Tennis Statistics with Applications to other Sports

Strategic Games
ABN: 64663038362
www.strategicgames.com.au

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. The course firstly introduces mathematical modelling in tennis to determine the chances of players winning as well as match duration. The Binomial theorem, summing an infinite series, Markov Chain theory, moment and cumulant generating functions, forwards and backwards recursion, distribution theory, probability theory, Normal distribution and Normal Power approximation are taught. Recursion formulas are particularly useful for obtaining numerical results in Excel without any programming and allows students to develop their own calculators. Applications of mathematical modelling in tennis are applied to performance aspects of prediction modelling, serving strategies, the analysis of tennis data and resource allocation; and regulation aspects of alternative scoring systems and the line call challenge system. Exponential smoothing, game theory and Bayesian updating are taught. Similar techniques used to model tennis can then be applied to table tennis and badminton. Volleyball and squash are more complicated to model than tennis due to the rotation of serve and volleyball is addressed in the next chapter, once again using recurrence formulas in Excel. Head-to-head matches with a fixed amount of playing time are relatively straight forward to implement in Excel and AFL is modelled as one example with alternative scoring systems to eliminate the possibility of a draw.

Unit Outcomes

- Learn concepts in Markov Chains, Binomial theorem, recursion formulas, generating functions, probability theory, distribution theory, Normal distribution, Normal Power approximation, exponential smoothing, game theory and Bayesian updating
- Become familiarized with Excel software by hands-on experience in building your own tennis, table tennis, badminton, squash, volleyball, AFL and other team sports calculators, which in turn assists in the understanding of probability and statistical concepts
- Apply operation research techniques to addressing real-world problems in tennis and AFL
- 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 four parts. Part 1 defines a mathematical model by calculating the chances of winning and duration of a tennis match. Part 2 focuses on applications in tennis through performance aspects of data analysis, predictions, serving strategies and resource allocation; and regulation aspects by analyzing scoring systems and the challenge system. Part 3 focuses on modelling volleyball and part 4 focuses on modelling AFL.

Resources

The textbooks for the course:

Barnett T and Brown A (2012). The Mathematics of Tennis. Strategic Games.

1. http://strategicgames.com.au/book.pdf

Barnett T (2019). Operations Research in Tennis. Strategic Games.

2. http://strategicgames.com.au/book-ORtennis.pdf

Additional articles for the course:

Barnett T, Brown A and Jackson K (2008). Modelling outcomes in volleyball. In proceedings of the Ninth Australian Conference on Mathematics and Computers in Sport.

3. http://strategicgames.com.au/article9.pdf

Barnett T (2014). Modelling outcomes in volleyball 2. Strategic Games.

4. http://strategicgames.com.au/article43.pdf

Barnett T and Pollard G (2011). Devising new Australian Rules football scoring systems. Journal of Quantitative Analysis in Sports 7(3).

5. http://strategicgames.com.au/article27.pdf

Barnett T (2019). Suggestions to Improve Scoring Systems Across a Range of Racket Sports. Strategic Games.

6. http://strategicgames.com.au/article63.pdf

Calculators for the course:

Java tennis calculator

1. http://strategicgames.com.au/TennisCalc.jar

Excel tennis calculator using analytic methods for the distribution of points played in a set

2. http://strategicgames.com.au/tennisdeuceanalytic.xlsx

Excel tennis calculator using simulation for the distribution of points played in a set/match

3. http://strategicgames.com.au/tennisdeucesim.xlsx

Excel point-by-point predictions for the Roddick vs El Aynaoui match plated at the 2003 Australian Open

4. http://strategicgames.com.au/RoddickElAynaoui.XLS

Excel tennis calculator with simulation for the distribution of points played in a set/match with the no-ad game

5. http://strategicgames.com.au/NoAdsim.xlsx

Excel tennis calculator using simulation for the distribution of points played in a set/match with the 50-40 game

6. http://strategicgames.com.au/5040.xlsx

Excel tennis calculator on the line call challenge system

7. http://strategicgames.com.au/tennischallenge.xlsx

Excel badminton, table tennis, volleyball, beach volleyball and squash calculator

8. http://strategicgames.com.au/sports.xlsx

Excel AFL calculator

9. http://strategicgames.com.au/AFL.xlsx

Data for the course:

