the above product as n tends to infinity:

$$\binom{6}{3} \cdot p^5 \cdot (1-p)^3 \textstyle\sum_{j=0}^{\infty} \{2p(1-p)\}^j$$

That's how we get the left hand side of the above equation.

Let's see how we go from $\sum_{j=0}^{\infty} \{2p(1-p)\}^j$ to $1/(1-2p+2p^2)$

$$\begin{aligned} & 2p(1-p) = 2p - 2p^2 = 0 \\ \Leftrightarrow & p = 0 \text{ or } p = 1, \text{ and } p^2 \text{ has a negative coefficient} \\ \Rightarrow & 2p(1-p) > 0 \text{ for } p \in]0,1[\end{aligned}$$

and

$$\begin{aligned} & \frac{d}{dp} 2p - 2p^2 = 2 - 4p = 2(1-2p) = 0 \\ \Leftrightarrow & p = 0.5 \text{ where } 2(0.5) - 2(0.5)^2 = 0.5 < 1 \end{aligned}$$

$$\begin{aligned} \Rightarrow & 0 < 2p(1-p) < 1 \text{ for } p \in]0,1[\\ \Rightarrow & \sum_{j=0}^{\infty} \{2p(1-p)\}^j = \frac{1}{1-(2p(1-p))} = \frac{1}{1-2p+2p^2} \text{ for } p \in]0,1[\end{aligned}$$

which is how we arrive at the right side of the above equation.

Above are described the probabilities if player A is serving and gets to win a game with score $4-k$, but also if the games are going into deuces.
<u>The probability of winning a game for player A is the sum of both probabilities above:</u>

$$W_A = \textstyle\sum_{k=0}^{2} \binom{3+k}{k} \cdot p^4 \cdot (1-p)^k + \left[\binom{6}{3} \cdot p^5 \cdot (1-p)^3\right] / (1-2p+2p^2)$$

The same formula applies for the other player B, when this one serves: W_B.

Let's consider the case if player A wins a set with a score of $6-k$, $k \in \{0,1,2,3,4\}$ the games won by player B (5-5 is treated further below). Player A starts serving, if k is even player B serves during the last game and if k is odd player A serves during the last game.

<u>If k is even, then $k \in \{0,2,4\}$:</u>

$$\textstyle\sum_{i=(6-k)/2}^{(6+k)/2} \binom{[(6+k)/2]-1}{6-i-1} \binom{(6+k)/2}{i} \cdot W_A^i \cdot (1-W_B)^{6-i} \cdot W_B^{i-[(6-k)/2]} \cdot (1-W_A)^{[(6+k)/2]-i}$$

<u>Justification:</u>
We will proceed with a proof by exhaustion.

- <u>k=0</u> so Player B doesn't win a game

  $$\begin{aligned}
  &\binom{[6/2]-1}{6-3-1} \binom{6/2}{3} \cdot W_A^3 \cdot (1-W_B)^{6-3} \cdot W_B^{3-[6/2]} \cdot (1-W_A)^{[6/2]-3} \\
  &= \binom{2}{2}\binom{3}{3} \cdot W_A^3 \cdot (1-W_B)^3 \cdot W_B^0 \cdot (1-W_A)^0 \\
  &= W_A^3 \cdot (1-W_B)^3 \\
  &= \textstyle\sum_{i=(6-k)/2}^{(6+k)/2} \binom{[(6+k)/2]-1}{6-i-1} \binom{(6+k)/2}{i} \cdot W_A^i \cdot (1-W_B)^{6-i} \cdot W_B^{i-[(6-k)/2]} \cdot (1-W_A)^{[(6+k)/2]-i}
  \end{aligned}$$
  with $k=0$

  We get the expected result, Player A wins three games while serving and three games while returning.

- <u>k=2</u> so Player B wins two games

  $$\begin{aligned}
  &\binom{[(6+2)/2]-1}{6-2-1} \binom{(6+2)/2}{2} \cdot W_A^2 \cdot (1-W_B)^{6-2} \cdot W_B^{2-[(6-2)/2]} \cdot (1-W_A)^{[(6+2)/2]-2} \\
  &+ \binom{[(6+2)/2]-1}{6-3-1} \binom{(6+2)/2}{3} \cdot W_A^3 \cdot (1-W_B)^{6-3} \cdot W_B^{3-[(6-2)/2]} \cdot (1-W_A)^{[(6+2)/2]-3} \\
  &+ \binom{[(6+2)/2]-1}{6-4-1} \binom{(6+2)/2}{4} \cdot W_A^4 \cdot (1-W_B)^{6-4} \cdot W_B^{4-[(6-2)/2]} \cdot (1-W_A)^{[(6+2)/2]-4} \\
  &= \binom{3}{3}\binom{4}{2} \cdot W_A^2 \cdot (1-W_B)^4 \cdot W_B^0 \cdot (1-W_A)^2 \\
  &+ \binom{3}{2}\binom{4}{3} \cdot W_A^3 \cdot (1-W_B)^3 \cdot W_B^1 \cdot (1-W_A)^1 \\
  &+ \binom{3}{1}\binom{4}{4} \cdot W_A^4 \cdot (1-W_B)^2 \cdot W_B^2 \cdot (1-W_A)^0 \\
  &= \textstyle\sum_{i=(6-k)/2}^{(6+k)/2} \binom{[(6+k)/2]-1}{6-i-1} \binom{(6+k)/2}{i} \cdot W_A^i \cdot (1-W_B)^{6-i} \cdot W_B^{i-[(6-k)/2]} \cdot (1-W_A)^{[(6+k)/2]-i}
  \end{aligned}$$
  with $k=2$

  We have different possibilities how player A can lose 2 games out of 8 games total. As player A serves first, the last game is served by player B and necessarily has to be won by player A.

