

Group 1: Predicting the Outcomes of Tennis Tournaments

A Monte Carlo Approach

Olivia Beck, Emma Lewis, Ryan Volkert

December 12, 2019

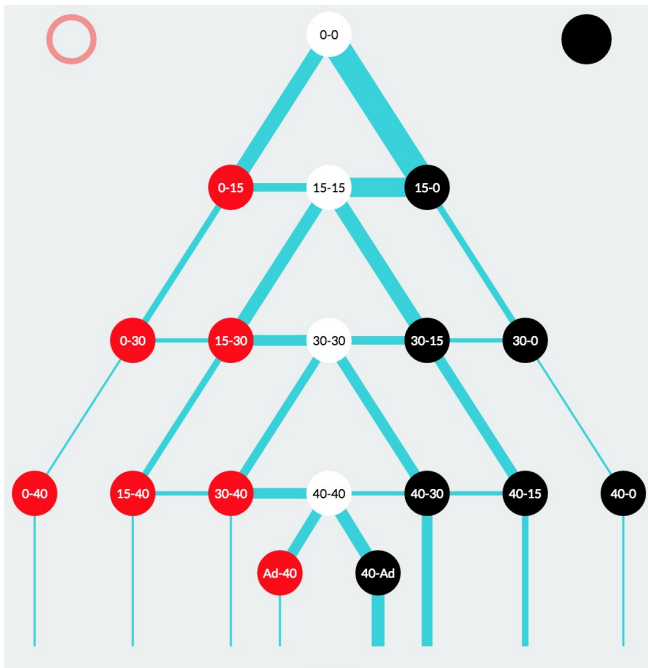
Project

- ▶ Project based off a paper by Paul K. Newton and Joseph B. Keller called Probability of Winning at Tennis 1. Theory and Data
- ▶ Paper discussed how the probability of winning a game, a set, and a match in tennis are computed based on each player's probability of winning a point on a serve
- ▶ Based on other papers the conclusion was made to treat serves in tennis as independent identically distributed (iid) random variables
- ▶ It is noted that points in tennis are not iid for most purposes
 - ▶ This is not a bad assumption as the divergence from iid is small
- ▶ Both two out of three and three out of five set matches were considered allowing for 13-point tiebreaker in each set if needed

The Rules of Tennis

- ▶ Point: Smallest unit of measurement (Love-15-30-40-game)
- ▶ Game: A game is one when a player reaches 4 points with at least a 2 point advantage
- ▶ Set: A set consists of 6 games and is won by the player who reaches 6 games first
- ▶ Advantage Set: If a game score of 6-6 is reached and advantage set rules are used, a player can only win a set with a 2 game lead
- ▶ Matches: Best of 3 set (for women) or 5 sets (for men)
- ▶ Tie-break game: If a game score of 6-6 is reached and tie-break set rules are used. In a tie-break game, a player/team must reach 7 points with a two point advantage to win

The Rules of Tennis



Values

Things we need to know:

- ▶ Empirical probability of winning a rally on serve
- ▶ Empirical probability of winning a game on serve

Values

Things we need to know:

- ▶ Empirical probability of winning a rally on serve
- ▶ Empirical probability of winning a game on serve

Things we need to calculate:

- ▶ Theoretical probability of winning a game on serve
- ▶ Theoretical probability of winning a set
- ▶ Theoretical probability of winning the tie-breaker
- ▶ Theoretical probability of winning a match

Theoretical Probabilities: Rally and Game

$$P(\text{A Winning a Rally} \mid \text{A Served}) = p_A^R = \frac{\text{Points Won On Serve}}{\text{Points Served}}$$

Theoretical Probabilities: Rally and Game

$$P(\text{A Winning a Rally} \mid \text{A Served}) = p_A^R = \frac{\text{Points Won On Serve}}{\text{Points Served}}$$

$$\begin{aligned} & P(\text{A Winning a Game} \mid \text{A Served}) \\ &= p_A^G = (p_A^R)^4 [1 + 4q_A^R + 10(q_A^R)^2] + 20(p_A^R q_A^R)^3 (p_A^R)^2 [1 - 2p_A^R q_A^R]^{-1} \end{aligned}$$

► Where Probability A loses a rally given A served:

$$q_A^R = 1 - \frac{\text{Points Won On Serve}}{\text{Points Served}}$$

Theoretical Probabilities: Set

P(A Winnning a Set | A Served)

$$= p_A^S = \sum_{j=0}^4 p_A^S(6, j) + p_A^S(7, 5) + p_A^S(6, 6) p_A^T$$

Theoretical Probabilities: Set

$$\begin{aligned} & P(\text{A Winning a Set} \mid \text{A Served}) \\ &= p_A^S = \sum_{j=0}^4 p_A^S(6, j) + p_A^S(7, 5) + p_A^S(6, 6) p_A^T \end{aligned}$$

Where, $p_A^S(i, j)$ is defined recursively as:

$p_A^S(0, 0) = 1$, $p_A^S(i, j) = 0$ if $i < 0$ or $j < 0$.

