Teaching Mathematics and Statistics Using Tennis

Reza Noubary

ABSTRACT: The widespread interest in sports in our culture provides a great opportunity to catch students' attention in mathematics and statistics classes. Many students whose eyes glaze over after few minutes of algebra will happily spend hours analyzing their favorite sport. As their teachers we may use this to enhance our teaching of mathematical and statistical concepts. Fortunately, many sports lend themselves to this. This article analyzes a tennis match with a view towards its use as an aid for teaching mathematical and statistical concepts. It shows that through sports students can be exposed to the basics of mathematical modeling and statistical reasoning using material that interests them. For those who plan to become mathematics teachers, in high school or college, it provides a source of material that could enrich their teaching. It has been my experience that using sports in the classroom increases students' interest in mathematics and statistics and the public's interest in university activities.

1. **INTRODUCTION**. The difficulties faced by educators teaching mathematics and statistics are well known. To help, many textbooks try to motivate students by introducing varied applications. This addresses both students' desire to see the relevance of their studies to the outside world and also their skepticism about whether mathematics and statistics have any value. This idea works mostly with students who are committed to a particular academic or career field. For typical students, applied examples may fail to motivate if they are not of immediate concern to them or they do not occur in their daily lives.

Fortunately, students have some common interests that we can build upon when teaching mathematics and statistics. Connecting their studies to something that interests or concerns them almost always works better. Unfortunately, it is not always easy to find something that will motivate the majority of students. I have tried several things and have concluded that games and sports are the best way to accomplish this. We can help students build their studies on a foundation, an understanding of a sport that they already possess. I think this is adaptable at all levels at which mathematics and statistics is taught, from junior high to graduate schools. In what follows we discuss other advantages of using sports.

General
⟨ Sports have a general appeal and scientific methods can be applied to them.
⟨ Sports are a part of everyday life, especially for young people.
⟨ Students usually enjoy sports and show a great deal of interest in mathematics and statistics applied to them.
⟨ A major part of the calculus and statistics sequences offered at college level can be taught using a sport.
⟨ Most students can relate to sports and can understand the rules and meanings of the statistics presented to them.

Specific

Sports data offer a unique opportunity to test methodologies offered by mathematics and statistics. I believe it is hard to find an area other than sports where one could collect reliable data with the highest precision possible. In addition to the quality measurements, here we have access to the names, faces, and life history of the participants and their coaches, trainers, and everyone involved. Almost all other data producing disciplines are susceptible to "data manipulation and data mining" and error, since unlike sports, they are not watched by millions of fans and reported on in the media. A theoretical result can be tested only when data is reliable and satisfies the conditions under which it was developed. If the validity of data produced or collected by an individual or an organization cannot be confirmed, one may end up being suspicious of the results obtained and the methodology applied.

Consider, for example, track and field. The nature and general availability of track and field data have resulted in their extensive use by researchers, teachers, and sports enthusiasts. The data are unique in that they:

(1) Possess a meaning that is apparent to most people.
(2) Are collected under very constant and controlled conditions, and thus are very accurate and reliable.
(3) Are recorded with great precision (e.g., to the hundredth of a second in races), and thus permit very fine differentiation of change or differences;
(4) Are both longitudinal (100 years for men's records) and cross-sectional (over different distances and across gender).
(5) Are publicly available at no cost. Thus, they provide wonderful data sets to test mathematical and statistical models of change.

2. **AN ILLUSTRATIVE EXAMPLE**. The focus of a lesson could be on a single concept based on examples from several different sports or on several different concepts based on a single sport. In this article we illustrate how a single sport, tennis, may be used to teach mathematical and statistical concepts.

The Quirks of Scoring. During its early years, tennis used a variety of scoring systems. By the time of the first championship at Wimbledon in 1877, the All England Croquet Club had settled on a scoring system based on court tennis. This system remained unchanged until the introduction of tie-breakers in 1970.

One quirk of tennis scoring is that strange names are used for points in scoring a game: *love, fifteen, thirty, forty, game*. Although no one knows the origin of this odd system, it has been proposed that *fifteen, thirty, forty-five, sixty* were originally used to represent the four quarters of an hour. Over the years the score *forty-five* became abbreviated as *forty*. (In informal play, *fifteen* is sometimes abbreviated as *five*.) It would be simpler to score the game: *zero, one, two, three, and four*. However, the weird point names give no advantage to either player.