He/She can win all the games where player B serves and two where he/she is the server. Player A loses two games where he is the server. We multiply by $\binom{4}{2}$ to represent the different ways in which player A can win two of the four games in which he/she serves.
Or
He/She can win 3 games while serving and 3 while returning. So he/she loses 1 game while returning and serving. We multiply by $\binom{4}{3}$ to represent the different ways in which player A can win three of the four games in which he/she serves. We also multiply by $\binom{3}{2}$ to account the ways in which player A can win two games of three while returning. We don't take into consideration the fourth game as he/she has to win the last game of the set, which is served by Player B.
Or
He/She can win all the games while serving and two where he/she is returning. Player A loses 2 games where he/she is not the server. We also multiply by $\binom{3}{1}$ to account the ways in which player A can win one game of three while returning. We don't take into consideration the fourth game as he/she has to win the last game of the set, which is served by Player B.

- k=4 so Player B wins four games
  We proceed as before with $i \in \{1, 2, 3, 4, 5\}$ and we get:

$$\binom{4}{4}\binom{5}{1} \cdot W_A^1 \cdot (1-W_B)^5 \cdot W_B^0 \cdot (1-W_A)^4$$
$$+\binom{4}{3}\binom{5}{2} \cdot W_A^2 \cdot (1-W_B)^4 \cdot W_B^1 \cdot (1-W_A)^3$$
$$+\binom{4}{2}\binom{5}{3} \cdot W_A^3 \cdot (1-W_B)^3 \cdot W_B^2 \cdot (1-W_A)^2$$
$$+\binom{4}{1}\binom{5}{4} \cdot W_A^4 \cdot (1-W_B)^2 \cdot W_B^3 \cdot (1-W_A)^1$$
$$+\binom{4}{0}\binom{5}{5} \cdot W_A^5 \cdot (1-W_B)^1 \cdot W_B^4 \cdot (1-W_A)^0$$
$$= \sum_{i=(6-k)/2}^{(6+k)/2} \binom{[(6+k)/2]-1}{6-i-1}\binom{(6+k)/2}{i} \cdot W_A^i \cdot (1-W_B)^{6-i} \cdot W_B^{i-[(6-k)/2]} \cdot (1-W_A)^{[(6+k)/2]-i}$$
with $k = 4$

If k is odd, then $k \in \{1, 3\}$:

$$\sum_{i=(6-k+1)/2}^{(6+k+1)/2} \binom{(6+k-1)/2}{6-i}\binom{(6+k-1)/2}{i-1} \cdot W_A^i \cdot (1-W_B)^{6-i} \cdot W_B^{i-[(6-k+1)/2]} \cdot (1-W_A)^{[(6+k+1)/2]-i}$$

- k=1 so Player B wins one game
  We evaluate the sum with $i \in \{3, 4\}$ and we get:

$$\binom{3}{3}\binom{3}{2} \cdot W_A^3 \cdot (1-W_B)^3 \cdot W_B^0 \cdot (1-W_A)^1$$
$$+\binom{3}{2}\binom{3}{3} \cdot W_A^4 \cdot (1-W_B)^2 \cdot W_B^1 \cdot (1-W_A)^0$$

$= \sum_{i=(6-k+1)/2}^{(6+k+1)/2} \binom{(6+k-1)/2}{6-i} \binom{(6+k-1)/2}{i-1} \cdot W_A^i \cdot (1-W_B)^{6-i} \cdot W_B^{i-[(6-k+1)/2]} \cdot (1-W_A)^{[(6+k+1)/2]-i}$ with $k=1$

We have different possibilities how player A can lose 1 game out of 7 games total. As player A serves first, the last game is served by him/her and necessarily has to be won by player A.
He/She can win 3 games while serving and 3 while returning. So he/she loses 1 game while serving. We multiply by $\binom{3}{2}$ to represent the different ways in which player A can win two of the three games in which he/she serves. We don't take into consideration the fourth game as he/she has to win the last game of the set, which is served by him/herself.
Or
He/She can win all the games where he/she serves and one where he/she is the returner. Player A loses one game while returning. We multiply by $\binom{3}{2}$ to represent the different ways in which player A can win two of the three games in which he/she returns.

- <u>k=3</u> so Player B wins three games
  We evaluate the sum with $i \in \{2,3,4,5\}$ and we get:

$$\binom{4}{4}\binom{4}{1} \cdot W_A^2 \cdot (1-W_B)^4 \cdot W_B^0 \cdot (1-W_A)^3$$
$$+ \binom{4}{3}\binom{4}{2} \cdot W_A^3 \cdot (1-W_B)^3 \cdot W_B^1 \cdot (1-W_A)^2$$
$$+ \binom{4}{2}\binom{4}{3} \cdot W_A^4 \cdot (1-W_B)^2 \cdot W_B^2 \cdot (1-W_A)^1$$
$$+ \binom{4}{1}\binom{4}{4} \cdot W_A^5 \cdot (1-W_B)^1 \cdot W_B^3 \cdot (1-W_A)^0$$
$= \sum_{i=(6-k+1)/2}^{(6+k+1)/2} \binom{(6+k-1)/2}{6-i} \binom{(6+k-1)/2}{i-1} \cdot W_A^i \cdot (1-W_B)^{6-i} \cdot W_B^{i-[(6-k+1)/2]} \cdot (1-W_A)^{[(6+k+1)/2]-i}$ with $k=3$

If we interchange W_A with $1-W_A$ and W_B with $1-W_B$ then we get the probability for the score $k-6$.

