Michelle Okereke

Mathematics of Sports

Leonid Pekelis

11 December 2014

Extended Abstract: First Server Advantage in Tennis

Unlike football, where each team starts one half kicking and the other half receiving, there is no guarantee in tennis that a player will get to serve first for any set. The order of **serve** and **receive** is determined at the start of the match with a coin toss. The winner of this coin toss either chooses who serves first, or which side of the court to start on. Typically, a winner either chooses to serve first, or implicitly chooses to serve second by opting to pick a side instead. If given the option, what should you choose? Will serving first or second put you in the best position to win? This is the question that this project sought to answer, using principles of **probability**, **combinatorics**, and **recursion**. By building a mathematical model for the likelihood of winning a set, it is determined that if you are more likely to win a game on serve than receive, then you should serve first—regardless of how good your opponent is. Conversely, if you are more likely to win on receive than serve, then you should serve second. Given the fact that the grand majority of players are better on serve than receive, it is advantageous to serve first. Using two simple statistics—the probability of holding a service game and the probability of breaking a receiving game—this model quantifies the magnitude of the **first server advantage**.

Rules of the Game

Currently, women play best of three matches (win two sets to win match) at all Women's Tennis Association (WTA) tournaments, as well as Grand Slams (which are organized by the

International Tennis Federation [ITF]). Men play best of three at all Association of Tennis Professionals (ATP) matches, and best of five (win three sets to win match) at Grand Slams. Across all these matches, there are two types of sets. **Advantage sets** are played in the deciding set (5^{th} set for men, 3^{rd} set for women) of the Australian Open, the French Open, and Wimbledon. **Tie-break sets** are played for all other sets of ATP, ITF, and WTA matches. Both types of sets use similar rules, with slight difference in scoring and service order.

Scoring

The first player to win 6 games by a margin of 2 (i.e. with the score 6-0, 6-1, 6-2, 6-3, or 6-4, 7-5) wins the match. In a tie-break set, if the score line reaches 6-6 without a winner being determined, then the players enter a tie-break game. The first server to win 7 points by a margin of two wins the tie-break game, and wins the set 7-6. In an advantage set, there is no option for a tie-break. Instead, the players continue until one wins by 2 games (i.e. 8-6, 9-7, 10-8…). This is meant to ensure the most consistent player wins, by eliminating the "luck" factor of tie-breaks.

Service Order

In both types of sets, if a player serves the last game, they serve second the next set. If a player receives last, they serve first the next set. So theoretically, a player could serve first in every set. This is the case when each set has an even number of total games. In a tie-break set, the first server for the match also serves first point of the tie-break. Player two serves the next two points, and player one serves the subsequent two points. This continues until a winner is determined. After a tie-break, the first server becomes the second server in the subsequent set.

Research Question

In an advantage set, is the "win by 2" rule enough to mitigate for the advantage of serving first? These types of sets are only played in the deciding set of all Grand Slams but the US Open.

Given the fact that these are the most prestigious tennis tournaments, and that matches that go the distance get the most attention, it is an important question to answer. It is generally accepted that the player who serves first in the deciding set has a "mental" advantage, but this advantage is rarely examined from a probabilistic perspective. If the first server advantage concretely exists, then tournaments should consider making changes to how they mitigate for it. Rather than doing one coin flip to start the match, a coin flip could be done at the beginning of each set.

Methodology

In order to solve this problem, a formula for winning an advantage set is determined, given that a player is serving first or second. This probability can be broken down into two primary components, the probability of winning in 10 games or less, and the probability of winning in over 10 games. Thus, we get the equation:

P(Win) = P(Win in 10 Games or Less) + P(Win in Over 10 Games)

The intuition behind this is that past 10 games, the problem can be analyzed recursively. Before this, a winning score line depends on non-recursive rules. In both cases, we define the probability "S" as P(1^{st} Server Holds a Service Games) and probability "B" as P(1^{st} Server Breaks a Receive Game). Consequently, we get the following probabilities:

- **P(1^{st} Server Wins an Odd-Numbered Game) = S**
- **P(2^{nd} Server Wins an Odd-Numbered Game) = 1 – S**
- **P(1^{st} Server Wins an Even-Numbered Game) = B**
- **P(2^{nd} Server Wins an Even-Numbered Game) = 1 – B**

These probabilities, as well as whether the total number of games is odd or even, is used to determine the formula for each score line.