A more important quirk is that a game must be won by two points. If players each score three points, the score is called *deuce*, rather than *40:40*. If the server wins the next point, the score becomes *advantage in*. If the server wins again, she wins the game, otherwise the score returns to *deuce*. If server loses the next point, the score becomes *advantage out*. If the server loses again, she loses the game, otherwise the score returns to *deuce*. This feature of tennis scoring increases the chance that the stronger player will win, as we shall see.

Consider a game of tennis between two players, A and B. The progression of the game can be used to teach many statistical concepts and critical thinking. Throughout, for any event E we use P(E) to denote the probability that E occurs.
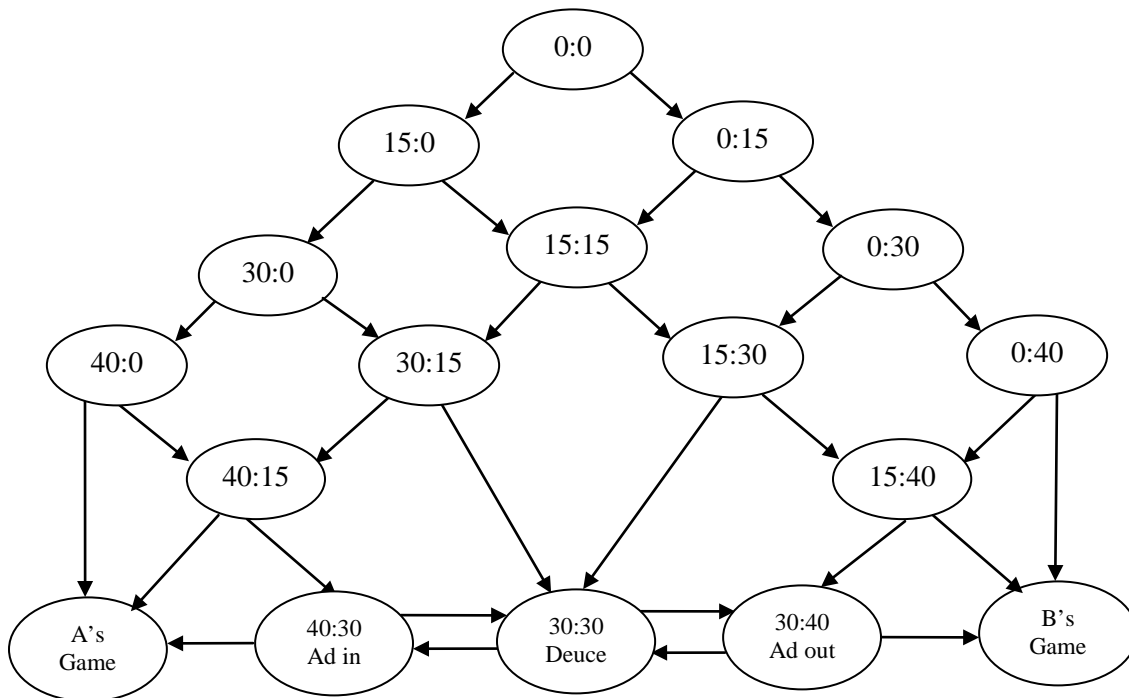
1. Let

$$x = \text{P (A wins a point)}$$
$$y = 1 - x = \text{P (B wins a point)}$$

To simplify the modeling and analysis of the game we may first assume that x is fixed. Do you think this is a reasonable assumption?

Objective: **To teach critical and logical thinking versus practical significance.**

2. Starting from 0:0 what are the possible outcomes after one exchange (one point), two exchanges, and so on?

Objective: **To teach concepts such as sample space (universal set), events (sets) and their algebra.**

3. How do you assign probabilities to the outcomes of the sample space after one exchange? The possible outcomes are: 0:15 and 15:0.

Objective: **To teach concepts such as quantification of uncertainty, probability (classical, objective, and subjective), and odds.**

4. Do you think winning a point will affect the probability of winning the next point?

   Objective: **To teach concepts such as conditional probability and independence.**

5. Suppose that the answer to question 4 is no. Let $x = 0.60$. How do you find the probabilities such as P (15:15) or P (30:30) or P (40:15), etc.? How do you find these probabilities if the answer to question 4 is yes?

   Objective: **To teach combinations, multiplication rule, Bernoulli, Binomial and Poisson distributions.**

6. Find   (*a*) P (A wins the game without deuce),
              (*b*) P (A wins the game after reaching deuce),
              (*c*) P (A wins the game).

