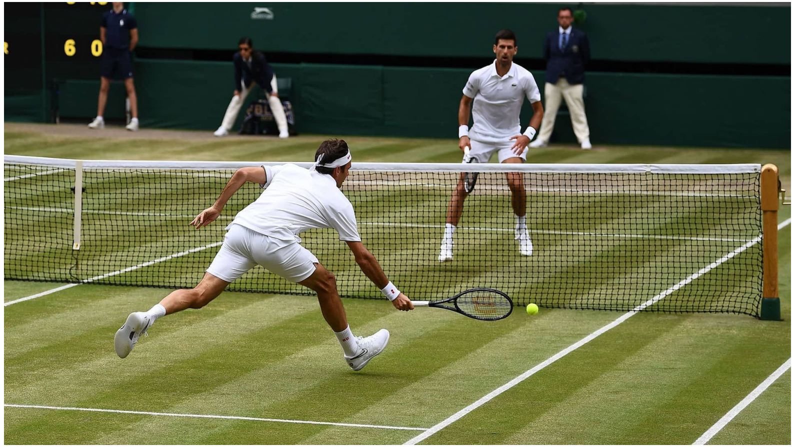
A Tennis Paradox

Justine Lai and Corrine Chen
Mentor: Jenny Jin

More Points = Victory?

Games won by Federer:
6+6+6+6+12=36

Games won by Djokovic:
7+1+7+4+13=32



Sets won by Federer: 2

Sets won by Djokovic: 3

Set	Winner	Federer games won	Djokovic games won
1	Djokovic	6	7
2	Federer	6	1
3	Djokovic	6	7
4	Federer	6	4
5	Djokovic	12	13

Concepts

Binomial Random Variables

- (def). A type of random variables, which model uncertainty
- Solves for the probability of **k successes** in **n trials**.
- Parameters: (n, p)
  - n = number of trials
  - p = probability of one successful trial
- Formula:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Markov Random Variables

- Markov Random Variables represent states at a given time

  → chains of states which depend solely on the previous state

- Transition matrices organize the probability of moving from one state to another
  - Example matrix:

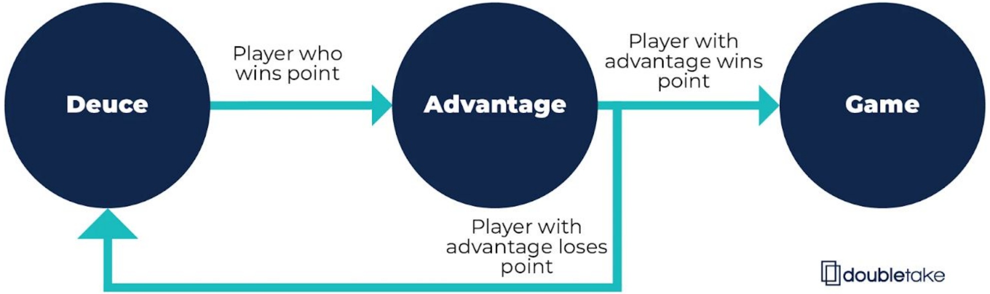
	$X_{t+1} = \text{sunny}$	$X_{t+1} = \text{cloudy}$	$X_{t+1} = \text{rainy}$
$X_t = \text{sunny}$	0.6	0.3	0.1
$X_t = \text{cloudy}$	0.3	0.3	0.4
$X_t = \text{rainy}$	0.2	0.4	0.4

The Problem

Tennis Rules

In tennis competitions, there are **points**, **games**, **sets**, and **matches**. High school tennis matches consist of:

- 2-3 sets in 1 match
  - Tiebreaker
- 6-7 games in 1 set
  - Tiebreaker
- 4+ in 1 game
  - Deuce



Our Problem

Originally, we set out to find the probability that a player **wins more points** but **loses an entire tennis match**. After some struggle... we decided to simplify the problem to determining the probability that **a player loses a set** but **won more points**.

