

Daniel Bregman

14/07/2011

Modelling Probabilities in Tennis

Introduction:

In this investigation I shall examine the possibilities for modelling the probabilities in tennis matches of varying complexity. I will begin by assuming a very simple game of ten points, played by players of a consistent known strength. I will then expand this into games following the real-life rules of tennis more closely, to see how these affect the probabilities involved. I will examine how any results may be generalized to players of different strengths, and how these strengths affect the odds of such players over matches.

A simple model:

Let us assume two players Adam and Ben (who will be referred to as A and B), with fixed probability of scoring a given point against each other. Let A win $\frac{2}{3}$ of points, and B win $\frac{1}{3}$ of points.

We can begin by simulating a 10-point game. As the probabilities are consistent, and there are only two possibilities for each point, the game is a series of Bernoulli trials and can be modelled with a binomial distribution. Let X denote the number of points scored by A. We can now state that:

$$X \sim B(10, \frac{2}{3})$$

and therefore that:

$$P(X = x) = \binom{10}{x} \left(\frac{2}{3}\right)^x \left(\frac{1}{3}\right)^{10-x}.$$

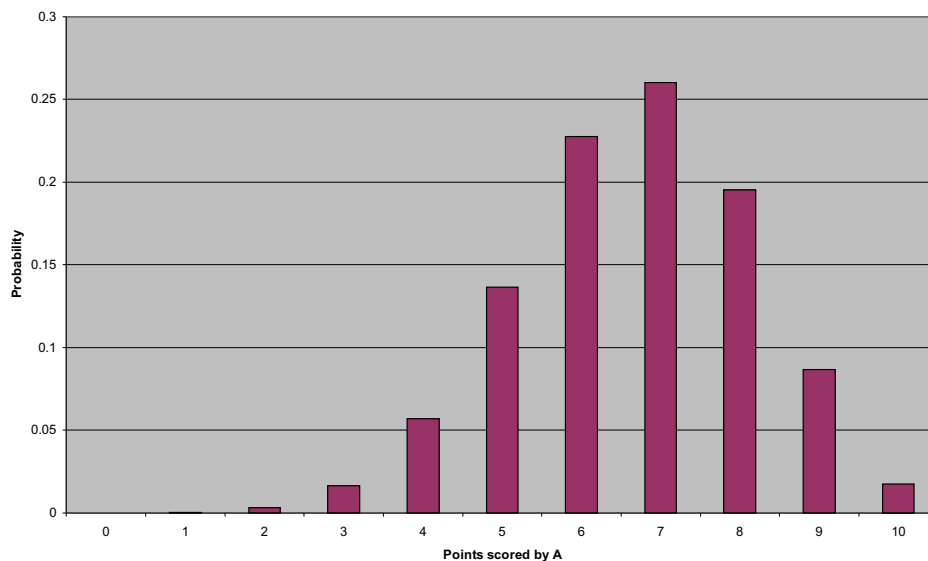
We can hence easily calculate the probability distribution for all possible values of X :

x	$P(X=x)$
0	0.000017
1	0.000339
2	0.003048
3	0.016258
4	0.056902
5	0.136565
6	0.227608
7	0.260123
8	0.195092
9	0.086708
10	0.017342

Daniel Bregman

14/07/2011

This can be presented as a histogram:



This shows that the modal score will be 7, with the highest individual probability. Based on the binomial distribution, we can also calculate the expected value and standard distribution:

$$E(X) = np = 10 \times \frac{2}{3} = \frac{20}{3} \approx 6.667$$

$$\sigma^2 = np(1-p) = \frac{20}{3} \left(1 - \frac{2}{3}\right) = \frac{20}{9}$$

$$\Rightarrow \sigma = \sqrt{\frac{20}{9}} \approx 1.49$$

Based on this we can see that most scores fall between 4 and 8, with the mean score being 6.6667.

It is important to remember the assumptions we are making here. In real play it is unlikely that a player's point probability would remain constant throughout play. We have not taken into account factors such as tiredness, confidence etc. which would seriously affect the game. With these caveats in mind, however, we can use the model to predict what might happen in an actual game.