Let's look at the match at <u>5-5 games</u> and the probability that Player A wins after this deuce.

$$\text{Prob}\{\text{score of 5-5}\} \cdot W_A \cdot (1-W_B) \cdot \sum_{j=0}^{n=\infty} [W_A \cdot W_B + (1-W_A) \cdot (1-W_B)]^j$$

<u>Justification:</u>

Let's look at the formula in several steps.

First of all:
Prob{score of 5-5} is the probability to get to 5-5 games each.

Then:
$W_A \cdot (1-W_B)$ is the probability that player A wins two consecutive games, one while serving

and one while returning, in order to avoid $6-6, 7-7, ...$ and win the set.

Last but not least:
$\sum_{j=0}^{n=\infty}[W_A \cdot W_B + (1-W_A)\cdot(1-W_B)]^j$: player A and B alternately win one game.

Justification:

- n=0: Player A immediately wins after 5-5 and there is no further deuce.

- n=1: There is one more deuce before player A wins the set.

  $[W_A \cdot W_B + (1-W_A)\cdot(1-W_B)]^1$

  There are two possibilities to win the games, either both players win their serve or both lose it.

- n=2: There are two more deuces before player A wins the set.

  $$[W_A \cdot W_B + (1-W_A)\cdot(1-W_B)]^2$$
  $$= W_A^2 \cdot W_B^2 + 2\cdot W_A \cdot W_B \cdot (1-W_A)\cdot(1-W_B) + (1-W_A)^2\cdot(1-W_B)^2$$
  $$= W_A \cdot W_B \cdot W_A \cdot W_B + (1-W_A)\cdot(1-W_B)\cdot(1-W_A)\cdot(1-W_B)$$
  $$+ W_A \cdot W_B \cdot (1-W_A)\cdot(1-W_B) + (1-W_A)\cdot(1-W_B)\cdot W_A \cdot W_B$$

  We have three different possibilities that lead to two new deuces:
  We know that player A starts serving. Either he loses or wins which leads to the following:
  If he wins both serves: $W_A \cdot W_B \cdot W_A \cdot W_B$
  If he loses both serves: $(1-W_A)\cdot(1-W_B)\cdot(1-W_A)\cdot(1-W_B)$
  If he wins his first serve and loses his second: $W_A \cdot W_B \cdot (1-W_A)\cdot(1-W_B)$
  If he wins his second serve and loses his first: $(1-W_A)\cdot(1-W_B)\cdot W_A \cdot W_B$

  In theory, this procedure can go on to infinity until one player finally wins two consecutive games. That's how we get the formula above.

3.3 Results and Conclusions

By looking at these formulas one can deduce that if two opponents are strong in their services the game is likely to get very long. That could have negative impacts on the players' performance and the spectators' attention span.
The authors therefore suggest a more frequent change of the server and to implement a "sudden death" model in case of a deuce in a set. It would work as follows: The player who wins more points in a period of time or wins a specific number of points wins the set.

These suggestions actually happened and today are known as the tie-break. The first player at seven/ ten points with at least a lead of two points wins or the first to lead by two points after 6-6/9-9. The server is changed every two points, except for the first one, then we have an immediate switch.
The seven points tie-break is either used at 6 games all and the ten points tie-break is used to replace the third set for women's encounters or the fifth set for men in certain tournaments.

4 S.L. George (1973) - Optimal Strategy in Tennis: A Simple Probabilistic Model

4.1 The main goal

Tennis players meet each other in tournaments each year. In order to increase their probability of winning, it would be great if one could find a strategy to maximize the chance of winning a point after serving.
A lot of people go with the following mentality: make the first serve a "strong" one and the second a "weak" one.
We will see if that is indeed the best available strategy.
This paper is similar to D. Gale's [1] and the results are indeed the same, but it follows a different approach to finding the best way to serve in tennis.

4.2 Calculations

Let's consider two opponents playing a match. The server has the possibility to use a strong or a weak serve, with corresponding probability of winning the point.
Let's use the following abbreviations:

- S - a good strong serve (good = in the serving area) with probability $P(S)$
- W - a good weak serve (good = in the serving area) with probabilty $P(W)$
- Q - the server wins the point after using a strong or weak serve with probabilites $P(Q|S)$ and $P(Q|W)$

Later on we utilize the probabilities with notation $P(QS) = P(Q \cap S)$ and $P(QW) = P(Q \cap W)$:

- $P(QS)= P(Q|S) \cdot P(S)$, the probability of serving a strong good serve and winning the point
- $P(QW)= P(Q|W) \cdot P(W)$, the probability of serving a weak good serve and winning the point
are going to be used.

We can establish the four possible strategies for serving in tennis and then calculate $P(Q)$, the probability that the server wins the point. To see which strategy is the most efficient one has to obtain the largest $P(Q)$.