Simplification we have made include:

- Reducing the match to only one set
- Eliminating special set outcomes such as 7-5 or 6-6.
- Assuming a constant probability that a player has to win a point

What is the probability someone loses a tennis set but wins more points in total?

Solution

Solution

First, we need to consider the individual cases of winning a single game.

Let p = probability of winning a point, q = probability of losing a point

The probability of winning with a score of 4-2 can be determined using binomial random variables:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$P(4 : 2) = \binom{5}{3} p^3 q^2 p = 10 p^4 q^2$$

- X = number of wins
- k = 3
- n = number of games = 5

Distributing 3 remaining wins across 5 remaining point opportunities

The last point must be a win

Solution (Contd.)

A special outcome of a game is the outcome of a **deuce** (40-40, or point-wise 3-3). In this case, we implement a Markov transition matrix:

	D	W	L	WW	LL
D	0	p	q	0	0
W	q	0	0	p	0
L	p	0	0	0	q
WW	0	0	0	0	0
LL	0	0	0	0	0

D: Deuce state

W: Wins one point

L: Loses one point

WW: Wins two points consecutively

LL: Loses two points consecutively

Note. P(X) = The probability of winning from state X

$$P(D) = p \cdot P(W) + q \cdot P(L)$$
$$P(W) = q \cdot P(D) + p \cdot P(WW)$$
$$P(L) = p \cdot P(D) + q \cdot P(LL)$$

......

$$\Rightarrow P(\text{Winning from Deuce}) = \frac{20p^5q^3}{1-2pq}$$

Solution (Contd.)

Using these two ideas, we were able to get expressions for the probability of winning/losing in all cases.

Note that we found the probabilities of winning. The probabilities of losing were determined by swapping p and q.

Game Outcome	Probability Equation
4-0	p^4
4-1	$4p^4q$
4-2	$10p^4q^2$
Deuce Win	$\frac{20p^5q^3}{1-2pq}$
0-4	q^4
1-4	$4q^4p$
2-4	$10q^4p^2$
Deuce Loss	$\frac{20q^5p^3}{1-2pq}$

Solution (Contd.)

Now, we consider outcomes for a set:

- 6-0
  - **not possible**
- 6-1
  - Player: $6 \cdot 4 + 1 \cdot 0 = 24$ points
  - Opponent: $6 \cdot 2 + 1 \cdot 4 = 16$ points
  - $16 < 24 \rightarrow$ **not possible**
- 6-2
  - Player: $6 \cdot 4 + 2 \cdot 0 = 24$ points
  - Opponent: $6 \cdot 2 + 2 \cdot 4 = 20$ points
  - $20 < 24 \rightarrow$ **not possible**
- 6-3
  - Player: $6 \cdot 4 + 3 \cdot 0 = 24$ points
  - Opponent: $6 \cdot 2 + 3 \cdot 4 = 24$ points
  - $24 = 24 \rightarrow$ **not possible, we are not counting tie scenarios**
- 6-4
  - Player: $6 \cdot 4 + 4 \cdot 0 = 24$ points
  - Opponent: $6 \cdot 2 + 4 \cdot 4 = 28$ points
  - $28 > 24 \rightarrow$ **possible**

In sum, it is only possible to win more points and lose a set if the set has a game score of 6-4. So, we only need to consider cases of the player winning 6 games and the opponent winning 4 games.

Due to the large amount of outcomes in the case of a 6-4 game, we ran a simulation to compute the overall probability.

Results and Conclusion

Results

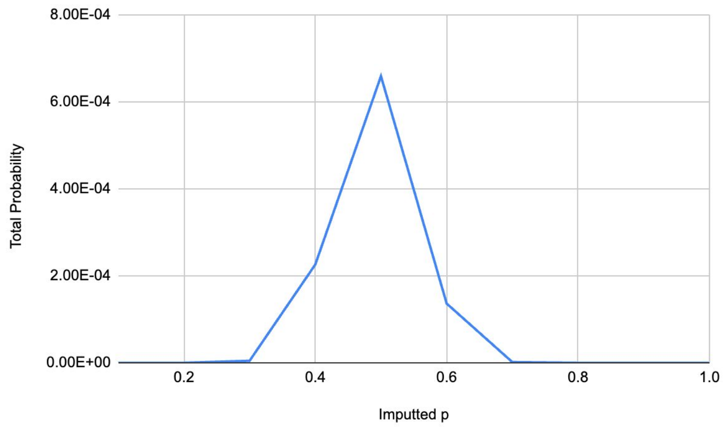
From our program, here are the probabilities depending on what p-value is chosen.