- ▶ if $i+j-1$ is even: $p_A^S(i, j) = p_A^S(i-1, j) p_A^G + p_A^S(i, j-1) q_A^G$
 - ▶ omit $i-1$ term if $j=6$ and $i < 6$
 - ▶ omit $j-1$ term if $i=6$ and $j < 6$
- ▶ if $i+j-1$ is odd: $p_A^S(i, j) = p_A^S(i-1, j) q_B^G + p_A^S(i, j-1) p_B^G$
 - ▶ omit $i-1$ term if $j=6$ and $i < 6$
 - ▶ omit $j-1$ term if $i=6$ and $j < 6$

Theoretical Probabilities: Tie Breaker

P(A Winning the Tie Breaker | A Served Initially)

$$= p_A^T = \sum_{j=0}^5 p_A^T(7, j) + p_A^T(6, 6) \sum_{n=1}^{\infty} p_A^T(n+2, n)$$

Theoretical Probabilities: Tie Breaker

P(A Winning the Tie Breaker | A Served Initially)

$$= p_A^T = \sum_{j=0}^5 p_A^T(7, j) + p_A^T(6, 6) \sum_{n=1}^{\infty} p_A^T(n+2, n)$$

Where, $p_A^T(i, j)$ is defined recursively as:

$p_A^T(0, 0) = 1$, $p_A^T(i, j) = 0$ if $i < 0$ or $j < 0$.

- ▶ if $i + j - 1 \bmod 4 \equiv 0$ or 3 :

$$p_A^S(i, j) = p_A^T(i-1, j)p_A^R + p_A^T(i, j-1)q_A^R$$

- ▶ omit $i-1$ term if $j=7$ and $i < 7$

- ▶ omit $j-1$ term if $i=7$ and $j < 7$

- ▶ if $i + j - 1 \bmod 4 \equiv 1$ or 2 :

$$p_A^T(i, j) = p_A^T(i-1, j)q_B^R + p_A^T(i, j-1)p_B^R$$

- ▶ omit $i-1$ term if $j=7$ and $i < 7$

Theoretical Probabilities: Match

Women's (Best of 3):

P(A Winning the Match | A Served First)

$$= p_A^M = (p_A^S)^2 + 2(p_A^S)^2 p_B^S$$

Theoretical Probabilities: Match

Women's (Best of 3):

P(A Winning the Match | A Served First)

$$= p_A^M = (p_A^S)^2 + 2(p_A^S)^2 p_B^S$$

Men's (Best of 5):

P(A Winning the Match | A Served First)

$$= p_A^M = (p_A^S)^3 + 3(p_A^S)^3 p_B^S + 6(p_A^S)^3 (p_B^S)^2$$

Where:

- ▶ p_A^S = Probability A wins a set given A served
- ▶ p_B^S = Probability B wins a set given B served

Data being used

Data used in paper:

- ▶ Men and Women semifinalists in the 2002 U.S Open tournament
- ▶ Men and Women semifinalists in the 2002 Wimbledon tournament

Data being used

Data used in paper:

- ▶ Men and Women semifinalists in the 2002 U.S Open tournament
- ▶ Men and Women semifinalists in the 2002 Wimbledon tournament

Data used in our simulation:

- ▶ Men and Women semifinalists in the 2019 Wimbledon tournament

Results of Game

Table 1: Data for the Womens Semifinalists in the 2002 U.S. Open Tournament

	P(Win a Rally)	Empirical P(Win a Game)	Paper P(Win a Game)	Our P(Win a Game)
S. Williams	0.69	0.71	0.89	0.89
V. Williams	0.63	0.80	0.79	0.79
L. Davenport	0.65	0.85	0.83	0.83
A. Mauresmo	0.63	0.77	0.79	0.79

Results of Game

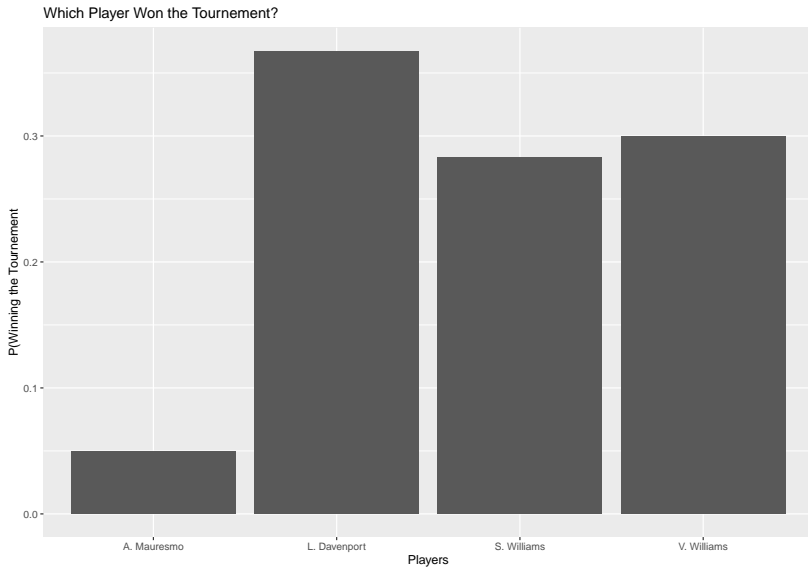
Table 1: Data for the Womens Semifinalists in the 2002 U.S. Open Tournament

