Victoria Olejarz
Peyton Pepkowski
Zuzu Trottier
Dr. Greg Matthews
STAT 321 001
10 December 2024

Simulation Project Report: Singles Tennis

Introduction

The game of tennis is an international pastime in which two individuals compete against each other until a winner is decided. In most sports, a winner is determined solely based on the individual or team who gains the greatest number of points throughout the duration of the match. In tennis, however, the overall score of a match is broken down into three subcategories: points, games, and sets. To earn a point, a player must win a single rally, or bout of back-and-forth hitting, by forcing the other player to miss their shot. To earn a game, a player must win four points and win those four points by at least two points more than their opponent, which may result in playing beyond four points until one player is up by two. To earn a set, a player must win six games by at least two games, else the set is decided by a tiebreak of seven points which, again, must be won by two points. An entire match is usually decided in a best two out of three set scenario, but in some cases, namely men's Grand Slam tournaments, a match is played best three out of five sets. Because of this scoring setup, it is possible for a player to win less than 50% of the *points* that make up a match and still win the match overall based on the distribution of the points they did win.

Every point in the game of tennis begins with one player serving the ball by tossing the ball above their head and swinging downward to send the ball into play on their opponent's half. Serve is usually an advantage in the sport, as most professional players have a higher win percentage when serving versus not serving. For example, Rafael Nadal, a recently-retired tennis great, has an overall point win percentage of 67% when serving compared to a 42% when not serving, or returning (*Rafael Nadal*). Aryna Sabalenka, the current women's World #1 player, has an on-serve point win percentage of 62.7% compared to an off-serve percentage of 45.3% (*Aryna Sabalenka*). Since the serve sets the pace for the rest of the point, it is an advantage for the serving player since they can dictate how they'd like the point to go based on speed, placement, amount of backspin put on the ball, and other factors. The serve also tends to be one the hardest-hit balls during a tennis point, making the serve returns one of the most difficult feats of the entire match.

Plenty of mathematical work has been done on the game of tennis and how its abnormal scoring plays into the outcomes of a match. Cooper and Kennedy (2021) looked into the length as well as probability of winning a tennis match by determining the different states of a match and the probability of getting to one from another based on a set p of Player 1 winning a point. This study focused more on the mathematical concepts of a tennis match and how probabilities affect length more than outcome. Kovalchik (2016) performed a systematic review on different forecasting models of predicting the outcomes of tennis matches, finding eleven different eligible models for prediction. Models are based on player-opponent ranking drives, which are representative of a player's win-loss record throughout the past year. Newton and Aslam (2009) developed a stochastic Markov chain model to predict a player's performance in a tennis match based on probability of winning a point and the consistency of a player's performance based on a standard deviation. Research has expanded on this topic due to the increased interest in the sport altogether as well as the uprising in sports betting on professional tennis tournaments.

This study aims to use simulation techniques to determine how frequently the winner of a singles tennis match wins less than 50% of the points in a match, as well as how different probabilities of each player winning a point affect those results. It will also attempt to determine how serve percentages and order may affect the outcomes and how much of an advantage serve order really may be. By adjusting point win percentages, serve percentages, serve order, length of a match, and various other factors, this study will attempt to find what it takes to win a tennis match by utilizing the least amount of effort in regards to winning individual points.

Methods

We began by creating the function play_point to simulate individual points in a tennis match. This function generates a random number between 0 and 1, and compares it to a given probability. If the random number is less than the probability, Player A wins the point; otherwise, Player B wins the point. This simple approach allows us to model the outcome of a single point based on the probability of each player winning that point.

Building on this, we developed the functions play_game, play_set, play_match, and simulate_matches, which simulate more complex aspects of a tennis match. These functions use Monte Carlo simulation, which involves generating random outcomes based on probabilities and then repeating the process many times to estimate statistical outcomes. By simulating multiple games, sets, and matches in this way, we can observe the overall behavior of matches under different conditions such as varying point probabilities. The Monte Carlo approach allows us to estimate the likelihood of various match outcomes, such as the probability of winning a set or a match, based on the probability of winning individual points.

In addition to looking at the likelihood of winning based on points won, we also looked at how serve percentages played into the probability of a player winning the match. We ran a simulation of serve percentage (50-100% by increments of 5%) versus the rank of the player's opponent (1-100 by increments of 1) and created a heatmap to display the probability of winning a match based on these factors.