   Objective: **To teach addition rule, infinite series, and geometric progression.**

7. Find   (*a*) P (A wins the set without tie-break),
              (*b*) P (A wins the set after a tie-break),
              (*c*) P (A wins the set).

   Objective: **To teach modeling and problem solving.**

8. Find   P (A wins the match).

   Objective: **To teach pattern identification and model building.**

9. Find general formulas for probabilities in questions 6, 7 and 8.

   Objective: **To teach functions, graphs, and function of functions (composite functions).**

10. Let $e = x - y$ represent the edge in one point. For example, $x = 0.51$ means player A has a 0.02 edge over player B in one point. Find edges in one game, one set, and the match. Hint: call the edge in one point Δx and edge in a game Δy.

    Objective: **To teach the derivative, chain rule, and differential equations.**

11. Let $r = x/y$. Since $0 \le x, y \le 1$ it follows that $0 \le r < \times$ and $r = 1$ for $x = y = 0.5$. Express the probabilities in question 9 in terms of r.

    Objective: **To teach transformations and homogeneous polynomials.**

12. Let 0, 1, 2, 3, and 4 represent the scores 0, 15, 30, 40, and the game respectively.
    Let $g(i, j) = $ P (A wins the game starting from the score $(i : j)$). Show that

$$g(i, j) = xg(i+1, j) + yg(i, j+1)$$

Also show that $g(3,3) = x^2/(x^2+y^2)$.

Objective: **To teach recursions and difference equations.**

13. Think about a game that has reached the state deuce (or 30: 30). There is no limit to how long the game could go on. From this point, the game could reach one of the five possible states. Let 1, 2, 3, 4, and 5 denote the states: A's game, B's game, Deuce, Advantage A, and Advantage B, respectively. The game moves from state to state until one player wins. The probabilities of moving from one state to another can be summarized as

$$
\begin{array}{c c c c c c}
 & 1 & 2 & 3 & 4 & 5 \\
1 & 1 & 0 & 0 & 0 & 0 \\
2 & 0 & 1 & 0 & 0 & 0 \\
3 & 0 & 0 & 0 & x & y \\
4 & x & 0 & y & 0 & 0 \\
5 & 0 & y & x & 0 & 0
\end{array}
$$

Objective: **To teach matrices, Markov chains and the states of a Markov chain.**

14. Suppose that the game is now in state deuce (state 3). This can be expressed as the state matrix;

$$
\begin{array}{c c c c c}
1 & 2 & 3 & 4 & 5 \\
\left[0 \right. & 0 & 1 & 0 & \left. 0\right]
\end{array}
$$

Show that after one and two exchanges the state matrices are respectively

$$
\begin{array}{c c c c c}
\mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} \\
\left[0\right. & \mathbf{0} & \mathbf{0} & x & \left. y\right]
\end{array}
\qquad
\begin{array}{c c c c c}
\mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} \\
\left[x^2\right. & y^2 & 2xy & 0 & \left. 0\right]
\end{array}
$$

Objective: **To teach matrix algebra.**

15. Starting from deuce
    a. How many exchanges (points) are expected to be played before the game ends?
    b. How many times is each state expected to be visited/ revisited before the game ends?

    Objective: **To teach stationary solution, inverse of a matrix, and the fundamental matrix.**

16. Suppose now that x_1, x_2 represent respectively the probabilities in part *a* and x_3, x_4 represent respectively the probabilities in part *b*.
    a.   P(A wins a point when serving)   and   P(A wins a point when receiving)
    b.   P(A wins a point after winning a point)   and   P(A wins a point after losing a point)
    Find the probabilities of winning a game, a set and the match for player A.

   Objective: **To teach basic concepts of modeling.**

17. Consider a tournament like the Davis Cup. Suppose that countries *A* and *B* each have three players represented as A_1, A_2, A_3 and B_1, B_2, B_3 respectively. Suppose that the following matrix represents their chances of winning or losing against each other.

	B_1	B_2	B_3
A_1	40%	52%	50%
A_2	40%	41%	30%
A_3	55%	45%	60%

For example, using this matrix, we have $P(A_1$ beats $B_1) = 40\%$. In the Davis Cup, each team decides which player plays the first, second, etc. game without knowing about the selection of the other team. How do you think teams should make their selection?

Objective: **To teach game theory.**

18. In tennis, the server gets a second chance to serve after missing the first one. Ordinarily, players go for a speedy (strong) but risky first serve and a slow but a more conservative second serve. Analyze all the possible serving strategies and their consequences.

