Fifteen-forty

Describe the real-world problem

Results: The Championships, Wimbledon 2015

Ladies' singles: *Serena Williams* def *Garbine Muguruza*: 6–4, 6–4

Gentlemen's singles: *Novak Djokovic* def *Roger Federer*: 7–6 (7–1), 6–7 (10–12), 6–4, 6–3

Some years ago a tennis commentator made the following remark during a Grand Slam tournament: 'A top male player has a fifty-fifty chance of winning a game from 15–40 on serve.'

Specify the mathematical problem

Evaluate this statement. That is: what probability can we assign to the outcome of winning a service game from a score of 15–40?

Formulate the mathematical model

Data

Students can be given the data table below or tasked with collecting the required data. The official Wimbledon website contains many years' worth of results, including the most up-to-date live scores. http://www.wimbledon.com/en_GB/scores.

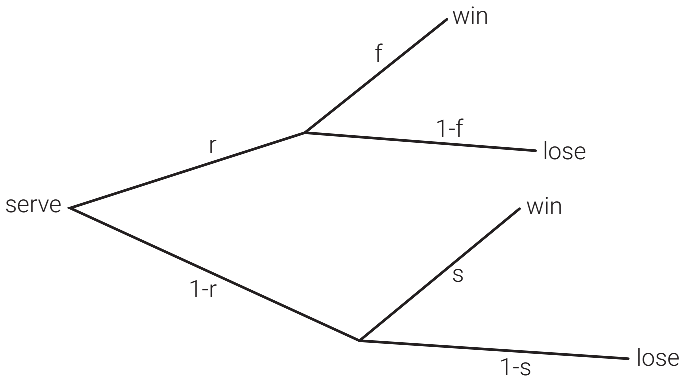
Other Grand Slam tournament websites also contain similar data.

Finals match data					
Wimbledon 2015	*Williams*	*Muguruza*		*Djokovic*	*Federer*
1st serve in (r)	37 of 68 54%	43 of 61 70%		95 of 145 66%	94 of 141 67%
Winning 1st serve (f)	29 of 37 78%	23 of 43 53%		70 of 95 74%	70 of 94 74%
Winning 2nd serve (s)	11 of 31 35%	6 of 18 33%		30 of 50 60%	23 of 47 49%

Situational assumptions

Given the long run frequency of serving outcomes we make the following assumptions:

- The probability of the outcome of any particular serve is estimated as the long run proportion for the outcome of serves as given in the table.
- The outcome of any point is independent of the result of the previous point. (Note that the reliability of these assumptions should increase with the quality of the players.)
- There are two mutually exclusive outcomes for each point (win or lose).



Outcomes of a service point

To win a point:
either the first serve is in (with probability given by r) and the service point is won (with probability f)
or the first serve is out (with probability 1 − r) and the second serve is won (with probability s).

Hence, as seen in the figure using the notation from the table, the probability of winning a service point (p) is given by
$p = rf + (1 - r)s$

Solve the mathematics

Using the data in the table, we obtain:
p (*Djokovic*) = $(0.66)(0.74) + (0.34)(0.60) = 0.69$
p (*Federer*) = $(0.67)(0.74) + (0.33)(0.49) = 0.66$

Note that $q = 1 - p$ is the probability of losing a service point, and that the probability of winning a point as receiver is the complement of the opponent's probability of winning as server.

Let Pr(G) be the probability that a player wins a service game from 15–40.

The server must win the next two points and then win from deuce.

So our required probability is $\Pr(G) = p^2 \times \Pr(D)$, where Pr(D) is the probability of winning a service game from deuce.

To win from deuce the server must win the next two points (probability p^2)
or return to deuce and win from deuce.

To return to deuce the server must win the first point and lose the second (pq) or vice-versa (qp).

Hence $\Pr(D) = p^2 + 2pq \times \Pr(D)$

Hence $\Pr(D) = p^2/(1 - 2pq) = p^2/(1 - 2p + 2\,p^2)$ since $q = 1 - p$.

