Analyzing Wimbledon: The Power of Statistics
by Franc Klaassen and Jan R. Magnus
Oxford University Press, 2014

last updated: December 20, 2013

Mathematical derivations

Almost all mathematical results in *Analyzing Wimbledon* are derived in the book; almost all, but not all. There are five formulas, listed in Appendix C of the book, that are not fully derived. The derivations of these five formulas are provided in the current document.

Chapter 2

(a) From point to game

Let p denote the probability that the server wins the point. We derive the probability g that the server wins the game, given that winning a point on service is an iid process. For convenience, we count points as $0, 1, 2, \dots$ rather than as $0, 15, 30, \dots$.

A game can be won in two ways: via deuce (40-40) or not via deuce. If a game does not reach deuce, then the server can win the game in precisely three ways, with probabilities

$$\begin{aligned}
\Pr(4,0) &= p \Pr(3,0) = p\binom{3}{3}p^3 = p^4, \\
\Pr(4,1) &= p \Pr(3,1) = p\binom{4}{3}p^3(1-p) = 4p^4(1-p), \\
\Pr(4,2) &= p \Pr(3,2) = p\binom{5}{3}p^3(1-p)^2 = 10p^4(1-p)^2,
\end{aligned}$$

where $\binom{h}{k}$ denotes the binomial coefficient

$$\binom{h}{k} = \frac{h!}{k!(h-k)!}$$

for any two integers $0 \leq k \leq h$, and $h!$ denotes the factorial of h:

$$h! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot h.$$

In the above three probabilities the binomial coefficient gives the number of paths by which the server can reach score 3; that is, 40 in tennis terminology. For example, there are $\binom{4}{3} = 4$ ways to reach 40-15. The probability of winning the game without reaching deuce is then

$$g_{\text{nodeuce}} = \Pr(4,0) + \Pr(4,1) + \Pr(4,2) = p^4 \left(10p^2 - 24p + 15\right).$$

Next we compute the probability of winning the game via deuce. The probability of deuce is

$$\Pr(3,3) = \binom{6}{3} p^3(1-p)^3 = 20p^3(1-p)^3.$$

Let c_p denote the probability that the game continues to another deuce, given that the current score is deuce and given the probability p that the server wins a point on service. Then $c_p = 2p(1-p)$, and hence

$$\begin{aligned}
\Pr(5,3) &= p^2 \Pr(3,3), \\
\Pr(6,4) &= p^2 \Pr(4,4) = p^2 c_p \Pr(3,3), \\
\Pr(7,5) &= p^2 \Pr(5,5) = p^2 c_p^2 \Pr(3,3),
\end{aligned}$$

and so on. Hence the probability of winning the game via deuce is

$$\begin{aligned}
g_{\text{deuce}} &= \Pr(5,3) + \Pr(6,4) + \Pr(7,5) + \cdots = p^2 \Pr(3,3)(1 + c_p + c_p^2 + \ldots) \\
&= \frac{p^2 \Pr(3,3)}{1 - c_p} = \frac{20p^5(1-p)^3}{p^2 + (1-p)^2}.
\end{aligned}$$

The probability g of winning a game is therefore as presented on page 15:

$$g = g_{\text{nodeuce}} + g_{\text{deuce}} = \frac{p^4(-8p^3 + 28p^2 - 34p + 15)}{p^2 + (1-p)^2}.$$

(b) Long matches: Isner-Mahut 2010

The probabilities of extremely long matches in Table 2.2 are based on the assumption that points are independent and identically distributed. We assume, in addition, that both players $\mathcal{I}$ and $\mathcal{J}$ have the same probability g to win their service game. Hence they are equally strong. This is the situation where such long matches can occur and hence the case of interest to us.

We observe that for a long match to occur two 'knots' must be passed. The players must reach 2-2 in sets and in the final set they must reach 5-5 in games. There is no other way that a long match can develop. As a result, we can do the calculations in three steps. First, the probability to reach 2-2 in sets is equal to 3/8, as explained on page 26. Step two concerns the calculation of the probability $\ell(5,5)$ of reaching the score 5-5 in a set. In step three we use $\ell(5,5)$ to compute the probability $\ell(a,b)$ that the set is decided with a score of a-b, where both $a \geq 5$ and $b \geq 5$ (a games for $\mathcal{I}$ and b games for $\mathcal{J}$). Obviously, winning the set requires a two-game difference; there is no tiebreak. We now explain steps two and three in some detail and derive the formula underlying Table 2.2.

Step two relies on the binomial probability distribution. The probability that one of the players, say player $\mathcal{I}$, wins his or her own service game precisely k times in 5 service games is

$$L_k = \binom{5}{k} g^k (1-g)^{5-k} \qquad (0 \leq k \leq 5).$$

Then the probability that $\mathcal{J}$ wins k times in 5 of his or her own service games is also L_k, and so is the probability that $\mathcal{I}$ breaks $5-k$ times in 5 of $\mathcal{J}$'s service games. The probability that $\mathcal{I}$ wins 5 of 10 games by winning k service games and breaking $5-k$ times is then L_k^2, and hence the probability of 5-5 is