Objective: **To teach basic concepts of decision analysis and its role in the game theory.**

19. How do you summarize statistics related to a tennis player, a team, and a tournament?

Objective: **To teach descriptive statistics.**

20. Suppose that you have data for the speed of player A's first serve. How do you calculate the probability that in the next match the average speed of A's first serves would exceed a certain value?

21. *Objective*: **To teach sampling distribution and central limit theorem.**

22. How do you compare two tennis players? How do you rank tennis players?

Objective: **To teach performance measures, measures of relative standing, z-score, etc.**

23. A claim is made about the performance of a tennis player. Using the player's statistics, how do you validate the claim?

Objective: **To teach hypothesis testing, Type I and Type II errors and P-value.**

24. How can you use the past statistics of a player to predict his or her future performance?

Objective: **To teach estimation (prediction), confidence intervals, regression, time series, and forecasting**.

25. Suppose that you have statistics on the speed of player A's first serves. How do you predict the next record speed and perhaps the maximum possible speed of A's serves?

Objective: **To teach theory of records, asymptotic theory of order statistics, extreme value theory, and threshold theory.**

26. How do you organize a tennis tournament?

Objective: **To teach planning and scheduling.**

27. The winner of men's tennis match must win three out of five sets. Each set has six games. Do you think the present scoring system is fair? For example, player A could win two sets 6-0 and lose three tie-break sets 6-7. So, A could win 30 games and lose only 21 games and yet lose the match. Do you have any suggestion to make the match more balanced?

Objective: **To teach methods for adaptive modeling.**

3**. ACTIVITIES**. We conclude with some activities that use tennis and some mathematical and statistical concepts.

Activity 1: Bouncing Ball

The balls used in different sports have a different amount of bounce. Even the balls used in the same sport may bounce differently because of their age, coverings, or simply because they contain different amounts of air. For consistency, a standard for bounciness must be established for the ball in each sport.

On way to measure the bounciness of a ball is through a quantity known as the coefficient of restitution (COR), defined as square root of the ratio of the rebound height to the initial height from which the ball was dropped:

$$COR = \left(H_{Re\,bound} / H_{Initial} \right)^{1/2}$$

Calculation
Provide answers to the following questions;

1. A tennis ball has a COR of 0.53. If this ball is dropped from a height of 8 feet, how high will it bounce?

2. From what height should a tennis ball with a COR of 0.54 be dropped so it will bounce 12 feet?

3. Suppose that a tennis ball with a COR of 0.55 hits a wall at a speed of 65 mph. With what speed will it rebound?

Critical Thinking
Make up a question pertaining to this lesson that you, the student, would ask if you were a teacher. For example;
⟨ You may want to know if bounciness can be quantified differently and, if so, what would be a consequence?
⟨ Is COR independent of speed? That is, if you throw two identical balls at a wall one with the speed twice the other, would speeds of the rebounds be 2 to 1 too?

Estimation, Modeling.

Select a tennis ball. Suppose that the ball is dropped from a point h feet high and has a COR equal to c.

(a) Develop a model to calculate the heights of the first, second, third, and other bounces.

(b) Think about the total distance traveled by the ball after one, two, etc. bounces. See if you could develop a model for this. Use the models you developed for a long run prediction. For example, what is the estimate for total distance traveled after n bounces?

(c) Suppose that c is unknown. How can you estimate it? For example, you may estimate COR by measuring the bounce n times and by averaging the results. Suggest a value for n.

(d) Estimate the second bounce by first directly measuring it n times and averaging as in part (c). Then apply the mathematical model in part (a) and predict the average of the second bounce. Which estimate do you prefer?

(e) Repeat part (b) for the third, fourth, and other bounces. Do you see any pattern?

Trajectories, Some Classical Functions, and Model Building

Suppose now that a ball will be hit by a racket at a certain angle (similar to serves in tennis). After hitting the ground it will bounce and follow a path like a parabola (similar to the trajectory of, for example, a lob shot in tennis). Model this first for a fixed angle and fixed initial speed. Repeat the process by keeping one variable fixed and letting the other vary. Finally, let both variables vary.

Activity 2: Applying Binomial Distribution, Matrices, Markov Chain, and Derivatives

Consider a match between two players, A and B. Suppose that player A has a 10% edge over Player B in one point. That is, the probability that player A will win a point is 55% and the probability that player B will win a point is 45%.

1. Show that the probability calculations before reaching deuce can be carried out using a Binomial distribution.

2. Find the probability that player A wins a game, a set, and the match given the edge A has in one point. Also calculate the edge in a game, a set, and a match given the edge in one point.