Ten Games or Less

For a game to end in 10 games or less, the final score line must be 6-0, 6-1, 6-2, 6-3, or 6-4. So:

P(Win in 10 or Less) = P(6-0) + P(6-1) + P(6-2) + P(6-3) + P(6-4)

A player can achieve each of these score lines by winning a combination of odd and even games. Moreover, each of these combinations can be achieved in a number of ways, depending on the number of odd and even games available. The winner of a set must win the final game of that set. So the number of odd and even games "available" is limited by whether the total number of games in that set is odd or even. If there are an odd number of games, then the final odd game of the set must go to the winner. If there is an even number of games, then the final even game of the set must go to the winner. This impacts the combinatorics used to determine the number of "ways" of getting each score line. As an example, we walk through analyzing the score 6-1:

- There are 4 odd games, 3 even games, and the total number of games—7—is odd.
- A player must either win 4 (lose 0) odd games and win 2 (lose 1) even game, or win 3 (lose 1) odd games and win 3 (lose 0) even games.
- The last game must be one of the odd games won. The order of the remaining games won can be varied.
- **P(1st Server Wins 4 Odd, 2 Even) =** $({}_{(4-1)}C_{(4-1)} \times {}_{3}C_{2})* (S^4(1-S)^0B^2(1-B)^1)$
- **P(1st Server Wins 3 Odd, 3 Even) =** $({}_{(4-1)}C_{(3-1)} \times {}_{3}C_{3})* (S^3(1-S)^1B^3(1-B)^0)$
- **P(1st Server Wins 6-1) =** $3(S^4(1-S)^0B^2(1-B)^1) + 3(S^3(1-S)^1B^3(1-B)^0)$
- For 2nd Server, replace all **S** with **(1-S)** and all **B** with **(1-B)** in final formula

The generic formula for this, along with each score line's combinations is summed in Excel.

Over Ten Games

Once the score line hits 5-5, a player must win by 2 games. This translates to winning 2 games in a row, or *avoiding losing* 2 games in a row until 2 games in a row are won.

We can find P(5-5) using same method as before:

- **P(5-5) = (P(5 Odd, 0 Even) + P(4 Odd, 1 Even) … + P(0 Odd, 5 Even)**

Then we can find the formula for P(Win 2 in a Row) recursively. For the first server we get:

- **W_1 = P(Win 2 in a Row) = SB + S(1-B) W_1 + (1-S)B W_1**

  → W_1 = (SB)/(1 – S – B + 2SB)

  → W_2 = ((1-S)(1-B))/(1 – (1-S) – (1-B) + 2(1-S)(1-B)) = (1-S-B+SB)/(1-S-B+2SB)

Combining this information, we find the formula:

- **P(Win in over 10) = P(5-5)*P(Win | 5-5) = P(5-5)*P(Win 2 in a Row)**

Results

By summing P(Win in 10 or Less) and P(Win in Over 10) for both first and second server in Excel, we get the following results. If a player is equally likely to win on serve as receive (S=B), then they are equally likely to win serving first or second. However, if they are more likely to win on serve than receive (S > B), then they are more likely to win if they serve first than if they serve second. This holds even if their overall likelihood to win serve or receive is lower than their opponent. Conversely, if (B > S), then a player is more likely to win if they serve second (i.e. receive first). Moreover, the first server is more likely to win whenever S > 0.5 and S+B ≥ 1. This means that whenever a first server is more likely to win than lose his own serve (which is almost always the case in real tennis matches), then he is more likely to win the set than his opponent, even if the probability of holding serve is the same for both players

(S+B=1). This means that in almost all real world cases, equally matched players are at an advantage if they serve first.

Conclusion

This analysis demonstrates that the first server advantage does, in fact, exist mathematically. As long as the probability of winning on serve is higher than 50%—which it typically is—then the first server is likely to win against an equally matched opponent. Even if the opponents are not equally matched, as long as a player is more likely to win on serve than receive, they have a better chance at overcoming the odds to win if they serve first, rather than second. So if you ever win a coin toss and you feel more confident serving than receiving, choose to serve first. Further analyses can be conducted to make this model more accurate and applicable. The model can be built out to include the first server advantage in tie-break sets or the first server advantage over the course of a match (with and without additional coin tosses). The model can be made more detailed by incorporating the means and variances of other statistics (such as break point conversions, winning percentage on first serve), in order to garner more accurate probabilities of winning games. Finally, to move the model beyond the theoretical, real world data could be brought in to quantify intangible variables, like the psychological effects of "momentum." However, this last extension would be difficult, due to limited statistics on the first server in any given set.