Statistical Insights into Striker Selection: Maximizing Leeds United's Performance with the Poisson Distribution

Determining the optimal combination of 3 stickers for Leeds United to increase their probability of winning and points per game using a Poisson Distribution

Pages: 19

Table of Contents:

1. Introduction

Leeds United, a football team based in Leeds, England, has been relegated from the Premier League (England's best league) at the end of the 22/23 season. The motivation behind this mathematical exploration was sparked by a debate I had with a group of my friends, this discussion was regarding what players Leeds should have bought at the end of the 21/22 season to prevent them from being relegated and improve their probability of winning games. We essentially said that the current players of the Leeds squad weren't good enough to prevent Leeds from getting relegated. To determine what players Leeds should have bought at the beginning of the 21/22 season I chose to focus on the offensive side of the Leeds team as it is the easiest to quantify and compare. When playing a football match both teams are allowed to have 11 players each on the field at one given moment, and these 11 players can be organised into many different formations where each player has a role and job, it is also important to note that formations can change based on whether the team is attacking or defending and where the ball is on the field. Since only the offensive side of Leeds is being considered the formation Leeds uses when in an attacking position must be regarded, the 4-3-3 formation, which can be represented by Figure 1.

Figure 1: Image representing Leeds's formation when they have the ball and are attacking.[1]

[1]Desmond, Rhys. "4-3-3 vs. 4-1-4-1: Tactical Flexibility." TheMastermindSite, TMS Football and Coaching, 30 Nov. 2020, themastermindsite.com/2019/04/10/4-3-3-vs-4-1-4-1-tactical-flexibility/.

The figure above shows all the positions for all the players in the Leeds team when they are attacking the goalkeeper is denoted by GK, the 4 defensive players are denoted by either CB, LB or RB (centre, left and right back respectively), the midfielders are denoted by CM and DM (central and defensive midfielder respectively), and the attacking players are denoted by LW, RW, and ST (left winger, right winger and striker respectively). It will assumed that the goals scored by Leeds United are solely dependent on the 3 attacking players: LW, ST, and RW, (this will be explained further when discussing the players). By altering Leeds's current attacking players we can, in theory, change how often goals are scored, and how many goals are scored per game, and hopefully increase the probability of victory, which will then affect their league position. A team's league position is determined by their performance in 38 games played, against other teams in the league, throughout the year, if a team wins a match they earn 3 points, if they draw they earn 1, and if they lose they earn 0. Overall the team with the most points at the end of the year win the league or championship. Finally, teams are allowed to buy players off other teams during both the winter and summer break, so this exploration aims to: Determine the optimal combination of 3 stickers for Leeds United to increase their probability of winning and points per game.

2. Methodology

To find out what combination of players will help Leeds gain the highest points per game ratio the probabilities that Leeds wins, draws and loses with these players must be calculated. To find these probabilities a probability tree will be constructed, this probability tree will first consider the possibilities that Leeds concedes (lets in) N number of goals, and then it will consider the probability that Leeds then scores more than, less than or equal to N number of goals by considering the possible combinations of goals scored by that attacking combination. To find these probabilities the mean number of goals scored by a player and goals conceded by Leeds will have to be calculated, using online data, then this value will be inputted into a Poisson distribution which will give the probability that a player scores N goals or that Leeds concedes N goals. After having constructed the probability tree using the technique mentioned previously, the probability that Leeds wins, loses or draws will then be calculated by multiplying the cases in which Leeds wins, loses, or draws in each respective branch, and then those numbers will be summed together, therefore giving a final probability that Leeds wins, loses or draws. Then using these probabilities a point-per-game ratio will be calculated that will make it possible to compare the various combinations of attacking players.