Suppose that the information regarding the players A and B is summarized in a rectangular array as;

	1	2	3	4	5
1	1	0	0	0	0
2	.55	0	.45	0	0
3	0	.55	0	.45	0
4	0	0	.55	0	.45
5	0	0	0	0	1

This is called a transition matrix. It includes the probabilities of moving from one state to another after a point. Here state 1 represents A won the game, state 2 represents advantage A, state 3 represents deuce, state 4 represents advantage B, and state 5 is B won the game.

3. Apply matrix algebra and interpret the results in the context of a tennis game.

4. If we look at the tennis game as a Markov chain, what are the states? Which states are non-recurrent? Which states are recurrent? Which states are absorbing?

5. Let x denote the probability that player A wins a point, and $y = 1 - x$ denote the probability that player B wins that point. It can be shown that

$$P \text{ (A wins the game)} = x^4[1 + 4y + 10y^2 + 20xy^3/(x^2 + y^2)].$$

Replace y by $1-x$ in this equation and find its derivative with respect to x. Calculate the value of the derivative at the point $x = 0.50$. For players close in ability (when the edge in one point is small, e.g., 1%) the resulting value provides the edge in one game. Compare the value obtained using the derivative with the actual value of the edge.

6. Consider the equation in problem 5. Replace y by $1-x$. The resulting function has several properties. For example, the function is symmetric with respect to $x = 0.5$. Study the other properties of this function.

7. Consider the formula in problem 5. Replace y by $1-x$. Suppose that P (A wins the game) = 0.60. Use numerical methods to find x.

8. Find the probability of winning a set as a function of x and show that it is an example of a function of a function. Use it to find the edge in a set both directly and by using the derivative (chain rule) as in problem 5.

Activity 3: Calculations Based on Normal Distribution
1. Suppose that the average speed of a tennis player's first serve is 117mph with a standard deviation of 5mph. What is the probability that this player's next first serve will be;
    a) Slower than 115?
    b) Faster than 120?
    c) Between 116 and 122?

2. Suppose that tennis balls are produced to have COR = 55.5% (target value). To see if the process is on target once a day 50 balls are tested. If the average COR falls outside the interval 53.5% and 57.5 the process is judged out of control. What is the probability that the process will be judged out of control incorrectly? Assume that the standard deviation is 1.5%.

Activity 4: Constructing Confidence Intervals and Testing Hypotheses
1. Look up statistics for the number of aces made in 36 matches by a tennis player of your choice. Construct a 95% confidence interval for the average number of aces for this player. Hint: Use the Central Limit Theorem.
2. A sample of 36 serves of a top player on hard court has mean of 107mph and standard deviation of 6.5mph. His coach claims that the average speed of his serves is 110mph. Check to see if data supports this claim. Use a 0.05 level of significance.

3. Another sample of 36 serves from the same player (problem 2) on a grass court has mean of 105mph and standard deviation of 10mph. Can we conclude that the mean speed of his serves on the hard court is greater (at the 0.05 level) than the mean speed of his serves on the grass court?

4. Construct a confidence interval for the difference between the population means in problems 2 and 3.

Activity 5: Applying Regression and Time Series for Prediction
1. Use statistics for a player of your choice who participated in the latest Wimbledon tournament. Use regression to predict the total points won using, for example, the number of opponents' unforced errors as predictor. Try other factors to see if you can find the best predictors.

3. If we cannot apply regression, we can still use smoothing techniques for prediction when the data form a time series. Time series refers to data with a time index. Use smoothing to predict the number of matches a player of your choice may win next year using the number of matches the player has won in previous years.

Activity 6: Research topics
The analysis of a tennis game can be expanded in many different directions. Examples include Canadian doubles or cut-throat, Australian doubles, regular doubles, and even a five-way game when server sits out for a game. Examples of research topics include determination of the size of a handicap to make a game a fair game and analyzing of the methods used for ranking tennis players.

REZA D.NOUBARY received his B.Sc. and M.Sc. in Mathematics from Tehran University, and M.Sc. and Ph.D. in statistics from Manchester University. His research interests include time series analysis, modeling and risk analysis of natural disasters, and applications of mathematics and statistics in sports. He is a fellow of the Alexander von Humboldt. He frequently teaches courses based on sports. His outside interests include soccer, racquetball, and tennis.
Department of Mathematics, Bloomsburg University, Bloomsburg, PA 17815, USA
rnoubary@bloomu.edu.