Strong + Weak:
For this strategy one has to sum:
-the probability that the first serve is a good strong serve and the server wins the point: $P(QS)$
-the probability that the second serve is a good weak serve and the server wins the point $P(QW)$ times the probability that the first serve is not inside the serving area $(1 - P(S))$

$$P(Q) = P(QS) + P(QW)(1 - P(S))$$

<u>Weak + Strong:</u>
The formula is the same as above but one has to exchange W and S:
-the probability that the first serve is a good weak serve and the server wins the point: $P(QW)$
-the probability that the second serve is a good strong serve and the server wins the point $P(QS)$ times the probability that the first serve is not inside the serving area $(1 - P(W))$

$$P(Q) = P(QW) + P(QS)(1 - P(W))$$

<u>Strong + Strong:</u>
The formula is composed by:
-the probability that the first serve is a good strong serve and the server wins the point: $P(QS)$
-the probability that the second serve is a good strong serve and the server wins the point $P(QS)$ which means that the first serve is not a good winning serve $(-P(QS)P(S))$

$$P(Q) = P(QS) + P(QS) - P(QS)P(S) = P(QS) \cdot \{2 - P(S)\}$$

<u>Weak + Weak:</u>
The formula is the same as above but one got to exchange W and S:
-the probability that the first serve is a good weak serve and the server wins the point: $P(QW)$
-the probability that the second serve is a good weak serve and the server wins the point $P(QW)$ which means that the first serve is not a good winning serve $(-P(QW)P(W))$

$$P(Q) = P(QW) + P(QW) - P(QW)P(W) = P(QW) \cdot \{2 - P(W)\}$$

If we want to compare which strategies are better we can look at the difference between their probabilities.
For example: Take the difference between Weak+Strong and Strong+Weak. Assuming Strong+Weak is preferable to Weak+Strong, their difference would be negative.

$$\begin{aligned}
&P(QW) + P(QS)(1 - P(W)) - P(QS) + P(QW)(1 - P(S)) < 0 \\
\Leftrightarrow\ & P(QW) + P(QS) - P(QS)P(W) - P(QS) + P(QW) - P(QW)P(S) < 0 \\
\Leftrightarrow\ & 2 \cdot P(QW) < P(QS)P(W) + P(QW)P(S) \\
\Leftrightarrow\ & P(QW) < \tfrac{P(QS)P(W)}{2-P(S)}
\end{aligned}$$

In a real tennis match, we find this inequality near universally satisfied, therefore making Strong+Weak preferable over Weak+Strong. The Weak+Strong strategy is never the best one and as a result will be excluded from our search.

Let's define $R = \frac{1}{1+P(S)-P(W)}$ and $Z = \frac{P(QW)}{P(QS)}$, the relation between the probability of winning a point using a weak serve and a strong serve.

The three remaining strategies are optimal if:

- $1 \leq Z \leq R$ for Strong + Weak
  This means $1 \leq Z \Rightarrow P(QS) \leq P(QW)$ and
  $Z \leq R \Rightarrow 1 + P(S) - P(W) \leq \frac{P(QS)}{P(QW)} \Rightarrow P(QW)(1 + P(S) - P(W)) \leq P(QS)$ with $P(S) - P(W) < 1$, so $P(W) > P(S)$

- $Z < 1 \leq R$ for Strong + Strong where $Z < 1 \Rightarrow P(QW) < P(QS)$ and $1 \leq R \Rightarrow 1 + P(S) - P(W) \leq 1 \Rightarrow P(S) \leq P(W)$

- $1 \leq R < Z$ for Weak + Weak which signifies $1 \leq R \Rightarrow P(S) \leq P(W)$ and $R < Z \Rightarrow P(QS) < P(QW)(1 + P(S) - P(W))$

These are represented in the figure below in terms of:
$P(AS) = P(QS)$ and $P(AW) = P(QW)$.

A graphical overview of the three strategies

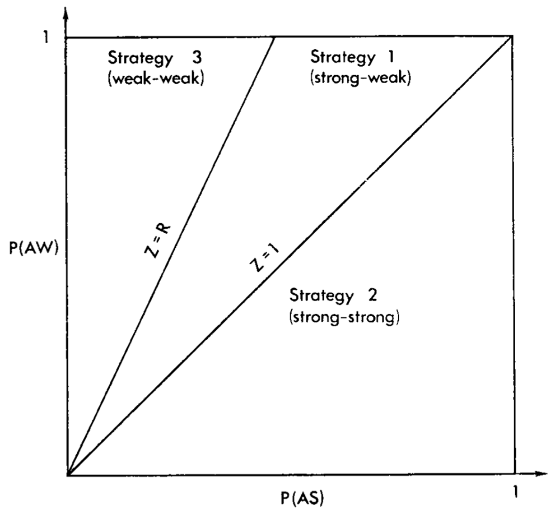


FIG. 1. Optimality regions.

S.L. George (1973) Optimal Strategy in Tennis: A Simple Probabilistic Model. Series C (Applied Statistics) Vol. 22, No. 1, pp. 97-104 - figure on p.99 section: 3. Probability Model

Explanation of the figure:
The horizontal axis represents $P(AS)$, the probability of serving a strong good serve and winning the point, with values lying between $[0;1]$.
The vertical axis represents $P(AW)$, the probability of serving a weak good serve and winning the point, with values lying between $[0;1]$.
Since $P(AS)$ and $P(AW)$ lie between $[0;1]$, all points lie in the square $[0;1] \times [0;1]$.
The relation between the probability of winning a point using a weak serve and a strong serve is defined above as: $Z = \frac{P(AW)}{P(AS)}$, which is the slope of all the lines. The diagonal of the square shows the case $Z = 1$.
Looking at the three service strategies above, we can observe that only for the case Strong+Strong: $Z < 1$. The outcome of that strategy therefore lies underneath the diagonal. In this case, $P(AS)$ is big and $P(AW)$ is small, therefore the player uses his/her strong serve twice.
The results of the other two methods lie above the diagonal.
Apart from Z, we used $R = \frac{1}{1+P(S)-P(W)}$ to define the three strategies. This additional information makes it possible to define the areas of the outcome of the two remaining strategies.
The segment $Z = R$ is drawn to distinguish between Weak+Weak where $R < Z$, therefore lying above the segment $Z = R$, and Strong+Weak where $Z \leq R$, therefore lying underneath the segment $Z = R$.
Optimal usage of the strategies results in the identification of the three regions corresponding to them.