ATP Tour

10. https://www.atptour.com/en/scores/results-archive

WTA Tour

11. https://www.wtatennis.com/scores

OnCourt: tennis software

12. http://oncourt.info/

Tennis Navigator: tennis software

13. http://www.tennisnavigator.com/

Australian Open

14. https://ausopen.com/

French Open

15. https://www.rolandgarros.com/en-us/

Wimbledon

16. https://www.wimbledon.com/index.html

US Open

17. https://www.usopen.org/index.html

Unit Schedule

Week	**Part 1: The Mathematics of Tennis**	Book/Article	Chapter	Calculator/Data
1	Winning a Game	1.	1.	1.
2	Winning a Match	1.	2./3.	1.
3	Duration of a Game	1.	4.	2.
4	Duration of a Match	1.	5./6./7.*	2./3.
	Part 2: Operations Research in Tennis			
5	Predictions	1.	8.	4.
6	Scoring Systems	2.	1.	5./6.
7	Challenge System	2.	2.	7.
8	Serving Strategies	2.	3.	
9	Tennis Data / Analysis of Tennis Data	2.	4./5.	10.-17.
10	Resource Allocation	2.	6.	
	Part 3: The Mathematics of Volleyball			
11	Winning a Set/Match, Duration of a Set/Match	3./4./6.		8.
	Part 4: The Mathematics of AFL			
12	Winning a Match, Margins, Duration of a Match, Scoring Systems	5.		9.
13	Revision			

*optional chapter

Assessment

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

Part 1: The Mathematics of Tennis

Week 1 - 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. Calculations are obtained for the probability of winning a 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 2 - 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 3 - Duration of a Game

Calculations are obtained for the duration of a game from the outset. 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. Calculations are obtained for the duration of a game from any point score within the 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 4 – 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. The Normal distribution has been widely studied, and tables of the probabilities for this distribution are readily available. The basic idea of the Normal Power approximation is to use these tables to estimate the tail probabilities of other distributions. This method uses the first four moments and produces a continuous approximation to the cumulative distribution. The approximation to the frequency distribution can be recovered using differences. Hence approximate frequency distributions are obtained for the number of points played in a set, the number of points played in a match and the time duration of a match.

Part 2: Operations Research in Tennis

Week 5: 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 1-4. Estimating these two parameters when two elite players meet on a particular surface is now obtained using exponential smoothing techniques, and a Bayesian 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 6: 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 7: 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.

Week 8: 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 9: Tennis Data / Analysis of 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. Using a match from the 2004 Australian Open, it is shown how the interpretation of player match statistics can affect player performance. By analyzing match statistics from commercial software providers, further statistics such as a player's average percentage of points won on serve (across many matches) can be obtained. Furthermore, these statistics could be conditioned by many factors such as the court surface. An example is given in a graphical format where a player's percentage wins on the return of the first serve across different court surfaces could be compared to another player and the average player. These types of comparisons in a numerical or graphical display could be useful to players and coaches for tactical and coaching regimes, and further 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 10: 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.

Part 3: The Mathematics of Volleyball

Week 11: Winning a Set/Match, Duration of a Set/Match

A Markov Chain model is applied to volleyball to calculate win probabilities and mean lengths with the associated variances, conditional on both the scoreboard and the server. A feature of this model is that it predicts outcomes conditional on both the scoreboard and the serving team. The inclusion of the serving team in the event space is an essential requirement of this model, and arises from the rule in volleyball that the winner of each point in a set must serve on the following point. The average probability of a team winning a point on serve is less than 0.5, and so rotation of serve is commonplace. The key to the analysis of an evenly contested set is the observation that, from the situation where the scores are level (after at least 46 points have been played), the team that wins the set must eventually win two successive points. If the two points are shared then the score is level once again, although a rotation of server has occurred. This scoring structure, combined with the method of rotating the serve, distinguishes volleyball from other racket sports such as tennis, badminton and table tennis. Results from the model indicate that it is advantageous to be the receiver on the opening point of a set and the team that wins the toss at the start of the fifth set (if the set score reaches 2-all), has an advantage for the remainder of the match. However, due to the rotation of serve after each set, there is no advantage for either side in being server or receiver at the start of the match.

Part 4: The Mathematics of AFL

Week 12: Winning a Match, Margins, Duration of a Match, Scoring Systems

A Markov Chain model is applied to AFL to calculate win probabilities, margin probabilities and the duration of a match. In every AFL season, there are usually one or two round matches that end in a draw. This situation is handled in finals matches by playing extra time. Whilst the scoring systems used in finals matches could be applied to round matches to eliminate draws, we devise new alternative systems that have better statistical properties by increasing the probability for the stronger team to win, increasing fairness and spectator interest, and reducing the likelihood of long matches occurring. Further, these new scoring systems could also be used in finals matches.

