Will Andy Murray win Wimbledon 2015?

Robert Johnson, Sporting Advantage Limited

PyDataLondon, 7th July 2015
Wimbledon Questions

- Murray won in 2013. Will he win again this year?
- Is Murray on better form this year than in 2013?
- How do the top players compare in the mens’ tournament?
Outline

1. A simple model for tennis
   - Data
   - Mathematical derivations
   - Implementation
Data sources.

- Tennis results are freely available: www.tennis-data.co.uk.
- I used all ATP matches from 2010 until now.
- Start estimating player abilities from January 2012.

Assume each player has an (unknown) ability parameter $\theta$.

Model $P(\text{Player } i \text{ defeats player } j)$ using a logistic function of $\theta_i - \theta_j$

$$P(\text{i defeats } j) = \frac{1}{1 + \exp(-(\theta_i - \theta_j))}$$
Likelihood function

\[ L(\theta, k = 1, \ldots, n) = \prod_{k=1}^{n} \frac{1}{1 + \exp \left( - (\theta_i - \theta_j) \right)} \]

- Player indices \( i \) and \( j \) in the equation above depend on match \( k \).
Let’s follow the method of Dixon and Coles (1997) and exponentially downweight older matches.

Gives us the pseudo-likelihood for each time point $t$:

$$L_t(\theta; \phi) = \prod_{k \in A_t} \left\{ \frac{1}{1 + \exp(- (\theta_i - \theta_j))} \right\}^{\phi(t - t_k)}$$

where $t_k$ is the time match $k$ is played, $A_t = \{k : t_k < t\}$ and $\phi$ is an exponential downweighting function.
Coded that all up in Python last Sunday
- Used **numpy**, **pandas**, **scipy**
- Numerically maximised the pseudo-likelihood using **minimize** function in **scipy.optimize**
- Fitted on a day by day rolling basis
- Plotted using **matplotlib**
Care needs to be taken to use the numpy vectorised operations for fitting the model

- \(\approx 4500\text{ms} \) to evaluate a single likelihood using simple loops
- Iterating over rows of a pandas DataFrame is a bad idea...
- \(\approx 2\text{ms} \) to calculate using numpy vectorised operations
Estimated tennis ability over time (2012-present)

- Federer
- Djokovic
- Murray
- Wawrinka
There appears to be value going long Federer, short Murray.

Screenshot taken from Betfair yesterday evening (6th July).
Wimbledon this year...

- Don’t bet on this!
- Betting markets are extremely efficient
  - This discrepancy is due to factors not included in my model
- However model suggests Murray and Federer have $\approx$ equal probabilities of winning Wimbledon
- Djokovic is the clear favourite to win
Many interesting applications of Data Science to the world of sport

We provide consultancy on sports modelling: see www.sporting-advantage.co.uk
M. Dixon and S. Coles. 
*Modelling Association Football Scores and Inefficiencies in the Football Betting Market.*

I. McHale and A. Morton. 
A Bradley-Terry type model for forecasting tennis match results.