Short game play:

Using the same players A and B from above, we may imagine a version of the tennis rules in which to win a player must have above four points (as the usual rules go) but in which, if both players have three points, deuce is called and the next point determines the winner. This scenario means that no game may go beyond 7 points.

Daniel Bregman

14/07/2011

We can model games as arrangements of 4 A's and 4 B's in 7 spaces, thus:

$$\text{Number of possible games} = \binom{7}{4} + \binom{7}{3} = 35 + 35 = 70$$

Some of these games finish earlier than 7 points, however. For instance, the arrangement 'AAAABBB' would finish at 4 points, with the score 4-0. We can therefore subtract such instances to work out the number of games of exactly y points.

$$y=4: 2 \times \binom{4}{4} = 2$$

$$y=5: 2 \times \binom{5}{4} - 2 = 8$$

$$y=6: 2 \times \binom{6}{4} - 8 - 2 = 20$$

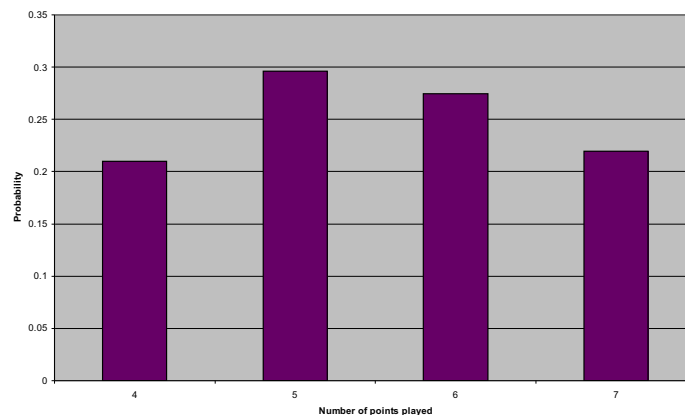
$$y=7: 2 \times \binom{7}{4} - 20 - 8 - 2 = 40$$

Note that each of these numbers represents the number of games of exactly y points, with either A or B winning. Half of these represent A winning, and half represent B winning. We can hence work out the probability of the game lasting y points, where n is the number of games of that point length:

$$P(Y = y) = \frac{n}{2} \left(\frac{2}{3} \right)^4 \left(\frac{1}{3} \right)^{y-4} + \frac{n}{2} \left(\frac{2}{3} \right)^{y-4} \left(\frac{1}{3} \right)^4$$

y	$P(Y=y)$
4	0.2099
5	0.2963
6	0.2195
7	0.2195

This can be shown as a histogram:



Daniel Bregman

14/07/2011

Based on this data we can calculate the probability that A wins the game:

$$P(A \text{ wins}) = \left(\frac{2}{3}\right)^4 + 4\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right) + 10\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^2 + 20\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^3 \approx 0.887$$

The odds that A wins are therefore:

$$\text{odds} = \frac{P(A \text{ wins})}{P(B \text{ wins})} \approx \frac{0.887}{0.113} = 4.704 : 1$$

For more general players C and D with point probabilities c and d , the probability that C wins is given by:

$$P(C \text{ wins}) = c^4 + 4c^4d + 10c^4d^2 + 20c^4d^3$$

Note that because on one point C and D's probabilities of scoring are mutually exclusive and exhaustive, we can say that:

$$\begin{aligned} c + d &= 1 \\ \Rightarrow d &= 1 - c \end{aligned}$$

In this way the probability of C winning can be written entirely in terms of C's probability.

Longer game play:

In the above examples the rules ensured that y , the number of points played, could not exceed 7. If instead the rules become that to win, a player must have 4 points and be 2 points ahead of their opponent, we can produce theoretically infinite games. The word 'deuce' will be used to refer to a state where the scores are equal and greater than 2, and the word 'advantage' will refer to the state after deuce where there is a 1 point difference.

To model this situation, we can consider two possibilities: a game with deuce and a game without deuce. We will discuss the latter first.