Week 13: Revision

Resources

Online Content

Strategic Games (html)
Franc Klaassen (html)
Minimax Play at Wimbledon by Mark Walker and John Wooders (pdf)

Publications

Proceedings of Mathematics and Computers in Sport (html)
Proceedings of Mathsport International
International Journal of Computer and Science in Sport (html)
Journal of Medicine and Science in Tennis (html)
Proceedings of the International Congress of Tennis Science and Technology (html)
ITF Coaching and Sport Science Review (html)
Science and Racket Sports
Journal of Quantitative Analysis in Sports (html)
Significance Magazine (html)
CHANCE (html)
International Journal of Sports Science and Coaching (html)
Journal of Sports Science & Medicine (html)
Journal of the Operational Research Society (html)
Foresight: The International Journal of Applied Forecasting (html)
International Journal of Forecasting (html)
Sports Forecasting edited by Herman Stekler and Leighton Vaughan-Williams (html)
Sports Forecasting edited by Ian McHale and Tim Swartz (pdf)
IMA Journal of Management Mathematics (pdf)
IMA Journal of Management Mathematics: Mathematics in Sport edited by David Percy, James Reade and Hugh Morton (html)
IMA Journal of Management Mathematics: Mathematics in Sport edited by John Norman and Philip Scarf (html)

Books

Statistics in Sport by Jay Bennett (html)
Figuring Sport by Graeme Cohen and Neville de Mestre (html)
The Hidden Mathematics of Sport by Rob Eastaway and John Haigh (html)
Analyzing Wimbledon: The Power of Statistics by Franc Klaassen and Jan Magnus (html)
Optimal Strategies in Sport by Shaul Ladany and Robert Machol (html)
Mathematics and Sports edited by Joseph Gallian (html)
Mathematics and Sports by LE Sadovskii and AL Sadovskii (html)
Analysing Oceania's qualification path to the World Cup: a retrospective ratings- based statistical analysis by Anthony Bedford and Cliff Da Costa (html)
Operational Research Applied to Sports edited by Mike Wright (html)
Gambling and Sport: A Statistical Approach by John Croucher (html)
Moneyball by Michael Lewis (html)
Duckworth Lewis: The Method and the Men behind it by Frank Duckworth and Tony Lewis (html)

Sports Math: An Introductory Course in the Mathematics of Sports Science and Sports Analytics by Roland Minton (html)
Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football by Wayne Winston (html)
Mathematics in Games, Sports, and Gambling: - The Games People Play by Ronald Gould (html)
The Math of Sports: Integrating Math in the Real World (Integrating Math in the Real World Series) by Hope Martin and Susan Guengerich (html)
Analytic Methods in Sports: Using Mathematics and Statistics to Understand Data from Baseball, Football, Basketball, and Other Sports by Thomas Severini (html)
Sports Math: Slam-Dunk Learning With Super-Fun Reproducible Activities That Build Essential Math Skills by Denise Kiernan (html)
Ball Game Math by Katie Marsico (html)
Basketball: The Math of the Game by Thomas Adamson (html)
Mathematics Is Not a Spectator Sport by George Phillips (html)
How to Take a Penalty: The Hidden Mathematics of Sport by Rob Eastaway and John Haigh (html)
Mathematics and Science for Exercise and Sport: The Basics by Craig Williams, David James and Cassie Wilson (html)
Fantasy Basketball and Mathematics: A Resource Guide for Teachers and Parents, Grades 5 and Up by Dan Flockhart (html)
Fantasy Football and Mathematics: Student Workbook by Dan Flockhart (html)
Basketball Analytics: Objective and Efficient Strategies for Understanding How Teams Win by Stephen Shea and Christopher Baker (html)
Mathematics in Sport by Bob Dengate (html)
Fantasy Baseball and Mathematics: A Resource Guide for Teachers and Parents, Grades 5 and Up by Dan Flockhart (html)
Economics, Management and Optimization in Sports edited by Sergiy Butenko and Jaime Gil-Lafuente (html)
Run Like You Stole Something: The Science Behind the Score Line by Damian Farrow and Justin Kemp (html)
Why Dick Fosbury Flopped: and answers to other big sporting questions by Damian Farrow and Justin Kemp (html)