3. Poisson Distribution

The Poisson distribution is a probability distribution that uses the following formula[2]:

$$P(X = x) = \frac{\lambda^{x} \times e^{-\lambda}}{x!}$$

This gives the probability that an event will happen x times in a given period, where the parameter Lambda (λ) represents the mean number of events, for example, the number of goals scored in 90 minutes. The Poisson distribution has the following assumptions:

- Each goal is independent, which means that the occurrence of one goal does not affect the probability of another goal occurring meaning that to find the probability that 2 goals were to both occur, their respective probabilities would have to be multiplied by each other
- The rate of each goal occurring is constant, so for example, if 3 goals are scored they will occur every 30 minutes, in mathematical terms this means that the mean and the variance are equal
- No goal can occur simultaneously (even though two goals can't occur at the same time as there is only one ball in play at a time, it is still something to consider)

It is possible to derive the Poisson distribution from the binomial distribution. Initially, it is considered that the probability that a player scores X goals in a game can be found using a binomial distribution this is because it is possible to consider the probability that a player scores as the probability of success and the number of trials can be interpreted as the amount of time a player has in one game to score. Meaning the probability that a player scores x goals can be denoted as the random variable X

[2] "Introduction to Poisson Distribution - Probability & Statistics." *YouTube*, YouTube, 16 June 2019, https://www.youtube.com/watch?v=m0o-585xwW0&pp=ygUUcG9pc3NvbiBkaXN0cmlidXRpb24%3D. Accessed 21 Sept. 2023.

and the number of trials is 90, representing the minutes in a game. Which can be made more accurate using the Poisson distribution which can be derived from the binomial distribution by considering the parameter of the Poisson distribution(λ) is equivalent to the mean, also known as the expected value, used in the binomial distribution. This relationship can be expressed as follows:

$$E(X) = n \times p$$

where n is the number of trials, for example, the time in a football match, p is the probability of x number of goals occurring per 90 minutes, and E(x) is the number of goals scored per 90 minutes. So we can say that λ is equal to the goals per 90 minutes, the number of trials is 90 minutes because that is how long a football match is and that is the number of possibilities in which a football player can score a goal, and the probability of success is λ over 90 which is the ratio of goals per 90 minutes in a game. This would lead to the following binomial distribution

$$P(X = x) = {}^{90}C_x \times (\frac{\lambda}{90})^x \times (1 - (\frac{\lambda}{90}))^{90-x}$$

However, the main problem with this binomial distribution is that it doesn't account for the fact that more than one goal can be scored per minute, and although it is possible to convert the minutes into seconds it still may not account for the very rare possibility of more than one goal being scored in that interval. This problem could be carried on to any interval (potentially not in football but in other instances, it could affect its reliability), to address this the limits to infinity for the intervals must be considered, which will now be denoted under the variable n, leading to the following distribution and simplification

$$P(X = x) = \lim_{n \to \infty} ({}^{n}C_x \times (\frac{\lambda}{n})^x \times (1 - (\frac{\lambda}{n}))^{n-x})$$

$$P(X = x) = \lim_{n \to \infty} (\frac{n!}{(n-x)! \times x!} \times \frac{\lambda^x}{n^x} \times (1 - (\frac{\lambda}{n}))^n \times (1 - (\frac{\lambda}{n}))^{-x})$$

$$P(X = x) = \lim_{n \to \infty} (\frac{n \times (n-1) \cdots (n-x+1)}{n^x} \times \frac{\lambda^x}{x!} \times (1 - (\frac{\lambda}{n}))^n \times ((1 - (\frac{\lambda}{n}))^{-x})$$

$$P(X = x) = \frac{\lambda^x}{x!} \times \lim_{n \to \infty} ((1 - (\frac{\lambda}{n}))^n \times (1 - (\frac{\lambda}{n})^{-x}))$$

The reason why $\lim_{n \to \infty} (\frac{n \times (n-1) \cdots (n-x+1)}{n^x})$ became 1 was that when the limit of that fraction is evaluated there would be an xth-degree polynomial in the numerator. Then the polynomial is separated into different fractions ($\frac{n^x}{n^x} + \frac{\alpha_1 \times n^{x-1}}{n^x} + ... + \frac{\alpha_x}{n^x}$(where all α's are constants)) the first fraction will be 1 as it is n^x divided by itself, and afterwards, it would be a sum of numbers that would just keep getting infinitely smaller due to the limit causing its simplification to become 1.