$$\ell(5,5) = \sum_{k=0}^{5} L_k^2.$$

In step three we first compute the probability that — given a 5-5 score in the set — player $\mathcal{I}$ wins the set with a score of $(7+h)$-$(5+h)$, where $h = 0, 1, 2, \ldots$. Let c_g denote the probability that the set continues from one equal game score to the next, such as from 5-5 to 6-6. Then $c_g = g^2 + (1-g)^2$, and hence player $\mathcal{I}$ wins the next two games with probability $(1-c_g)/2$. The probability that $\mathcal{I}$ wins by $(7+h)$-$(5+h)$ given 5-5 becomes

$$\Pr(7+h, 5+h \mid 5\text{-}5) = c_g^h (1-c_g)/2.$$

The probability that $\mathcal{J}$ wins by the same score is equal to that. Therefore, the probability $\ell(a,b)$ that the set is decided with a score of a-b becomes

$$\ell(a,b) = \ell(5,5) \Pr(7+h, 5+h \mid 5\text{-}5), \qquad h = (a+b-12)/2.$$

Table 2.2 concerns probabilities of scores *at least as* extreme as a-b. The probability that player $\mathcal{I}$ wins by at least $(7+h)$-$(5+h)$ given 5-5 is

$$\sum_{k=h}^{\infty} \Pr(7+k, 5+k \mid 5\text{-}5) = \sum_{k=h}^{\infty} c_g^k (1-c_g)/2 = c_g^h/2.$$

Thus the probability of playing a fifth set with at least $12 + 2h$ games is

$$2 \times (3/8) \times \sum_{k=0}^{5} \left(\binom{5}{k} g^k (1-g)^{5-k} \right)^2 \times \left(g^2 + (1-g)^2 \right)^h /2.$$

The results in Table 2.2 are based on this formula for $h = 63$.

(c) Rule changes: the no-ad rule

If a rule were introduced of playing only one point at deuce (the no-ad rule), then the probability of winning a game without deuce remains g_{nodeuce} as given on page 2 above. However, the winning probability via deuce changes, namely to

$$\tilde{g}_{\text{deuce}} = \Pr(4,3) = p \Pr(3,3) = 20p^4(1-p)^3,$$

where $\Pr(3,3)$ is also given on page 2. The probability $\tilde{g}$ of winning a game thus becomes

$$\tilde{g} = g_{\text{nodeuce}} + \tilde{g}_{\text{deuce}} = p^4(-20p^3 + 70p^2 - 84p + 35),$$

corresponding to the formula on page 28 of the book.

Chapter 4

Big points in a game

The importance of the score 40-30 depends on the probability of winning the game from deuce (3-3). This probability is given by

$$g(3,3) = \frac{p^2}{p^2 + (1-p)^2},$$

which can be obtained by dividing the probability of winning a game via deuce (g_{deuce}) by the probability of deuce, as given on page 2 above. Hence, the importance of the score 40-30 is

$$imp(40,30) = g(4,2) - g(3,3) = 1 - g(3,3) = \frac{(1-p)^2}{p^2 + (1-p)^2}.$$

Likewise, the importance of the score 30-40 is

$$imp(30,40) = g(3,3) - g(2,4) = g(3,3) - 0 = \frac{p^2}{p^2 + (1-p)^2}.$$

Chapter 9

Impact on the paycheck

Serving efficiently increases the probability of winning the match and thus the expected tournament paycheck. On page 153 of the book we claim that the expected paycheck for the efficient player will rise by 18.7% for men and by 32.8% for women. This can be derived as follows.

We consider a tournament with 128 players (seven rounds, as in a grand slam tournament). All players in this hypothetical tournament have the same strength, except one, say player $\mathcal{I}$. Hence, in each match the probability of winning is 0.5 for each player, except for $\mathcal{I}$ who has a probability of $0.5+\epsilon$ of winning a match and for his or her opponent who has a probability of $0.5-\epsilon$ of winning a match against $\mathcal{I}$. In our experiment we take

$$\epsilon = 0.027\,(\text{men}), \qquad \epsilon = 0.044\,(\text{women}).$$

In addition, we assume that the paycheck doubles each round, more specifically, a player receives \$1 when losing in round 1, \$2 when losing in round 2, \$4 when losing in round 3, and so on until \$64 when losing in round 7 (the final) and \$128 when winning the final.

We then find for player $\mathcal{I}$:

$$\begin{aligned} \Pr(\text{losing in round } k) &= (0.5+\epsilon)^{k-1}(0.5-\epsilon) \qquad (k=1,\dots,7) \\ \Pr(\text{winning the tournament}) &= (0.5+\epsilon)^7. \end{aligned}$$

The expected gain for player $\mathcal{I}$ is therefore

$$G(\epsilon) = \sum_{k=1}^{7} 2^{k-1}(0.5+\epsilon)^{k-1}(0.5-\epsilon) + 128(0.5+\epsilon)^7.$$

We find

$$G(0) = \$4.5, \qquad G(0.027) = \$5.343404, \qquad G(0.044) = \$5.974438.$$

The relative increases are therefore

$$\frac{5.343404}{4.5} = 1.187, \qquad \frac{5.974438}{4.5} = 1.328,$$

and this corresponds to the percentages 18.7 and 32.8 mentioned on page 153.

Comments and suggestions

Comments and suggestions are welcome and should be sent to Jan Magnus or to Franc Klaassen.