Interpretation of the figure:
Assume we want to advise a player on an optimal serving strategy. We observe how likely it is that he/she wins the point after a strong serve, $P(AS)$, and after a weak serve, $P(AW)$. We can then associate the two dimensional point $p := (P(AS), P(AW))$ with this player. Depending on the zone in which this point lies, we can recommend the appropriate strategy for that athlete. For example, assume $p = (0.6, 0.2)$. We can see this player is far more likely to win a point after a strong serve than a weak one. In this case, p lies in Strategy 2, where the optimal serving strategy for this athlete would then be Strong-Strong.

4.3 Results and Conclusions

The Strong-Weak strategy is not the only one with a large surface but it is the strategy used by most players as they try to avoid double faults. They could also use the Strong-Strong strategy. But in fact for most situations $Z > 1$ (more likely to get the weak serve inside the service area) and R is large enough so one can understand that most players use the Strong-Weak strategy. By looking at the figure we can conclude that the classic strategy is not the only perfect solution. Depending on the situation and the fixed probabilities as well as Z and R players could go for a different strategy.

5 J.S. Croucher (1986) - The Conditional Probability of Winning Games of Tennis

5.1 The main goal

A tennis match consists of a lot of ball exchanges, resulting in points to earn. But not every point has the same importance with regard to winning the game. In order to rank the points, Croucher set up the conditional probabilities of a server winning a game from any possible scoreline.

5.2 Calculations

We suppose that Player A faces Player B.
Furthermore, Player A serves and wins each point with a constant probability p or loses with probability $1 - p = q$. The points are independent from one another.
From Hsi and Burych [2] we know these two formulas:
The probability that player A wins a game with a score of $4 - k$ and $k \in \{0; 1; 2\}$ is:

$$\boxed{\binom{3+k}{k} \cdot p^4 \cdot (1-p)^k}$$

and

the probability that A wins the game after a deuce:

$$\boxed{\binom{6}{3} \cdot p^5 \cdot (1-p)^3 \sum_{j=0}^{\infty} \{2p(1-p)\}^j = \left[\binom{6}{3} \cdot p^5 \cdot (1-p)^3\right] / (1 - 2p + 2p^2)}$$

Let's establish the individual probabilities of Player's A chance of winning a game from a particular score.
Let's start by looking at the probabilities where one knows that Player A wins the game with 0,1,2 lost points or wins via deuce.

- $P(A$ wins the game and loses 0 points/ wins to 0$)= p^4$
- $P(A$ wins the game and loses 1 point/ wins to 15$)= 4p^4q$
- $P(A$ wins the game and loses 2 points/ wins to 30$)= 10p^4q^2$
- $P(A$ wins the game via deuce$)= (20p^3q^3p^2)/(p^2 + q^2)$

As a reminder:

$$\begin{aligned}
&(20p^3q^3p^2)/(p^2 + q^2) \\
&= (20p^5q^3)/(p^2 + (1-p)^2) \\
&= (20p^5q^3)/(p^2 + (1 - 2p + p^2) \\
&= \left[\binom{6}{3} \cdot p^5 \cdot (1-p)^3\right] / (1 - 2p + 2p^2)
\end{aligned}$$

We get: $P(A \text{ wins the game}) = p^4 + 4p^4q + 10p^4q^2 + (20p^3q^3p^2)/(p^2 + q^2)$

Now we want to establish the individual probabilities from a particular score.
Note: the author refers to $p^* = p^2/(p^2+q^2)$ winning from deuce, it is a part of the probability of winning via deuce.

Table 1
Conditional Probability Formulas for a Server Winning a Game

Current score	Probability server wins game
0–0	$p^4(1 + 4q + 10q^2) + 20p^3q^3p^*$
15–0	$p^3(1 + 3q + 6q^2) + 10p^2q^3p^*$
30–0	$p^2(1 + 2q + 3q^2) + 4pq^3p^*$
40–0	$p(1 + q + q^2) + q^3p^*$
0–15	$p^4(1 + 4q) + 10p^3q^2p^*$
15–15	$p^3(1 + 3q) + 6p^2q^2p^*$
30–15	$p^2(1 + 2q) + 3pq^2p^*$
40–15	$p(1 + q) + q^2p^*$
0–30	$p^4 + 4p^3qp^*$
15–30	$p^3 + 3p^2qp^*$
30–30	p^*
40–30	$p + qp^*$
0–40	p^3p^*
15–40	p^2p^*
30–40	pp^*
40–40	p^*

Note. $q = 1 - p$ (probability of receiver winning the point)
p = probability of server winning point.
p^* = probability that server wins game from a score of deuce

J.S. Croucher (1986) The Conditional Probability of Winning Games of Tennis. Research Quarterly for Exercise and Sport, 57:1, pp. 23-26 - table on p.24

In this table are presented all the possible scores and the probabilities that Player A wins from that score.
Note that these probabilities are equivalent and therefore not in the table:

- $30-40$ and *"advantage Player B/receiver"*
- $40-30$ and *"advantage Player A/server"*

  Justification:

  Let's calculate some of these probabilities to show how these are put together.

