The Tennis Formula: How it can be used in Professional Tennis

A. James O’Malley
Harvard Medical School
Email: omalley@hcp.med.harvard.edu

September 29, 2007
Overview

- Tennis scoring
- The tennis formula and its properties
- Other tennis-related formulas
- Applications of tennis formula
- Future work
Scoring in tennis

• points: love, 15, 30, 40, deuce, advantage.

• games: love, 1, 2, 3, 4, 5, 6, (7).

• sets: love, 1, 2, (3).

• first to win four points or more by margin of two wins the game.

• first to win six games by margin of two or otherwise seven games wins the set (tiebreaker at six all).

• first to win two (or three) sets wins the match.
“Tennis Formula”

• Let $p$ denote the probability that a player wins a single point serving.

• Assume probability is fixed throughout game (match).

$$\Pr(\text{Win game}) = p^4 + 4p^4(1 - p) + 10p^4(1 - p)^2$$

$$+ 20p^3(1 - p)^3 \cdot \frac{p^2}{1 - 2p(1 - p)}$$

$$= p^4 \left( 15 - 4p - \frac{10p^2}{1 - 2p(1 - p)} \right)$$
Tennis formula, its derivative, and integral functions

Pr(Game)

Derivative function

Integral function

Pr(win if p<0.5) = 0.0616  Pr(win if p>=0.5) = 0.4384
Properties of tennis formula

- Asymmetric - point of inflection at \( p = 0.5 \).
- Monotone increasing
- Derivative function reveals where improve performance is most beneficial.
  \[
  \frac{d Pr(p)}{dp} = 20p^3 \left( 3 - p + \frac{5p^3 - 3p^2 + 4p^4}{(1 - 2p(1 - p))^2} \right)
  \]
- Integral function gives probability of winning when serving probability selected at random.
  \[
  \int_0^p Pr(x) \, dx = -\frac{2}{3}p^6 + 2p^5 - \frac{5}{4}p^4 - \frac{5}{6}p^3 + \frac{5}{4}p + \frac{5}{8} \log(1 - 2p(1 - p))
  \]
  - Average over whole range \( \int_0^1 Pr(x) \, dx = 0.5 \).
Other probabilities

• Probability of winning:
  – tie-breaker.
  – set or match.
  – from a break down in final set.

• Derive similarly to the tennis formula; using tree diagram/dynamic programming approach.
Probability of winning tiebreaker

- Tie-breaker is longer than a regular service game.
  - Involves both players serving, $q =$ opponents probability of winning point on serve.
  - When $q = 1 - p$ expect curve to be steeper than for the tennis formula.
Probability of winning tie–breaker

Pr(Win receiving point) = 0.5

Pr(Win serve point) = Pr(Win receiving point)

Pr(Win serve point) − Pr(Opponent wins serve point) = 0.02
Probability of winning set and match

• Functions of game and tie-breaker winning probabilities.
  – Thus, also of point-winning probabilities.

• Interested in how steeply odds favor better player.
Probability of winning match

Pr(Win receiving point) = 0.5

Pr(Win serve point) = Pr(Win receiving point)

Pr(Win serve point) − Pr(Opponent wins serve point) = 0.02
Difference of probabilities over match duration

$Pr(Win | 5 \text{ sets}) - Pr(Win | 3 \text{ sets})$
Comparison of tennis formula to empirical data?

- Formula’s are based on assumptions:
  - Independence between points.
  - Homogeneous probabilities.

- Obtained data from Wimbledon 2007 (Mens singles).

- Compare empirical game winning percentages to predictions.
Predicted v. Actual Proportion of Service Games Won

Mean predicted = 0.8236
Mean observed = 0.8309

$t$-stat of difference = −0.16
Lack of homogeneity of points across game

- 118 saves out of 208 break points, $p_{\text{save}} = 0.549$.
- 2,101 out of 3,156 service points won at other stages of game, $p_{\text{other}} = 0.666$.
- P-value of difference $\approx 0.0053$. 
Applications of tennis formula

• By players to focus training efforts.
• By players to evaluate where to concentrate match preparation.
• By commentary teams to make broadcast more interesting.
• Useful in determining effect of a rule change.
Training and match preparation

• Compute proportion of points one on serve and while receiving against all opponents.

• Evaluate corresponding probabilities of winning a match.

• Determine if more beneficial to improve serve or return game.

• Work on improving that aspect of game.

• Could extend this by averaging over types of opponents (left-handers, right-handers) to obtain more accuracy.

• Before playing a match analyze head-to-head data.
Example

- Probability win service point = 0.65.
- Probability win receiving point = 0.37.
- Probability win 3 set match = 0.5985.
- Suppose focused training could improve serve probability by 1.1 percentage points or return by 1 percentage point. Where to focus effort?
  - If improve service by 10%: Pr(match) = 0.6497.
  - If improve return by 10%: Pr(match) = 0.6466.
- Better to improve serve!
Making tennis commentary more interesting

• Report likelihood that each player wins match if:
  – Current point-winning percentage is maintained.
  – Players revert to historical winning proportions.
  – Probabilities became equal.
  – Stopped playing and tossed a coin.

• Calibrate statement “match is effectively over if player A breaks serve”.

### Chance of winning when break down in final set.

<table>
<thead>
<tr>
<th>Situation</th>
<th>Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p = 0.62$</td>
</tr>
<tr>
<td>4-5</td>
<td>0.1420</td>
</tr>
<tr>
<td>3-5</td>
<td>0.0546</td>
</tr>
<tr>
<td>2-5</td>
<td>0.2033</td>
</tr>
<tr>
<td>3-4</td>
<td>0.0850</td>
</tr>
<tr>
<td>2-3</td>
<td>0.1675</td>
</tr>
</tbody>
</table>
Rule change

• In 1999 a change in the scoring of tennis was proposed.
• Replace deuce-advantage system with sudden death.
• At deuce the next point decides the game.
• Pete Sampras was against, Andre Agassi supported, the change.
New Tennis formula

• Probability of winning game under new scoring system changes to:

\[ \text{pr(game - new)} = p^4 + 4p^4(1 - p) + 10p^4(1 - p)^2 + 20p^4(1 - p)^3 \]

• Compute change in probability of winning match.
### Sampras-Agassi Data (from 1999)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Sampras</th>
<th>Agassi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serving point</td>
<td>0.709</td>
<td>0.657</td>
</tr>
<tr>
<td>Return point</td>
<td>0.371</td>
<td>0.418</td>
</tr>
<tr>
<td>Pr(Win match - new)</td>
<td>0.8210</td>
<td>0.8092</td>
</tr>
<tr>
<td>Pr(Win match - old)</td>
<td>0.8331</td>
<td>0.8296</td>
</tr>
<tr>
<td>Net gain</td>
<td>-0.0121</td>
<td>-0.0205</td>
</tr>
</tbody>
</table>
Future work

• More realistic models - allow probabilities to vary through stages of match.
  – At deuce, on break- or set-points, between sets.

• Use models to examine player performance at crucial stages of a match.
  – When to be most wary or optimistic against certain opponents.