	P(Win a Rally)	Empirical P(Win a Game)	Paper P(Win a Game)	Our P(Win a Game)
S. Williams	0.69	0.71	0.89	0.89
V. Williams	0.63	0.80	0.79	0.79
L. Davenport	0.65	0.85	0.83	0.83
A. Mauresmo	0.63	0.77	0.79	0.79

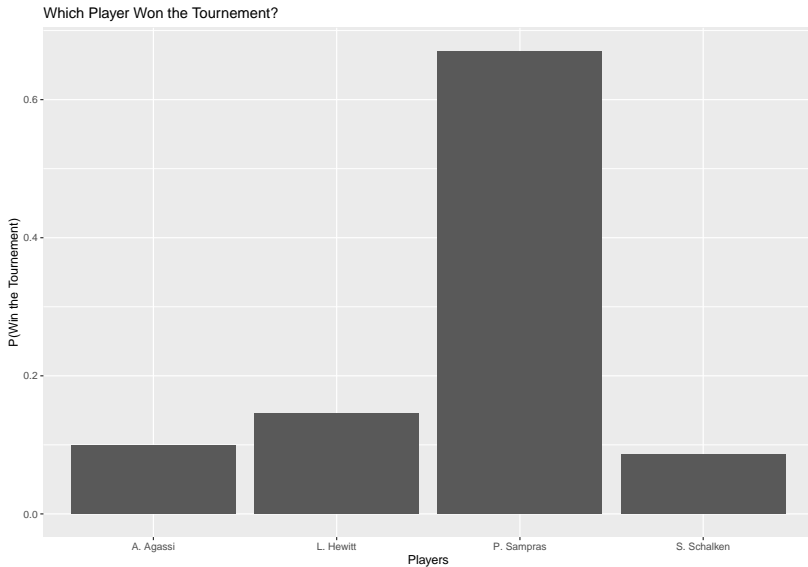
Table 2: Data for the Mens Semifinalists in the 2002 U.S. Open Tournament

	P(Win a Rally)	Empirical P(Win a Game)	Paper P(Win a Game)	Our P(Win a Game)
P. Sampras	0.73	0.95	0.93	0.93
A. Agassi	0.66	0.87	0.85	0.85
L. Hewitt	0.67	0.85	0.86	0.86
S. Schalken	0.68	0.90	0.88	0.88

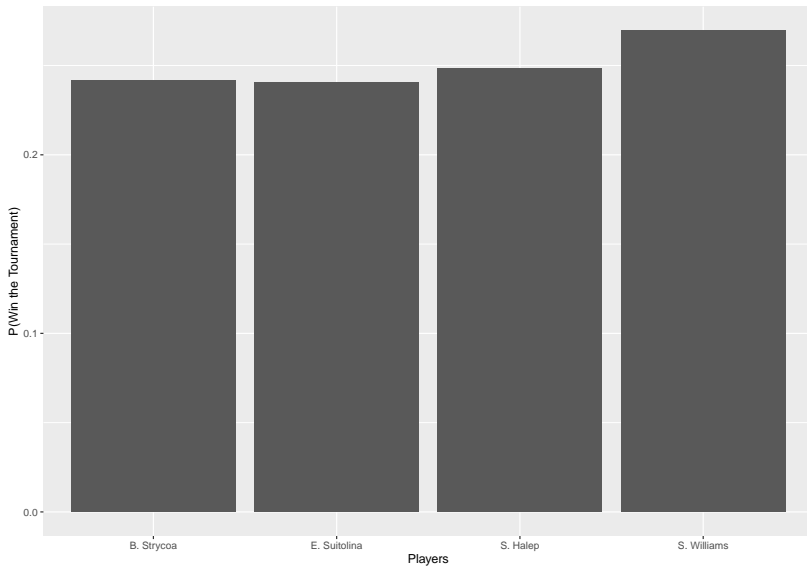
Tournament Results: Women



Tournament Results: Men



2019 Wimbledon: Women



```
## $title
```

```
## [1] "Which Player Won the Tournament?"
```

2019 Wimbledon: Men

```
## Parsed with column specification:  
## cols(  
##   Name = col_character(),  
##   `Points Won on Serve` = col_double(),  
##   `Total Points on Serve` = col_double()  
## )
```