If the game is to occur without deuce, it must end at 4-0, 4-1, 4-2, or the reverses of these. Any other score would either be too low to win, or would have involved deuce (e.g. a score of 4-3 could only occur after a score of 4-2, which is already accounted for, or 3-3, which is deuce). The probability of each score can be calculated as follows:

$$P(4-0 \cup 0-4) = \left(\frac{2}{3}\right)^4 + \left(\frac{1}{3}\right)^4 \approx 0.299$$

$$P(4-1 \cup 1-4) = 4\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right) + 4\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^4 \approx 0.293$$

$$P(4-2 \cup 2-4) = 10\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^2 + 10\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^4 \approx 0.293$$

Daniel Bregman

14/07/2011

We can use a similar calculation to find A's chance of winning in a no n-deuce game:

$$P(A \text{ wins} \mid \neg \text{deuce}) = \left(\frac{2}{3}\right)^4 + 4\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right) + 10\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^2 \approx 0.694$$

We must also consider a game in which deuce occurs. After deuce, the next point brings the game into advantage either for A or for B. The point after that either ends the game or returns to deuce. For A to win he must score twice after deuce, for B to win he must score twice, and for the game to return to deuce each player must score once.

Two points after deuce, the probabilities are therefore as follows:

$$P(A \text{ wins} \mid \text{deuce}) = \left(\frac{2}{3}\right)^2 = \frac{4}{9}$$

$$P(B \text{ wins} \mid \text{deuce}) = \left(\frac{1}{3}\right)^2 = \frac{1}{9}$$

$$P(\text{deuce} \mid \text{deuce}) = 2 \times \left(\frac{2}{3}\right)\left(\frac{1}{3}\right) = \frac{4}{9}$$

Two points after that, the probabilities are multiplied by $\frac{4}{9}$ (because deuce must have occurred again, and so on. The probability of A winning $2z$ points after deuce is first called is given by:

$$P(A \text{ wins} \mid \text{deuce}) = \left(\frac{4}{9}\right)^z$$

The probabilities form a geometric sequence with first term $\frac{4}{9}$ and common ratio $\frac{4}{9}$. Because $\left|\frac{4}{9}\right| < 1$ we know it is a convergent sequence, and we can calculate its sum to infinity:

$$P(A \text{ wins} \mid \text{deuce}) = \sum_{z=1}^{\infty} \left(\frac{4}{9}\right)^z = \frac{\frac{4}{9}}{1 - \frac{4}{9}} = \frac{4}{5}$$

Therefore A's probability of winning once deuce is called become $\frac{4}{5}$ over a long game. To calculate A's probability of winning in a game with deuce we must calculate the probability of the original deuce occurring. The score for this must be 3-3, so the probability is:

$$P(\text{deuce}) = \binom{6}{3} \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^3 \approx 0.296$$

Daniel Bregman

14/07/2011

The probability of winning in a game with deuce is thus:

$$P(A \text{ wins} \cap \text{deuce}) = 0.295 \times \frac{4}{5} \approx 0.176$$

A's probability of winning overall is the sum of his probabilities with and without deuce:

$$P(A \text{ wins}) = 0.604 + 0.176 = 0.880$$

A and B winning are mutually exclusive and exhaustive, so they sum to 1. The odds of A winning are therefore:

$$\text{odds} = \frac{0.880}{1 - 0.880} = 5.94 : 1$$

It can be seen that A's odds of 2:1 on a single point give a much greater advantage over extended play, as he now has odds of winning of almost 6:1.

We can model these probabilities more generally again for players C and D with point probabilities c and d respectively. In a game without deuce:

$$P(C \text{ wins} \cap \neg \text{deuce}) = c^4 + 4c^4d + 10c^4d^2$$

The probability of deuce occurring to begin with is:

$$P(\text{deuce}) = \binom{6}{3} c^3 d^3 = 20 c^3 d^3$$

Once deuce has occurred, the probability of C winning is given by the sum to infinity as above (note that $|c| < 1$ must hold; the below equation fails if C is certain to score) :

$$P(C \text{ wins} \mid \text{deuce}) = \frac{c^2}{1 - 2cd}$$

The probability of C winning the game is therefore:

$$P(C \text{ wins}) = c^4 + 4c^4d + 10c^4d^2 + (20c^3d^3) \left(\frac{c^2}{1 - 2cd} \right)$$

