Tennis Statistics

Course Convenor
Tristan Barnett

Unit Description
Tennis has evolved as a leading international spectator sport. It has a unique scoring structure in the accumulation of points, games and sets to ultimately win the match. Tennis can be uncertain as to when the match will finish; as was the case in the 11 hour match between John Isner and Nicholas Mahut at the first-round of Wimbledon 2010. Consequently this match led to controversial debates over scoring systems. The challenge system for close line calls introduced at the 2006 US Open has also led to much controversy. This course uses quantitative methods to analyse such topics and obtain solutions that could potentially assist in the decision-making process within tennis regulations. Other aspects addressed are predictions, tennis data, serving strategies and resource allocation; that could potentially assist in elite player performance.

Unit Outcomes

- Learn concepts in Markov Chains, Binomial theorem, Recursion formulas, generating functions, probability theory, exponential smoothing and game theory
- Become familiarized with Excel software by hands-on experience in building your own tennis calculator, which in turn assists in the understanding of probability and statistical concepts
- Apply operation research techniques to addressing real-world problems in tennis
- Have a greater appreciation of the sport of tennis through its history and unique scoring structure
- Prepare students with quantitative skills for careers in tennis and sport abroad

Structure of the course
The course is divided into two parts. Part 1 defines a mathematical model by calculating the chances of winning and duration of a match. Part 2 focuses on applications in performance aspects of tennis data, predictions, serving strategies and resource allocation; regulation aspects by analysing scoring systems and challenge the system.

Resources
The textbook for the course is:
Barnett T and Brown A (2012). The Mathematics of Tennis. Strategic Games.
www.strategicgames.com.au/tennisbook.pdf

Tennis calculator
www.strategicgames.com.au/tennisdeucesim.xlsx

Unit Schedule

Week		Book	References
1	Introduction		
	Part 1: Mathematics		
2	Winning a game	Ch 1	
3	Winning a game	Ch 1	
4	Winning a match	Ch 2,3	
5	Duration of a game	Ch 4	
6	Duration of a game	Ch 4,5	
7	Duration of a match	Ch 5,6,7*	[9]
	Part 2: Applications		
8	Tennis Data		[1],[2]
9	Predictions	Ch 8	
10	Serving Strategies		[3]
11	Resource Allocation		[4]
12	Scoring Systems		[5], [8]
13	Challenge System		[6],[7]

*optional chapter

Assessment

Item	**Amount**
Tutorial Exercises	20%
Mid-Semester Exam	40%
Semester Exam	40%

References

[1] Barnett T (2013). Summarizing tennis data to enhance elite performance. To appear in Coaching & Sport Science Review.

[2] Barnett (2010). How the interpretation of match statistics affects player performance. Journal of Medicine and Science in Tennis 15(2), 23-27.

[3] Barnett T (2012). Game theoretic solutions to tennis serving strategies. Coaching & Sport Science Review 56.

[4] Barnett T (2010). Using game theory to optimize performance in a best-of-N set match. Journal of Quantitative Analysis in Sports 6(2).

[5] Barnett T (2012). Analyzing tennis scoring systems: from the origins to today. Journal of Medicine and Science in Tennis 17(2), 68-77.

[6] Barnett T (2010). Applying strategies to the tennis challenge system. Journal of Medicine and Science in Tennis 15(1), 12-15.

[7] Barnett T (2011). Challenges with the tennis challenge system. Significance - web exclusive.

[8] Barnett T (2014). Some alternative men's grand slam singles scoring systems. Strategic Games.

[9] Barnett T (2014). A recursive approach to modelling the amount of time played in a tennis match. Strategic Games.

Part 1: Mathematics

Week Two - winning a game

Calculations are obtained for the probability of winning a game from the outset. The techniques to obtain these calculations consist of counting paths, Binomial theorem and Markov Chain theory.

Week Three - winning a game

Calculations are obtained for the probability of winning a game. This consists of winning the game from any point score within the game (i.e. winning the game from 30-0) and winning the game to a specific point score from any point score within the game (i.e. winning the game to 0 from 30-0). The techniques to obtain these calculations consist of backward recursion and forward recursion.

Week Four - winning a match

Backward and forward recursion calculations are applied to obtain the probability of winning a tiebreak game from any point score within the game, winning a tiebreak and advantage set from any point and game score within the set; and winning an all tiebreak set and final set advantage match from any point, game and set score within the match. Forward recursion formulas generalize to obtain probabilities of reaching a specific point score from any point score within a game, reaching a specific point score from any point score within a tiebreak game, reaching a specific game score from any point and game score within a tiebreak and advantage set, and reaching a specific set score from any point, game and set score within an all tiebreak set and final set advantage match.

Week Five - duration of a game

Calculations are obtained for the duration of a game. This consists of two distributions; namely the distribution of the total number of points played in a game from the outset and the distribution of the number of points remaining in the game from the outset. Calculations are also obtained for the parameters of distribution consisting of mean, variance, coefficient of skewness and coefficient of kurtosis from the outset. The techniques to obtain these calculations consist of the Binomial theorem and generating functions.

