

Tennis Challenge Solution

David Starkey

August 10, 2018

1 Statistics of Tennis

In the following short study, I investigate the probability of various outcomes of a set of tennis. The set is won when either of two players have reached 6 games AND are ahead by two clear games. A game is won when a player wins four rallies AND are ahead by two clear points. The investigation will examine how the probability of winning a set of tennis (the set-win probability) is affected by a players point-win probability (the probability of the player winning a single rally). The statistical framework for the study is described in the following sections and the accompanying code will allow the interested reader to duplicate the figures and analysis below for any desired point-win probability. I aim to:

- Examine landscape of the 2D posterior probability distribution as a function of all possible scores.
- Examine how the set-win probability evolves with point-win probability (in other words how much more likely are you to win a set given a particular probability of scoring a rally).

I hope you have fun!

2 Probabilty of Winning Game

This is the sum of two possible outcomes.

$$P(wg) = P(wbd) + P(g2d) \times P(wod) \quad (1)$$

where the probabilities of winning before deuce, of the game reaching deuce, and of winning a deuce are given by $P(wbd)$, $P(g2d)$ and $P(wod)$ respectively.

2.1 Probability of Reaching Deuce: P(g2d)

A game will reach deuce if C(W W W L L L), where C denotes the events can happen in any order, and yet the same event is described by reversing the positions of two W's. We

therefore need to avoid double counting of events using the factorials below. This has the probability given by

$$P(\text{g2d}) = \frac{6!}{3!3!}p^3(1-p)^3. \quad (2)$$

2.2 Probability of Winning Before Deuce: $P(\text{wbd})$

This will occur in any of the following circumstances.

$$\begin{aligned} C(\text{WWWW}) &= p^4, \\ C(\text{WWWL})W &= \frac{4!}{3!1!}p^3(1-p)p, \\ C(\text{WWWLL})W &= \frac{5!}{3!2!}p^3(1-p)^2p. \end{aligned} \quad (3)$$

The probability of winning before deuce is then the sum of these

$$P(\text{wbd}) = p^4 \left(1 + \frac{4!}{3!1!}(1-p) + \frac{5!}{3!2!}(1-p)^2 \right). \quad (4)$$

2.3 Probability of Winning Deuce: $P(\text{wod})$

To proceed further we need the probability of returning to deuce $P(\text{b2d})$.

$$P(\text{b2d}) = C(WL) = 2p(1-p) \quad (5)$$

The game can now potentially last forever, continuously returning to deuce with diminishing probability. The probability of winning after n deuces is.

$$P(\text{wnd}) = P(\text{b2d})^{n-1}p^2. \quad (6)$$

Since we can win on any of the n deuces up to infinity we must evaluate the sum

$$P(\text{wod}) = \sum_{n=1}^{\infty} P(\text{b2d})^{n-1}p^2. \quad (7)$$

For a geometric series of the form ar^{n-1} , the sum to infinity is $a_1/(1-r)$, with first term and common ratio a_1 and r . Combining this with Equation 7, we get the probability of winning a deuce (or of winning game by two clear points),

$$P(\text{wod}) = \frac{p^2}{1 - 2p(1-p)}. \quad (8)$$

Combining Equations 1, 8 and 2 gives us the explicit probability of winning a game.

$$P(\text{wg}) = p^4 \left(1 + \frac{4!}{3!1!}(1-p) + \frac{5!}{3!2!}(1-p)^2 \right) + \frac{6!}{3!3!}p^3(1-p)^3 \frac{p^2}{1 - 2p(1-p)}. \quad (9)$$

This gives us the probability of winning a game. We now do a similar procedure, taking $P(\text{wg})$ forward to calculate the set win probability.