We also wanted to determine if serve order plays a role in the outcome of a tennis match. We simulated a tennis match as a function of the probability the "server" (versus the "returner") wins the point, game, set, and match. We set the probability at 0.5, meaning both players had an equal probability of winning the point. We then simulated multiple trials and found the average percentage of match wins based on which player served first, with the "server" remaining consistent throughout the simulations. This gave us the proportion of wins when the “server” served first and the proportion when the “returner” served first of both the server and returner winning the match.

Results

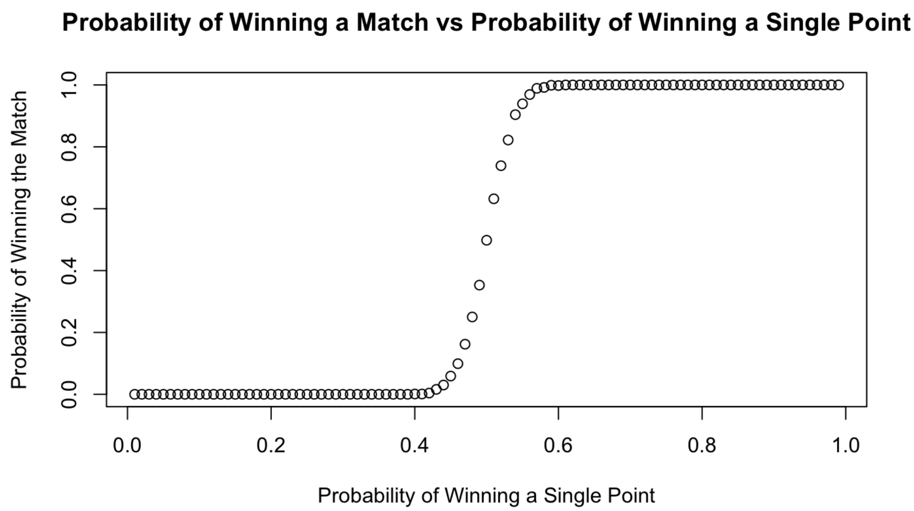
The simulation, which ran 10,000 matches with both players having a 50% chance of winning a point, provides insight into how games, sets, and matches unfold. On average, the winner wins about 53.43% of the points during a match, but this varies depending on the match. In some matches, the winner may only win 45.05% of the points, while in others, they win as much as 69.23%. This suggests that while the winner generally wins more points, the outcome of a match isn’t solely determined by the total number of points won. Instead, it depends more on winning the right games and sets at critical moments.

Additionally, the simulation shows that in about 7.97% of the matches, the winner ends up winning less than 50% of the total points. This demonstrates that a player can still win a match even if they don’t win the majority of points, as long as they perform better in the key moments, like winning crucial games and sets. Overall, these results highlight how matches in tennis are shaped more by winning the right games and sets rather than simply accumulating points.

Adjusting the probability each player wins a point can show us the relationship between probability of winning a point and percentage of points won during a match by the winner. When the probabilities are adjusted to Player A having a 60% probability of winning a point and Player B having a 40% probability, the average percentage of points won by the winner shoots up to 60.90%, with Player A winning 60.92% of the points on average when they win and Player B winning an average of 49.82% of the points when they win. This shows that Player A wins a majority of the time since their percentage is closer to the average of both players, and that in the fewer instances Player B wins the match, they actually win less than half of the points. When the percentages are shifted further to 70% and 30%, Player A wins approximately 70% of the points and wins every single match simulated in the 10,000 draws. These changes in proportion of points won based on probability a player wins a point show that even though a player may be of a higher ability than another, it is still possible for the underdog to win by winning the most important points of a match.

Understanding how the probability of winning individual points affects the overall outcome of a match can provide valuable insights into competitive strategies. The plot below visualizes the relationship between the probability of winning a single point and the probability of winning a match.