- $0-0$: We have calculated that probability above.
- $30-0$: We have to look at different cases:

  $P(A$ wins to 0 | current score 30-0)
  $= P(A$ wins the next two points)
  $= p^2$

  $P(A$ wins to 15 | current score 30-0)
  $= P(A$ wins exactly two of the next three points, one of them is for sure the last one)
  $= 2p^2q$

  $P(A$ wins to 30 | current score 30-0)
  $= P(A$ wins exactly two of the next four points, one of them is for sure the last one)
  $= 3p^2q^2$

  $P($ Score reaches deuce | current score 30-0)
  $= P(A$ wins exactly one of the next four points)
  $= 4pq^3$

$P(A$ wins via deuce | current score 30-0)
$= 4pq^3p^*$

So $P(A$ wins a game | current score 30-0)
$= p^2 + 2p^2q + 3p^2q^2 + 4pq^3p^*$
$= p^2(1 + 2q + 3q^2) + 4pq^3p^*$

- $15 - 15$ We have to put different probabilities together:

  $P(A$ wins to 15 | current score 15-15)
  $= P(A$ wins the next three points)
  $= p^3$

  $P(A$ wins to 30 | current score 15-15)
  $= P(A$ wins exactly three of the next four points, one of them is for sure the last one)
  $= 3p^3q$

  $P($ Score reaches deuce | current score 15-15)
  $= P(A$ wins exactly two of the next four points)
  $= 6p^2q^2$

  $P(A$ wins via deuce | current score 15-15)
  $= 6p^2q^2p^*$

  So $P(A$ wins a game | current score 15-15)
  $= p^3 + 3p^3q + 6p^2q^2p^*$
  $= p^3(1 + 3q) + 6p^2q^2p^*$

- $15 - 30$:

  $P(A$ wins to 30 | current score 15-30)
  $= P(A$ wins the next three points)
  $= p^3$

  $P($ Score reaches deuce | current score 15-30)
  $= P(A$ wins exactly two of the next three points)
  $= 3p^2q$

  $P(A$ wins via deuce | current score 15-30)
  $= 3p^2qp^*$

  So $P(A$ wins a game | current score 30-0)
  $= p^3 + 3p^2qp^*$

- $30 - 40$: $P($ Score reaches deuce | current score 30-40)
  $= P(A$ wins the next point)
  $= p$

  $P(A$ wins via deuce | current score 30-40)
  $= pp^*$

- $30 - 30$ and $40 - 40$: They are exactly p^* as the only way for player A to win the game he/she has to win two consecutive points or otherwise the game continues in deuces.

Now that we have found the formulas let's turn to the importance of the points.
In the beginning of this section we have set p as a constant probability. In the real world every tennis player has his/her own value p. Let's therefore look at the outcome of the above established formulas for values of p between .30 and .80.

Table 2
Probability of Server Winning Game

Current Score	Probability server wins each point										
	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80
0–0	.099	.170	.264	.377	.500	.623	.736	.830	.901	.949	.978
15–0	.211	.311	.424	.542	.656	.758	.842	.905	.949	.976	.990
30–0	.412	.523	.631	.729	.813	.879	.927	.960	.980	.991	.997
40–0	.710	.787	.850	.900	.938	.963	.980	.990	.996	.998	.999
0–15	.051	.095	.158	.242	.344	.458	.576	.689	.789	.870	.930
15–15	.125	1.96	.286	.389	.500	.611	.714	.804	.875	.928	.964
30–15	.284	.381	.485	.589	.688	.775	.847	.903	.944	.970	.986
40–15	.586	.672	.751	.819	.875	.919	.951	.972	.986	.994	.998
0–30	.020	.040	.073	.121	.188	.271	.369	.477	.588	.696	.795
15–30	.056	.097	.153	.225	.313	.411	.515	.619	.716	.802	.873
30–30	.155	.225	.308	.401	.500	.599	.692	.775	.844	.900	.941
40–30	.409	.496	.585	.671	.750	.820	.877	.921	.953	.975	.988
0–40	.004	.009	.020	.037	.063	.100	.150	.213	.290	.380	.481
15–40	.013	.028	.049	.081	.125	.181	.249	.328	.414	.506	.602
30–40	.047	.079	.123	.180	.250	.329	.415	.504	.592	.675	.753
40–40	.155	.225	.308	.401	.500	.599	.692	.775	.844	.900	.941

J.S. Croucher (1986) The Conditional Probability of Winning Games of Tennis. Research Quarterly for Exercise and Sport, 57:1, pp. 23-26 - table on p.25

<u>Justification:</u>

Let's calculate some values of scores using the formulas from Table 1. Let's look at $p = .40$ and $q = .60$:

- $0 - 0$
  $0.40^4 \cdot (1 + 4 \cdot 0.60 + 10 \cdot 0.60^2) + 20 \cdot 0.40^3 \cdot 0.60^3 \cdot 0.40^2/(0.40^2 + 0.60^2)$
  $= 0.1792 + 0.8507076923$
  $= 0.2642707692$
  $\cong 0.264$

- $0 - 30$
  $0.40^4 + 4 \cdot 0.40^3 \cdot 0.60 \cdot 0.40^2/(0.40^2 + 0.60^2)$
  $= 0.0256 + 0.04726153846$
  $= 0.07286153846$
  $\cong 0.07$

- $30 - 40$
  $0.40 \cdot 0.40^2/(0.40^2 + 0.60^2)$
  $= 0.1230769231$
  $\cong 0.123$

These values are easy to calculate: take the corresponding p value and calculate $1 - p = q$, then insert those in the formulas from Table 1.
Some interesting values are colored.
<u>$15 - 40 < 0 - 30$</u> and <u>$30 - 40 < 15 - 30$</u>
<u>$30 - 0 < 40 - 15$</u> and <u>$40 - 30 < 30 - 15$</u>
<u>$30 - 30 < 15 - 15$ if $p > .50$</u> and <u>$15 - 15 < 30 - 30$ if $p < .50$</u>

Those inequalities are very intuitive, let's consider for example the first one: $15-40 < 0-30$. The probability that the server wins a point is more likely if his/her opponent is not so close to winning the game. The player is less stressed about doing a mistake and goes "all in", without holding back. If the server plays as good as he/she can, he/she makes it more difficult for his/her opponent to win a point. Player A risks losing a point at $0-30$ in comparison to losing a game at $15-40$. At $15-40$ the server plays safer and makes it easier for player B to win the point than at $0-30$.