Also since it is known that,

$$\lim_{n \to \infty} (1 - (\frac{\lambda}{n})^n) = e^{-\lambda}$$

and that

$$\lim_{n \to \infty} (\frac{\lambda}{n}) = 0$$

the equation can be simplified into the following:

$$P(X = x) = e^{-\lambda} \times (1 - 0)^{-x} \times \frac{\lambda^x}{x!}$$

Which then leads to this final equation, otherwise known as the general formula for the Poisson distribution.

$$P(X = x) = \frac{\lambda^x}{x!} \times e^{-\lambda}$$

The reason why the Poisson distribution was chosen for this project specifically was because it could calculate the probability that a specific event happened x number of times, this made it a natural fit for a project discussing football, as it can consider the scoring or conceding of a goal as an event. The original idea was to use the binomial distribution but after having discovered the Poisson distribution it was chosen over the binomial due to its accuracy.

4. Leeds United

The first step is to determine the current probability that Leeds United will win, draw or lose, this will be done by considering how many times Leeds has won, drawn and lost during the 21/22 season.

After consulting the official premier league website[3] Leeds's league performance over 38 games can be represented using Table 1. It was chosen to round the probabilities to 3 decimal places for Leeds United as it was felt that it was the most appropriate way to present these probabilities since they were still relatively accurate while being clear and readable.

Outcome	Frequency	Probability (to 3 decimal places)
Win	9	$\frac{9}{38} = 0.237$
Draw	11	$\frac{11}{38} = 0.289$
Lose	18	$\frac{18}{38} = 0.474$

Table 1: Table showing Leeds United's league performance over 38 games in the 21/22 season.

The next step is to find how many goals Leeds concedes (let in), and the probability of this happening, which can be done using a Poisson distribution. After consulting the stats section of the premier league website it was determined that the average amount of goals conceded by Leeds during the 21/22 season is 2.08, leading to the following Poisson distribution:

$$P(X = x_x) = \frac{2.08^x \times e^{-2.08}}{x!}$$

The results obtained from using the formula above are presented in Table 2.

[3]"Official Website." *Leeds United*, Leeds United, www.leedsunited.com/. Accessed 21 Sept. 2023.

Goals Conceded	Poisson Probabilities (3 dp)
0	0.125
1	0.260
2	0.270
3	0.187
4	0.097
5	0.041
6	0.014
7	0.004

Table 2: A table representing the goals conceded by Leeds, and the respective probability during the 21/22 season.

It will be assumed that the mean amount of goals Leeds United concede next season will remain unchanged since only the teams' attack will be modified. But to ensure that this exploration provides an objective result it would be prudent to exclude any cases where Leeds conceded more than 4 goals as the events outside this range only occurred once, and will be considered as anomalies. So the probabilities will be normalised to ensure that the sum of the probabilities that Leed concedes x number of goals (assuming $x \in \mathbb{Z} \cap [0, 4]$) is equal to 1. The normalised probabilities of Leeds conceding x number of goals in 90 minutes is represented by Table 3:

Goals Conceded	New Poisson probability (3 decimal places)
0	0.133
1	0.276
2	0.288
3	0.199
4	0.104

Table 3: Table representing the regulated probabilities that Leeds United Concedes x goals in 90 minutes.

Some of the limitations that come with the acquisition of these values include the fact that each minute is not equal, as generally, the tempo of the game increases over time, and that a football match is not continuous due to the break at half-time. Additionally, it is also somewhat incorrect to assume that Leeds will concede no more than 4 goals, but for simplicity and considering that it is incredibly rare that a team concedes more than 4 goals for the rest of this report it will be assumed Leeds don't concede more than 4 goals.

5. The Players

Now it's time to discuss the three attacking players that Leeds should have signed during the 21/22 summer transfer window, as discussed previously there are 3 attacking positions and while it is not uncommon that players sometimes play in multiple positions (for example Cristiano Ronaldo has played both LW and ST) it will be assumed that the players mentioned will only play in the position assigned to them by the Transfermarkt website[4] and that each player mentioned throughout this exploration will be out of contract. Meaning that if Leeds did want to buy this player they would have been willing and able to leave their club to join Leeds and that they are all from a premier league club

[4] "Transfermarkt" *Transfermarkt*, Axel Springer SE, www.transfermarkt.com . Accessed 21 Sept. 2023.