3 Probability of Winning Set: $P_s(X, X)$

To win a set, we need to get to 6 points and be in the lead by 2 clear points. If both players reach 6, a tie break game determines the final set and winner. Using the same reasoning as Equation 3 I calculate the possible winning scores and probabilities as

$$\begin{aligned}
P_s(6, 0) &= C(\text{WWWWWW}) = p_g^6, \\
P_s(6, 1) &= C(\text{WWWWWL})W = \frac{6!}{5!1!}p_g^5(1-p_g)p_g, \\
P_s(6, 2) &= C(\text{WWWWLW})W = \frac{7!}{5!2!}p_g^5(1-p_g)^2p_g, \\
P_s(6, 3) &= C(\text{WWWLWW})W = \frac{8!}{5!3!}p_g^5(1-p_g)^3p_g, \\
P_s(6, 4) &= C(\text{WWLWWW})W = \frac{9!}{5!4!}p_g^5(1-p_g)^4p_g, \\
P_s(7, 5) &= C(\text{WWLWWL})WW = \frac{10!}{5!5!}p_g^5(1-p_g)^5p_g^2, \\
P_s(7, 6) &= P(\text{g266})p_g,
\end{aligned} \tag{10}$$

where the point probability p has been replaced by the probability of winning a game $p_g = P(\text{wg})$ (see Equation 1). These are calculated by suitably combining probabilities and noting that the some ordering of W and L are the same event. I try to encapsulate this meaning in the $C()$ functions. The interesting case is that of the tie break $P_s(7, 6)$. To win a tie break, we first have to reach a score of 5,5. This is given by $P(\text{g255})$ where

$$P(\text{g255}) = C(\text{WWLWWL}) = \frac{10!}{5!5!}p_g^5(1-p_g)^5. \tag{11}$$

The probability of reaching a score of 6,6 is then

$$P(\text{g266}) = P(\text{g255})p_g(1-p_g) + P(\text{g255})(1-p_g)p_g = 2P(\text{g255})p_g(1-p_g), \tag{12}$$

since we can first win and then lose or vice versa. Combining Equations 12 and 10 gives us the full range of probabilities for an input point probability p and score. To consider losing the set, I simply replace p_g with $1-p_g$ in Equation 10 and consider a win W now to be a loss L .

4 Posterior Probability Landscape

The statistical analysis below can be coded into a Python script without too much hassle and allows us to investigate the probability of various game scores for the set. Figure 1 shows this landscape for a game between Player 1 and Player 2 when Player 1 has a point-win probability of 0.45 (and Player 2 has a point-win probability of 0.55), 0.50 and 0.55 respectively.