So $\Pr(G) = p^4/(1 - 2p + 2p^2)$

Check:
If $p = 0$, $\Pr(G) = 0$ and if $p = 1$, $\Pr(G) = 1$
as should occur.

Interpret the solution

If we take the case of *Federer* then $p = 0.66$

Substituting this value for p gives $\Pr(G) = 0.34$
That is, a 34 per cent chance of winning the game.

Verify that for *Djokovic* the figure is around 40 per cent.

We can find the value of 'p' that will meet the commentator's criterion as follows.

We need to solve $p^4/(1 - 2p + 2p^2) = 0.5$
That is,
$2p^4 - 2p^2 + 2p - 1 = 0$ which we can solve with the help of Computer Algebra System (CAS) software.
$p \approx 0.75$ is the single positive root

Check:
When $p = 0.75$, $\Pr(G) = 0.506$ (approximately 50 per cent chance of winning).

It seems the commentator was a bit over optimistic!

But, Wimbledon finals is the very best playing against the very best. If the opponent is a lesser player (which happens most of the time) then the opinion might be closer to the mark.

The Wimbledon website contains complete statistics for all seven matches played by finalists. To obtain them click on the names of players.

The collated data for *Djokovic* gives outcomes for 728 service points across seven opponents.

Using the total data for the same variables shown in the table we calculate:
$r = 0.707$ (0.71)
$f = 0.77$
$s = 0.636$ (0.64)
giving $p = 0.731$ (0.73)

With $p = 0.73$ we obtain $\Pr(G) = 0.469$ (around 47 per cent chance of winning).

So the estimate is looking quite a reasonable one, especially when we consider that many opponents on the circuit will be lesser players than those accepted for a Grand Slam tournament.

Evaluate the model

We can also take a statistical look at outcomes.

Assume that the choice of the seven opponents in the draw is sufficiently random to support the estimation of confidence intervals. (The sample is probably best regarded as representative rather than random.)

Consider the 728 service points from Wimbledon 2015 as a sample from the population of many thousands that a quality player, such as a Grand Slam winner, is involved with while at the top of his or her form over a period of years.

Considering the probability of winning a point as given by the proportion of service success, then the standard error of the sampling distribution of proportions is given by $(\sqrt{(p^*(1-p^*)/n)})$ where p^* is the value (0.73) from the sample ($n = 728$).

Then with (approximately) 95 per cent confidence we estimate the population value of p to lie in the interval:
$p^* - 2(\sqrt{(P^*(1 - p^*)/n)} \leq p \leq p^* + (2\sqrt{(P^*(1 - p^*)/n)}$

That is $0.698 \leq p \leq 0.764$ which translates to
0.409 (41%) ≤ Pr(G) ≤ 0. 531 (53%)

Given that some opponents will not be of Grand Slam quality, we might argue that this outcome is probably still conservative at the upper end.

It is interesting that the original Wimbledon data generated a p-value (0.69) that lay just outside the lower 95 per cent confidence limit. It is probably no surprise that such an extreme occurrence took place with Roger Federer as the opponent.

The first evaluation suggested further avenues to explore, which were not envisaged at the start. This is an illustration of how interim results can often stimulate further modelling activity in which aspects of the modelling cycle are re-activated, sometimes a number of times.

Report the solution

This analysis has focused around Novak Djokovic as a case study. There is opportunity to test the commentator's claim using a selection of top players, and ideally this should be pursued.

For example, what are the corresponding outcomes for top women players? A similar approach can be used to investigate parallel outcomes for women players, using data from the website.

Perhaps Serena Williams is just about the best of all time? Use available data to compare her serving performance with that of other top women players.

Further problems could also be explored. For example, what are reasonable probabilities of winning a game from each of the positions from 0–40 to 40 –0?

Using the same notation, show that the probability of winning a game from 40–0 can be expressed as:
$Pr(G) = p(1+ q + q^2) + q^3\, Pr(D)$
$Pr(G) = 0.998$ when $p = 0.73$
and so on.

ACER