(at the end of the 21/22 season). Additionally, it will also be assumed that the transfer price will be determined by the market value assigned to them through the Transfermarkt website and that there will not be any extra fees added to the acquisition of the player for Leeds. Finally, for simplicity, this exploration will only consider 7 total players: 3 strikers, 2 left and 2 right-wingers. The 3 strikers that will be considered for this exploration are Cristiano Ronaldo, Jaime Vardy, and Harry Kane. The 2 left-wingers are Heung Min Son and Raheem Sterling. The 2 right-wingers are Mohamed Salah and Riyad Mahrez. In general and in this mathematical exploration football players are referred to using their last names, for example, Heung Min Son will be referred to as Son, Riyad Mahrez as Mahrez, and so forth. To determine the best combination of football players for Leeds to purchase the football player's respective probabilities of scoring or assisting x number of goals per game must first be determined, to make it possible to calculate the probability of Leeds winning, losing and drawing to therefore achieve the aim of this exploration. To find these probabilities a Poisson distribution must once again be used, but first, the mean number of goals and assists provided by the players must be procured. An assist in football is a cross or pass made by a player which then leads to a goal. The reason why assists will also be looked at is to cancel out any goals made by non-attacking players, this will be done by assuming that if anyone who is not an attacking player scores a goal it will be a direct result of an attacking player's assist. Goals and assists may also be written in the form G/A. This will then account for all the goals scored by Leeds throughout each game of the season. To calculate the mean G/A of a player, the premier league website[5] will be visited, to check 3 of the player's statistics, their assists, goals and appearances. Appearances are the number of times a player has been part of a game, regardless of whether or not they came on or off as a substitute or were in the starting 11. It will be assumed that an appearance is of equal length for all the games and for every player, then the sum of the goals and assists will be divided by their appearances. This will give the player's average G/A per game which will be used as the λ value. For example, in the case of Mohamed Salah[6], it was found

[5] "Premier League Football News, Fixtures, Scores & Results." *Premier League*, www.premierleague.com/home. Accessed 21 Sept. 2023.

[6] "Premier League Football News, Fixtures, Scores & Results." Premier League, www.premierleague.com/home. Accessed 21 Sept. 2023.

that he had 23 goals and 13 assists in 35 appearances during the 21/22 season. Then his goals and assists were summed to obtain 36 G/A and then divided by 35, and when it is rounded to 3 decimal places it provides a mean of 1.03 goals and assists per game. Then the respective mean (or lambda value) is applied to a Poisson distribution to find the probability of each player scoring and assisting x number of goals each game. Continuing with the example of Mohamed Salah his respective Poisson distribution can be expressed using the following formula:

$$P(Salah\ scores\ and\ assists\ x\ goals\ per\ game) = \frac{1.03^{x} \times e^{-1.03}}{x!}$$

If the Poisson distribution is used on its own it may not yield accurate and reliable data, because in theory Salah could potentially have 100 goals and assists in one game. But to ensure this test was fair Salah's highest G/A value in one premier league game was checked, which was a G/A of 4 against Manchester United where he scored 3 goals and assisted 1. Salah’s G/A of 4 was the highest G/A per game recorded amongst all the 7 potential players. This then means that similarly to the Leeds Poisson Salah’s Poisson distribution will be capped at four and then the probabilities will be normalised accordingly to ensure they sum up to 1. After having carried out the Poisson distribution using the formula above the following table can be constructed.

G/A per game	P(G/A per game = x)(to 3 decimal places)
0	0.357
1	0.368
2	0.189
3	0.065
4	0.017

Table 4: Table representing the probability of Mohamed Salah scoring or assisting x goals, derived from a Poisson distribution.