Though the chance of this is extremely small, we should take into account the probability of a 6-4 outcome by itself:

$$P(6-4) = \frac{\binom{6-1}{3}}{\binom{6-1}{0}+\binom{6-1}{1}+\binom{6-1}{2}+\binom{6-1}{3}} = 0.385$$

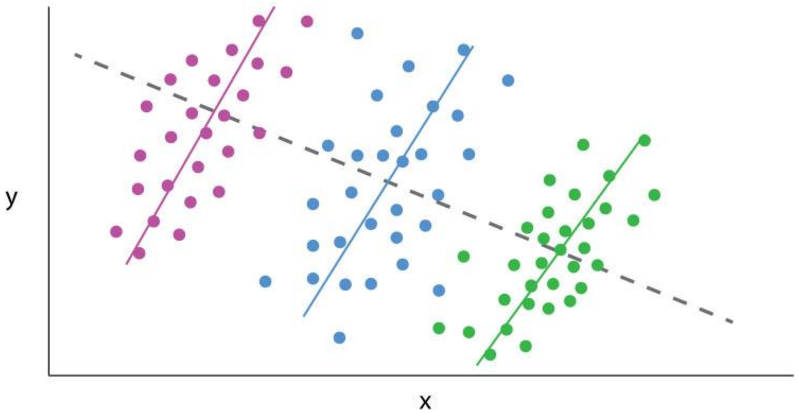
In addition, from these results, we can conclude that the phenomenon we were studying happens mostly with equally matched players.

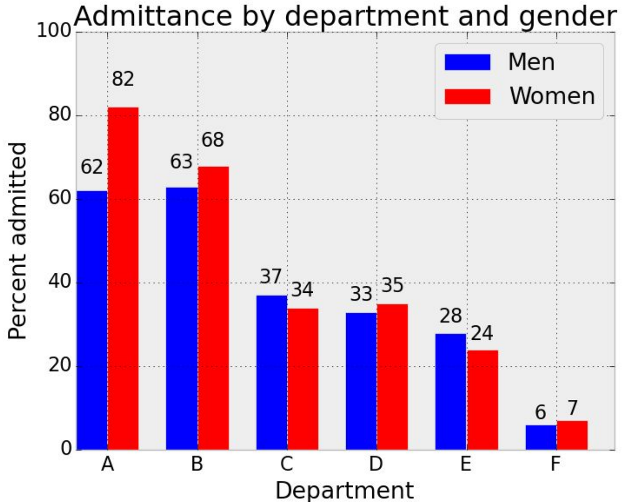
Inputted p	Total Probability
1	0.0
0.9	$2.682 \cdot 10^{-17}$
0.8	$3.505 \cdot 10^{-10}$
0.7	$1.563 \cdot 10^{-6}$
0.6	0.0001363
0.5	0.00065943
0.4	0.000226
0.3	$4.58 \cdot 10^{-6}$
0.2	$2.1142 \cdot 10^{-9}$
0.1	$4.8966 \cdot 10^{-16}$
0	0.0



Conclusion

Simpson's Paradox: a statistical phenomenon where an association between two variables is different between subpopulations and a whole population





References

IMG Video Archive. (2019). IMG Video Archive. https://imgvideoarchive.com/client/the_wimbledon_archive/results/gentlemens_singles?season=2019

Ross, S. (2014). *A First Course in Probability* (10th ed.). Pearson.

Sprenger, J., & Weinberger, N. (2021, March 24). *Simpson's Paradox*. Stanford.edu. https://plato.stanford.edu/archives/sum2021/entries/paradox-simpson/

Thank you!