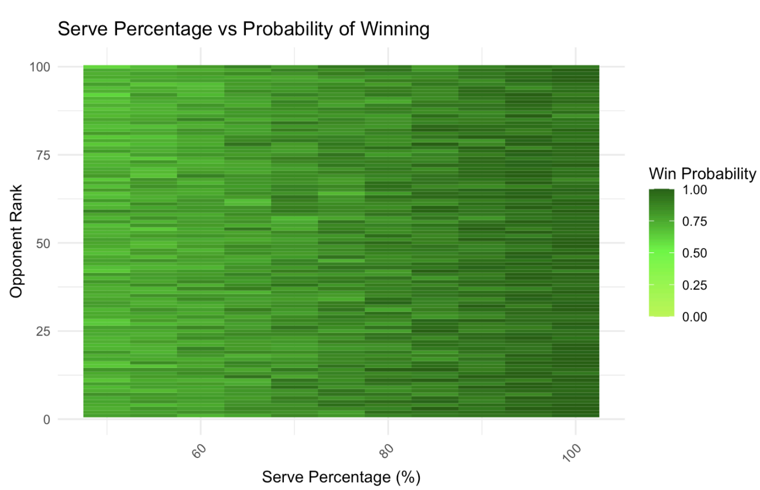
Figure 1



In the graph, the x-axis represents the probability of winning a single point. The y-axis shows the probability of winning the match. The trend is clear: at low probabilities of winning a single point, the chance of winning the match is nearly zero. However, as the probability of winning a single point increases, particularly around the critical threshold of 50%, the probability of winning the match rises significantly. This creates an S-shaped curve, demonstrating that small improvements in the probability of winning points around this 50% mark can dramatically increase the likelihood of winning the entire match. The plot underscores the importance of consistently winning points and how even marginal gains in this area can lead to significant competitive advantages.

In regards to how serve affects the outcome of a tennis match, the heat map displaying serve percentage vs probability of winning is shown here.

Figure 2

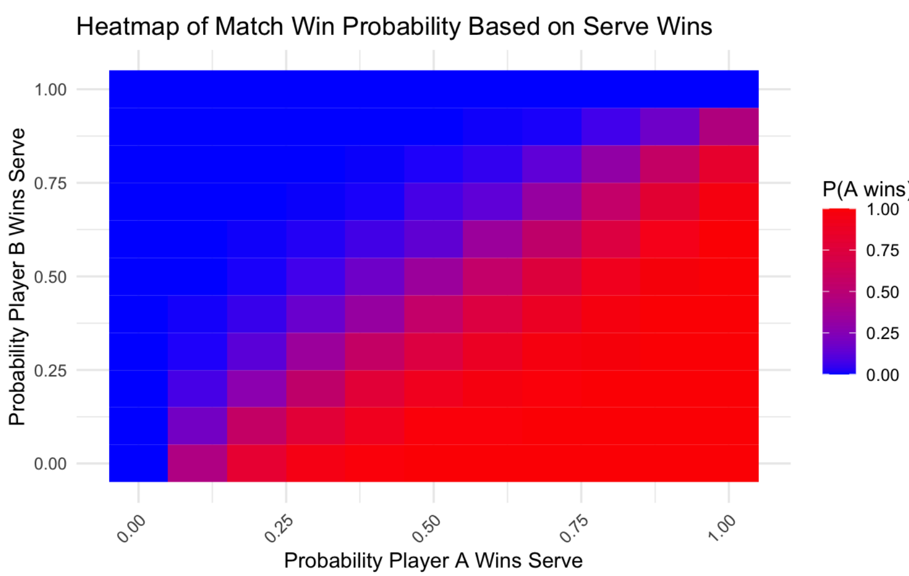


Based on the results of the heat map, it can be seen that as serve percentage increases, win probability increases alongside it. The map clearly shows a darker hue towards the right side, correlating with a higher win probability overall. This effect is seen almost despite the rank of the opponent, showing the powerful effect of a player being consistent on their serve and getting the point started at their chosen pace. Win probability is also very low towards the lower end of serve percentage (50%), again showing the importance of serve for a player's game. It almost seems as though any influence of the rank of the opposing player is overridden by the player's serve percentage, as the hue of the map shifts from light to dark across the board.

When looking at serve order and its effects on the outcome of a match, there did not seem to be a significant difference in the proportion of wins based on which player served first. When the "server" served first, their win proportion was 0.504 compared to when the "returner" served first, their win proportion was 0.518. When the "returner" served first, their proportion was 0.482 versus 0.496 when the "server" served first. A z-test was run on these results which gave a p-value of 0.531, providing evidence for the insignificance of serve order on match outcome.

As for the difference between Player A winning their serve versus Player B, the heat map below displays the probability of Player A winning the serve on the x-axis, Player B winning the serve on the y-axis, and the probability that Player A wins the match based on the probability the serve was won.

Figure 3



This heat map reveals that as the probability that Player A wins their serve increases, the probability that they win the match increases. At a probability of 1 for Player A winning their serve, the probability that they win the match is also 1. This is the same for Player B when they win their serve at a probability of 1. There looks to be a linear trend throughout the middle of the heat map. When both Player A and Player B have a similar probability to win their serve, the outcome of the match tends to be more balanced; there is not a strong bias towards either of the players. When both players have a similar likelihood to win the serve, the outcome of the match is generally more unpredictable.