Week Six – duration of a game

Calculations are obtained for the duration of a game. This consists of two distributions; namely the distribution of the total number of points played in a game from any point score within the game and the distribution of the number of points remaining in the game from any point score within the game. Calculations are also obtained for the parameters of distribution consisting of mean, variance, coefficient of skewness and coefficient of kurtosis from any point score within the game. The techniques to obtain these calculations consist of forward recursion and backward recursion.

Week Seven – duration of a match

Calculations are obtained for distributions of the total number of points played and the number of points remaining from any point score within a tiebreak game, distributions of the total number of games played and the number of games remaining from any game score within a tiebreak and advantage set, and distributions of the total number of sets played and the number of sets remaining from any set score within an all tiebreak set and final set advantage match. Calculations are also obtained for the mean, variance, coefficient of skewness and coefficient of kurtosis of the

number of points remaining in a tiebreak game from any point score within the game, the number of games remaining in a tiebreak and advantage set from any point and game score within the set, and the number of sets remaining in an all tiebreak set and final set advantage match from any point, game and set score within the match.

Part 2: Applications

Week Eight -tennis data

A classification of online and commercially available data for the men's and women's singles professional tennis circuit is obtained. The various data sources are categorized by data type (e.g. ratings, point-by-point data, match statistics), tournament type (e.g. grand slam, Olympics, ITF Circuit) and the year commencing. Through data analysis it is shown how the various data types could be used in building decision support tools that could be readily accessed by coaches in preparation for an upcoming match.

Week Nine - predictions

By assigning two parameters, the constant probabilities of player A and player B winning a point on serve; the probability of winning and duration can be determined using the methods outlined in weeks 2-5. Estimating these two parameters when two elite players meet on a particular surface is now obtained using exponential smoothing techniques, and an updating rule is derived for the match in progress. The method is demonstrated by focusing on the `long' men's singles match between John Isner and Nicholas Mahut at the 2010 Wimbledon Championships. The appeal of how predictions in sports multimedia can be used is also presented.

Week Ten - serving strategies

Risk-taking on serve is analysed to maximize a player's chances of winning a point on the second serve by either serving a common low risk second serve (with a high second serve percentage) or a high risk second serve by decreasing the second serve percentage but increasing the proportion of points won if the second serve goes in. The notion of "importance" of points is defined and there is evidence to suggest that servers could be encouraged to take more risk on the more "important" points. The results could be used by coaches to help determine how much risk their players should take on the second serve. A working example between Andy Roddick and Rafael Nadal is given to support the results.

Week Eleven-resource allocation

A best-of-N set match is analysed, where both players/teams are given the opportunity to increase their probability of winning a set (increase in effort) on one particular set. To gain insight to the problem, a best-of-3 set match (as typically used in tennis) is analysed. Using game theory to obtain an optimal solution, the results indicate that both players should use a mixed strategy, by applying their increase in effort at each set with a probability of one third. A conjecture is devised to obtain an optimal solution for a best-of-N set match. Some applications are given to the theoretical results, which could be used by coaches and players to optimize performance.

Week Twelve - scoring systems

Tennis scoring systems are investigated that have been used throughout history — from Royal Tennis in 1490 to the most recent change to doubles Lawn Tennis in 2006. By identifying how the game has changed (such as technology in equipment), helps in establishing "reasonable" scoring systems that could be used for today. Based on this information and obtaining mathematical results of scoring systems, recommendations are given for men's and women's singles and doubles events. Actual matches are given to demonstrate why changes in many scoring systems are necessary.

Week Thirteen - challenge system

The challenge system for close line calls in tennis has been used on the ATP and WTA tour for grand slam events since the 2006 US Open, and was designed to increase fairness for players by obtaining accurate line calls and enhance spectator interest through video technology. In the current system, players have unlimited opportunity to challenge, but once three incorrect challenges are made in a set, they cannot challenge again until the next set. By applying the 'importance' concept it is demonstrated how the challenge system could be potentially improved for overall player fairness.

Resources

Online

Strategic Games – www.strategicgames.com.au
RMIT Sports Statistics - www.rmit.edu.au/mathsgeo/research/systemmodelling/sportstats
Franc Klaassen's Homepage - www1.fee.uva.nl/pp/klaassen/
Stephen Clarke's Homepage - www.swinburne.edu.au/lss/staff/view.php?who=sclarke

Books

Cohen G and de Mestre N (2007) Figuring Sport. UTS Printing Services, Sydney.
Ladany S and Machol R (1977) Optimal Strategies in Sport. North-Holland, Amsterdam.
Bennett J (1998) Statistics in Sport. Arnold, London.
Eastaway R and Haigh J (2011) The Hidden Mathematics of Sport.
Sadovskii LE and Sadovskii AL (1993) Mathematics and Sports. American Mathematical Society.
DasGupta A (2010) Fundamentals of Probability: A First Course. Springer, New York

Publications

MathSport - www.anziam.org.au/Mathsport
ITF Tennis Science & Technology - www.itftennis.com/home.aspx
ITF Coaching & Sport Science Review - www.itftennis.com/home.aspx
Journal of Medicine and Science in Tennis - www.stms-web.org/
International Journal of Computer Science in Sport - www.iacss.org/index.php?id=30
Journal of Quantitative Analysis in Sports - www.degruyter.com/view/j/jqas
Significance - www.significancemagazine.org/