The odds can be calculated as:

$$\text{odds} = \frac{P(C \text{ wins})}{1 - P(C \text{ wins})}$$

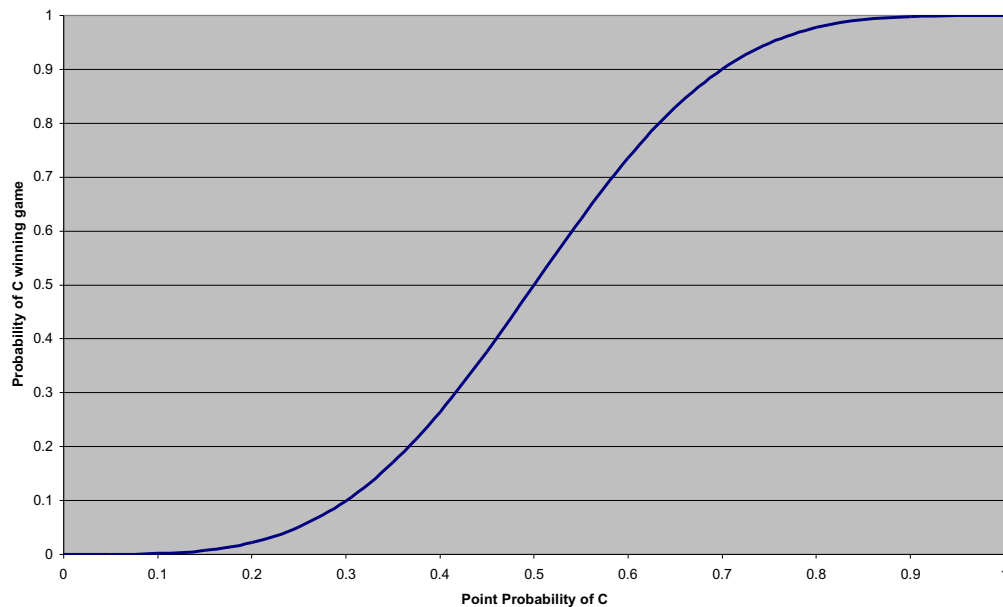
Daniel Bregman

14/07/2011

The probability distribution for C winning can be calculated for any value of c :

c	P(Cwins)	odds
0	0.000000	0:1
0.05	0.000092	1:10830
0.1	0.001447	1:689.70
0.15	0.007137	1:139.11
0.2	0.021779	1:44.916
0.25	0.050781	1:18.692
0.3	0.099211	1:9.0795
0.35	0.170355	1:4.8701
0.4	0.264271	1:2.7840
0.45	0.376852	1:1.6536
0.5	0.500000	1:1.0000
0.55	0.623149	1.6536:1
0.6	0.735729	2.7840:1
0.65	0.829645	4.8701:1
0.7	0.900989	9.0795:1
0.75	0.949219	18.692:1
0.8	0.978221	44.916:1
0.85	0.992863	139.11:1
0.9	0.998552	689.70:1
0.95	0.999908	10830:1
1.0	1.000000	1:0

The relationship can be shown graphically:



Daniel Bregman

14/07/2011

This shows that when the point probabilities are close to each other (i.e. near 0.5) a small change can make a large difference in the overall game probabilities, whereas if the point probabilities are far apart a change will not make much difference.

In practical terms, this means (perhaps unsurprisingly) that matches with closely-matched players are the most exciting to watch, because a small change in the player's relative performance can change the likelihoods dramatically, making the game less predictable.

Conclusions:

These results must be interpreted based on the limitations identified at the start. The results hold for cases with perfectly consistent players, but this is not the case in real life. The changes in performance that would result from tiredness as the game goes on, or from confidence (or the lack thereof) as one player succeeds and the other does not, would significantly change the probabilities in ways that this model cannot predict. Because of this the model would need to be made much more complex if it were to serve as a predictor of real-life situations. However, as a tool to explore the mechanisms at play in a game of tennis, the simple model that we have developed shows clearly how the players' relative strengths affect the overall result, and also how the rules of play that allow a game to go on longer significantly enhance a strong player's advantage, so the model is not entirely useless.