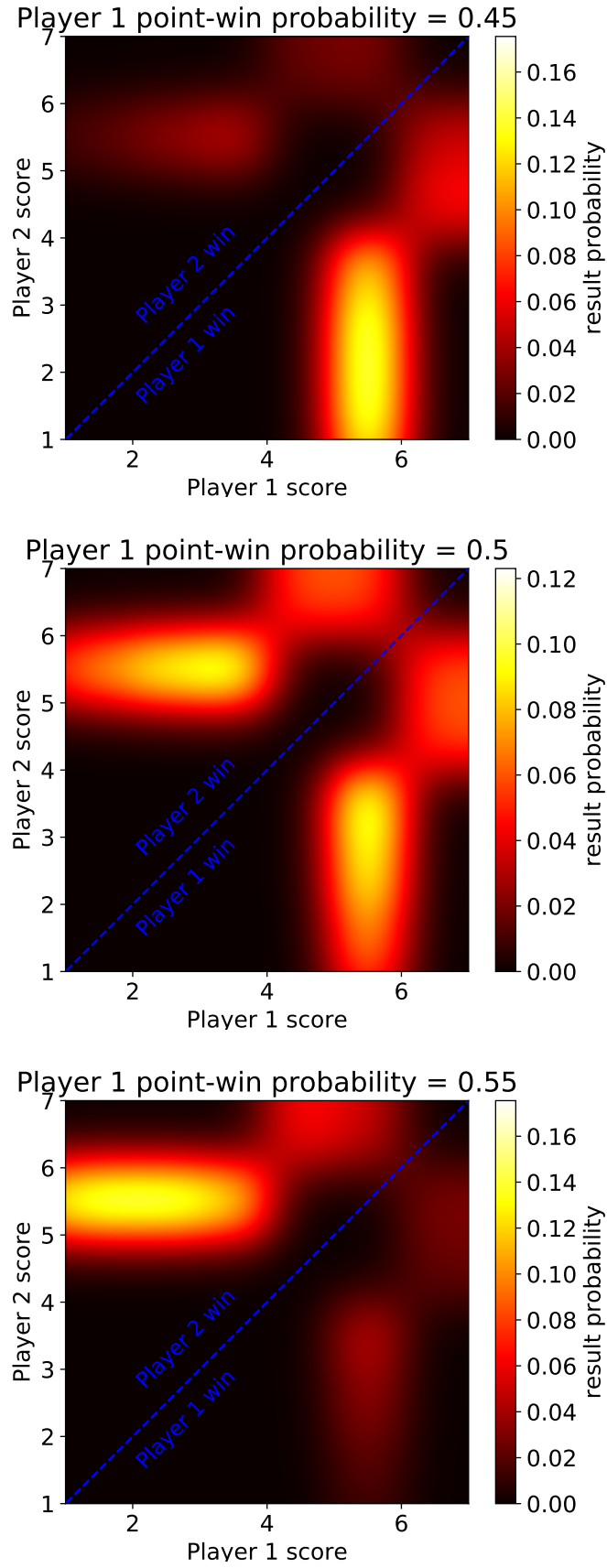


Figure 1: The 2-D posterior probability landscape as a function of the various scores. Black points indicate invalid scores where the blue line divides a Player 1 victory from a Player 2 victory. The image headers show the input point-win probability for Player 1.

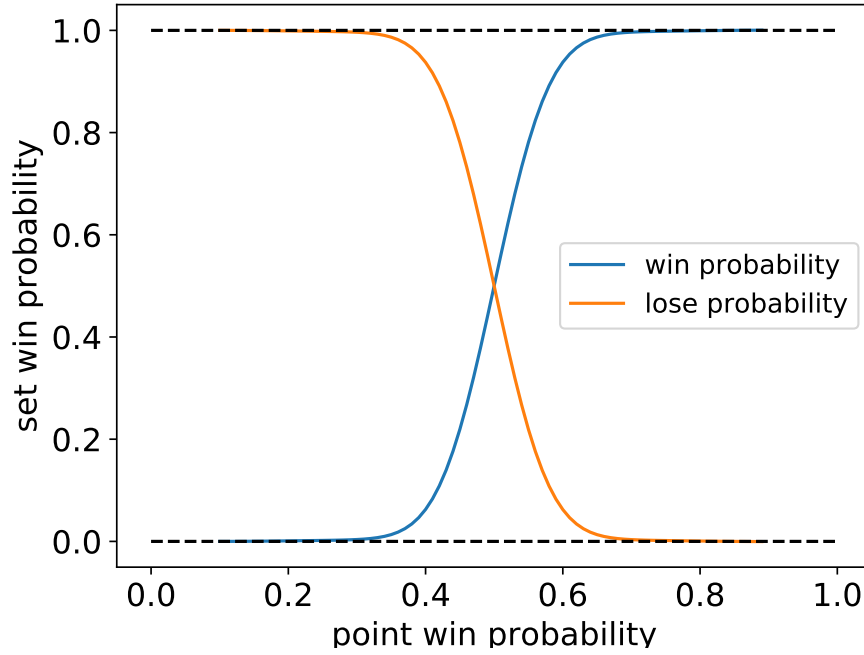


Figure 2: The set-win probability vs the point-win probability. Dashed lines indicate the lower and upper limits zero and one.

5 Conclusion

A number of interesting features can be concluded from the posterior probability data in Figures 1 and 2:

- Figure 1 shows that the posterior probability landscape is very sensitive to the point-win probability; a slight imbalance of the point-win probability is all that is needed to make a Player significantly more likely to win the set.
- The above point is illustrated more dramatically in Figure 2. This shows the probability of winning a set given the probability of winning a point. This is not linear (a 10 % increase in the probability of winning a rally does not correspond to the same increase in winning a set). The steep increase of the curve around 0.4 shows how quickly a point-win advantage conveys a much higher set-win probability.

Future considerations in this study might build on the set-win probability study and investigate the match-win probability. This is a very simple rule addition to the code where match is the best of three (or five) sets and to win by two clear sets. One could also adapt the rules to other similar court games such as squash. The statistics of games is a very interesting field of study.