Conclusion

We began our study by creating functions to simulate a tennis game, set, and match using Monte Carlo simulation. Using these functions, we simulated 10,000 matches with both players having an equal chance of winning the point, revealing that the match winner won about 53.43% of the points during the match, but results varied between a range of 45.05% to 69.23%. The given results indicate that the match result is not determined only by the total number of points won, but winning games and sets at pivotal moments. The simulation also revealed that the match winner won less than half of the total points 7.97% of the time, suggesting that a player can win a match even when they do not win most of the points.

Additionally, we adjusted the probabilities of certain players winning a point. When Player A had a 60% chance to win the point and Player B 40%, the average percentage of points won by the match winner increased to 60.90%. This showed that Player A would win most of the time since their win percentage was closer to the average points won by the match winner; when Player B would win the match, they would win less than half of the points. A similar result was found when the percentages were shifted to 70% and 30%, which further supported the conclusion that a match can still be won by a player who won fewer points.

We then created a graph showing the trend between the probability of winning a single point against the probability of winning the match. The graph indicated that at the lower probabilities of winning a single point, the probability of winning the match was almost zero. As the probability of winning a single point increased, the probability of winning the match did as well, creating an S-shaped curve. When the probability of winning a single point is around 50%, the probability of winning the match differs significantly. However, the plot does not account for a player consistently winning points, which can lead to significant advantages in a match.

Furthermore, we generated a heat map to visualize how serve percentage and opponent rank contributed to the probability of winning. This chart revealed that as serve percentage increased, win probability increased as well, almost disregarding the rank of the opponent. This indicated that the more consistent a player is with their serve can be instrumental in the match outcome, no matter the ranking of the opponent. The trend is similar at serve percentages of 50% and below, suggesting a lower win probability associated with a lower serve percentage, again overlooking the opponent's rank.

When looking at the effect of serve order on the outcome of a match, we found no indication of significance. When the designated "server" served first, their win probability was 0.504; when the "returner" served first, their win probability was 0.518. Then when the "returner" served first, their win probability was 0.482 compared to the "server," with a probability of 0.496. We ran a z-test with results evidently suggesting an insignificance on the order of serving on match outcome.

Lastly, we analyzed the difference in win probability based on which player won their serve. We created a heat map that indicated when a player has a probability of winning their serve that is close to 1, the probability that they win the match is also close to 1. There was a linear trend where both Player A and Player B had a similar chance to win their serve, conveying a more balanced match outcome. When both Players have a similar probability to win their serve, it is more difficult to clearly forecast who will win the match.

Overall, our study revealed that a player does not have to win a majority of the points in a tennis match to win the match itself, with serve order not playing a significant role but winning the serve seeming to be important. In the future, it would be beneficial to analyze the certain strengths and weaknesses of different players - such as how well they play offensively versus defensively - and note how those characteristics impact the result of a match as well. It may also be interesting to look at the consistency in one's level of play and if it differs based on certain factors, such as the court surface. Court surface could play a role in many other aspects of the match as well, so this would be an intriguing element to explore. Additionally, we ran our simulations using only three sets, but in men's Grand Slam tournaments a match is played to the best three out of five sets. Increasing the number of sets to five could make a difference in our results and in future analyses. We also did not look into how a tiebreak format could affect the outcome of a match, so future work should include tiebreakers to see how those, which are decided solely on number of points won, play into the proportion of matches won by each player.

References

Aryna Sabalenka: Player stats & more – WTA official. Women's Tennis Association. (n.d.). https://www.wtatennis.com/players/320760/aryna-sabalenka

Cooper, C., & Kennedy, R. E. (2021). Expected length and probability of winning a tennis game. *The Mathematical Gazette*, *105*(564), 490–500. doi:10.1017/mag.2021.117

Kovalchik, S. (2016). Searching for the GOAT of tennis win prediction. *Journal of Quantitative Analysis in Sports*, *12*(3), 127-138. https://doi.org/10.1515/jqas-2015-0059

Newton, P. & Aslam, K. (2009). Monte Carlo Tennis: A Stochastic Markov Chain Model. *Journal of Quantitative Analysis in Sports*, *5*(3). https://doi.org/10.2202/1559-0410.1169

Rafael Nadal: Player stats: ATP tour: Tennis. ATP Tour. (n.d.). https://www.atptour.com/en/players/rafael-nadal/n409/player-stats?year=all&surface=all