5.3 Results and Conclusions

We have gathered enough information to be able to calculate the importance and rank the points according to their value in each column for the value of p and $q = 1-p$. These give the players A and B the importance of the points according to their constant probability p or q.

Table 3
Importance of each Point (Ranking in Parentheses)

Current score	Probability server wins each point				
	.30	.35	.40	.45	.50
0–0	.160 (10)	.216 (11)	.266 (10)	.300 (11)	.313 (8)
15–0	.287 (8)	.327 (8)	.346 (7)	.340 (7)	.313 (8)
30–0	.426 (4)	.406 (6)	.366 (6)	.311 (10)	.250 (11)
40–0	.414 (5)	.328 (7)	.249 (12)	.181 (14)	.125 (15)
0–15	.105 (13)	.156 (13)	.213 (13)	.267 (12)	.313 (8)
15–15	.228 (9)	.285 (9)	.332 (8)	.364 (6)	.375 (5)
30–15	.431 (3)	.448 (3)	.443 (4)	.418 (4)	.375 (5)
40–15	.591 (2)	.504 (2)	.415 (5)	.329 (8)	.250 (11)
0–30	.052 (14)	.087 (14)	.133 (14)	.189 (13)	.250 (11)
15–30	.141 (12)	.197 (12)	.258 (11)	.320 (9)	.375 (5)
30–30	.362 (6)	.417 (4)	.462 (2)	.490 (2)	.500 (1)
40–30	.845 (1)	.775 (1)	.692 (1)	.599 (1)	.500 (1)
0–40	.014 (16)	.028 (16)	.049 (16)	.081 (16)	.125 (15)
15–40	.047 (15)	.079 (15)	.123 (15)	.180 (15)	.250 (11)
30–40	.155 (11)	.225 (10)	.308 (9)	.401 (5)	.500 (1)
40–40	.362 (6)	.417 (4)	.462 (2)	.490 (2)	.500 (1)

J.S. Croucher (1986) The Conditional Probability of Winning Games of Tennis. Research Quarterly for Exercise and Sport, 57:1, pp. 23-26 - table on p.26

Table 3 (continued)
Importance of each Point (Ranking in Parentheses)

Current score	Probability server wins each point					
	.55	.60	.65	.70	.75	.80
0–0	.300 (11)	.266 (10)	.216 (11)	.160 (10)	.105 (10)	.060 (10)
15–0	.267 (12)	.213 (13)	.156 (13)	.105 (13)	.063 (13)	.033 (13)
30–0	.189 (13)	.133 (14)	.087 (14)	.052 (14)	.028 (14)	.013 (14)
40–0	.081 (16)	.049 (16)	.028 (16)	.014 (16)	.006 (16)	.002 (16)
0–15	.340 (7)	.346 (7)	.327 (8)	.287 (8)	.232 (8)	.169 (8)
15–15	.364 (6)	.332 (8)	.285 (9)	.228 (9)	.169 (9)	.113 (9)
30–15	.320 (9)	.258 (11)	.197 (12)	.141 (12)	.094 (12)	.056 (12)
40–15	.180 (15)	.123 (15)	.079 (15)	.046 (15)	.025 (15)	.012 (15)
0–30	.311 (10)	.366 (6)	.406 (6)	.426 (4)	.422 (4)	.392 (4)
15–30	.418 (4)	.443 (4)	.448 (3)	.431 (3)	.394 (5)	.339 (5)
30–30	.490 (2)	.462 (2)	.417 (4)	.362 (6)	.300 (6)	.235 (6)
40–30	.401 (5)	.308 (9)	.225 (10)	.155 (11)	.100 (11)	.059 (11)
0–40	.181 (14)	.249 (12)	.328 (7)	.414 (5)	.506 (3)	.602 (3)
15–40	.329 (8)	.415 (5)	.504 (2)	.591 (2)	.675 (2)	.753 (2)
30–40	.599 (1)	.692 (1)	.775 (1)	.845 (1)	.900 (1)	.941 (1)
40–40	.490 (2)	.462 (2)	.417 (4)	.362 (6)	.300 (6)	.235 (6)

J.S. Croucher (1986) The Conditional Probability of Winning Games of Tennis. Research Quarterly for Exercise and Sport, 57:1, pp. 23-26 - table on p.26

The most important point/s is/are highlighted in red.
The next essential point/s is/are covered in yellow.
The third most crucial point is coloured in green.

Justification:

Let's calculate the importance for some points using Table 2.

- Importance of the point: $30 - 15$ with $p = .30$
  $P(A$ wins game from $40 - 15) - P(A$ wins game from $30 - 30)$
  $= .586 - .155$
  $= .431$

- Importance of the point: $40 - 30$ with $p = .40$
  $P(A$ wins game after scoring the point $40 - 30)$ - $P(A$ wins game from $40 - 40)$
  $= 1 - .308$
  $= .692$

- Importance of the point: $40 - 40$ with $p = .50$
  $P(A$ wins game from $40 - 30) - P(A$ wins game from $30 - 40)$
  $= .750 - .250$
  $= .500$

- Importance of the point: $30 - 40$ with $p = .75$
  $P(A$ wins game from $40 - 40)$
  $= .900$

We have the conditional probabilities for player A to win the game and the corresponding importance. We can now evaluate these two values during a match by constantly adjusting p.

6 Conclusion

Tennis research has made significant progress and undergone substantial development over time. I am confident that the expansion of tennis research will continue as technology advances and more accurate data becomes accessible. Such progress will be advantageous for both players and coaches, enabling them to make more informed decisions regarding training and reducing the risk of injuries. Additionally, further research advancements will likely lead to enhanced predictions, which is another positive outcome.

References

[1] D. Gale (1971) Optimal Strategy for Serving in Tennis. *Mathematics Magazine Volume 44 - Issue 4, pp. 197-199*

[2] B.P. Hsi and D.M. Burych (1971) Games of Two Players. *Journal of the Royal Statistical Society Series C Vol.20, No.1, pp. 86-92*

[3] S.L. George (1973) Optimal Strategy in Tennis: A Simple Probabilistic Model. *Journal of the Royal Statistical Society. Series C (Applied Statistics) Vol. 22, No. 1, pp. 97-104*

[4] J.S. Croucher (1986) The Conditional Probability of Winning Games of Tennis. *Research Quarterly for Exercise and Sport, 57:1, pp. 23-26*

[5] S.R. Clarke, D. Dyte (2000) Using offical ratings to simulate major tennis tournaments. *International Transactions in Operational Research 7(6), pp. 585-594*

[6] Y. Liu (2001) Random walks in tennis. *Missouri Journal of Mathematical Sciences 13(3), pp. 154-162*

[7] F. Klaassen, J. Magnus (2003) Forecasting the winner of a tennis match. *European Journal of Operational Research 148 (2), pp. 257-267*

[8] P.K. Newton, G.H. Pollard (2004) Service neutral scoring strategies for tennis. *Proceedings of the Seventh Australasian Conference on Mathematics and Computers in Sport, Massey University Palmerston North, New Zealand, pp. 221-225*

[9] T. Barnett and S.R. Clarke (2005) Combining Player Statistics to Predict Outcomes of Tennis Matches. *IMA Journal of Management Mathematics 16(2), pp. 113–120*

[10] P.K. Newton, J.B. Keller (2005) Probability of Winning at Tennis I. Theory and Data. *Studies in Applied Mathematics Volume 114 Issue 3, pp. 241-269*

[11] T. Barnett, A. Brown and S. Clarke (2006) Developing a model that reflects outcomes of tennis matches. *Proceedings of the 8th Australasian Conference on Mathematics and Computers in Sport, Coolangatta, Queensland, 3-5 July 2006, pp. 178-188*

[12] G. Hunter, A. Shihab and K. Zienowicz (2008) Modelling Tennis Rallies Using Information from both Video and Audio Signals. *available on https://eprints.kingston.ac.uk/id/eprint/8394*

[13] P.K. Newton, K. Aslam (2009) Monte Carlo Tennis: A Stochastic Markov Chain Model. *Journal of Quantitative Analysis in Sports Volume 5, Issue 3 Article 7*

[14] A. Bedford, T. Barnett, Gr. Pollard, Ge. Pollard (2010) How the Interpretation of Match Statistics Affects Player Performance. *Journal of Medicine and Science in Tennis 15(2), pp. 23-27*

[15] D. Paindaveine, Y. Swan (2011) A stochastic Analysis of some Two-Person Sports. *Studies in Applied Mathematics 127, pp. 221-249*

[16] F. Radicchi (2011) Who is the best player ever? A complex network analysis of the history of professional tennis. *available on https://doi.org/10.1371/journal.pone.0017249*

[17] T. Barnett, D. O'Shaughnessy, A. Bedford (2011) Predicting a tennis match in progress for sports multimedia. *OR Insight 24(3), pp. 190-204*

[18] W.J. Knottenbelt, D. Spanias, A.M. Madurska (2012) A common-opponent stochastic model for predicting the outcome of professional tennis matches. *Computers and Mathematics with Applications Volume 64, Issue 12, pp. 3820-3827*

[19] C. Roure (2014) What are the key points to win in Tennis? *ITF Coaching and Sport Science Review 2014; 64(22), pp. 14-15*

[20] M. Bevc (2015) Predicting the Outcome of Tennis Matches From Point-by-Point Data. *University of Glasgow, School of Computing Science*

[21] C. Gray (2015) Game set and stats. *Significance Volume 12 Issue 1, pp.28-31*

[22] C. Cooper and R.E. Kennedy (2021) Expected length and probability of winning a tennis game. *Cambridge University Press, pp. 490-500*

[23] A. Sarcevic, M. Vranic, D. Pintar (2021) A Combinatorial Approach in predicting the Outcome of Tennis Matches. *International Journal of Applied Mathematics and Computer Science, Vol. 31, No. 3, pp. 525-538*

[24] J.C. Yue, E.P. Chou, M.-H. Hsieh, L.-C. Hsiao (2022) A study of forecasting tennis matches via the Glicko model. *PLoS One 17(4) e0266838*

[25] S.A. Kovalchik, J. Albert (2022) A Statistical Model of Serve Return Impact Patterns in Professional Tennis. *available on arXiv:2202.00583 [stat.ME]*

[26] F. Rothe, M. Lames (2022) Simulation of Tennis Behaviour Using Finite Markov Chains. *IFAC PapersOnLine 55-20, pp. 606-611*

[27] C. Ley, Y. Dominicy (2023) Statistics Meets Sports: What we can learn from Sports Data. *Cambridge Scholars Publishing; 1st edition (March 1, 2023) ISBN: 1-5275-9273-1*