After having obtained these probabilities, the next step is to sum them:

0.357 + 0.368 + 0.189 + 0.065 + 0.017 = 0.996

Then we must divide each probability from Table 4 by 0.996 making it possible to create the following table:

G/A per game	P(G/A per game = x)(to 3 decimal places)
0	$\frac{0.357}{0.996} = 0.358$
1	$\frac{0.368}{0.996} = 0.369$
2	$\frac{0.189}{0.996} = 0.190$
3	$\frac{0.065}{0.996} = 0.065$
4	$\frac{0.017}{0.996} = 0.017$

Table 5: Table representing the adjusted probability of Mohamed Salah scoring or assisting x goals. (Appendix A)

After summing the following values the final result is 0.999, which can then be rounded to one. The reason why the values sum to 0.999 and not exactly one is due to the choice to use 3 decimal places to represent all the probabilities for this exploration and to follow a uniform exploration the same will be done with Mohamed Salah too. Since relatively small numbers like 0.189, 0.065, and 0.017 were being divided by another number that is relatively close to 1 (0.996) the changes made to them were so minuscule that they did not affect the overall 3 decimal point value. Hence why 0.065 and 0.017 remained constant after the division. These steps were then repeated for all the other 6 players (Appendix A).

Now that all of the players' probability of having x G/A per game has been calculated it is possible to then move on to find the best possible combination of attacking players for Leeds United. As seen previously there are 3 attacking positions namely: LW, RW and ST otherwise known as left-winger, right-winger and striker respectively. Of the 7 potential players, there are 2 right-wingers, 2 left-wingers and 3 strikers, it then can be deduced that there are 12 different combinations, and that number is derived using the following thought process:

It can be said that there are 2 players to play RW, 2 to play LW, and 3 to play ST, then it can be imagined that when one RW remains constant he can play with 3 different strikers and 2 different left-wingers, amounting to 6 different possibilities. Consequently, each possibility with the first RW can be done with the other RW, amounting to another 6 possibilities. After summing the two quantities of possibilities total of 12 possible combinations is found. Now all of the possible combinations must be listed, this can be done in several ways, but it was chosen to do it manually and the different combinations are listed below:

1. Sterling - Kane - Salah
2. Son - Kane - Salah
3. Son - Kane - Mahrez
4. Sterling - Kane - Mahrez
5. Sterling - Ronaldo - Mahrez
6. Son - Ronaldo - Mahrez
7. Son - Ronaldo - Salah
8. Sterling - Ronaldo - Salah
9. Sterling - Vardy - Salah
10. Son - Vardy - Salah
11. Sterling - Vardy - Mahrez
12. Son - Vardy - Mahrez

Next, their respective probabilities that Leeds, win, lose or draw must be calculated. This can be achieved using a probability tree, as Leeds can concede 0 to 4 goals and the combinations of attacking players have to score more than what Leeds concede to win. To help carry out the calculations quickly Google Sheets were used through the creation of a general system that would calculate the probability of winning, drawing, and losing for any combination of attacking players. For example let's consider the combination number 8: Sterling - Rondalo - Salah who have the following probability distributions:

	A	B	C	D
1	0.533	0.499	0.358	P(G/A) = 0
2	0.337	0.35	0.369	P(G/A) = 1
3	0.107	0.122	0.19	P(G/A) = 2
4	0.023	0.029	0.065	P(G/A) = 3
5	0	0	0.017	P(G/A) = 4
6	Sterling	Ronaldo	Salah	

Table 6: Image showing the probability distributions of combination 8: Sterling - Ronaldo - Salah.

To find the probability that Leeds wins the different amounts of goals that Leeds concedes must be considered, as shown in Table 3 the probability that Leeds concede 0 goals is 0.133. If Leeds doesn't concede any goals then there can only be 2 outcomes, a draw (if Leeds scores 0 goals) or a victory (if Leeds scores more than 0 goals), and the only way Leeds can score 0 goals is the 3 attacking players all have a G/A of 0. To find the probability that every attacking player has a G/A of 0 the product of the probability that each player has a G/A of 0 must be taken since it is assumed that each possibility is independent, leading to the following operation:

$0.533 \times 0.499 \times 0.358 = 0.095$ (to 3 decimal places)

Therefore the probability that Leeds score more than 1 goal is:

$1 - 0.095 = 0.905$ (to 3 decimal places)

It is important to consider that the previous two probabilities do not provide the probability that Leeds win or draw. To find that the product of the two probabilities and 0.133 (which is the probability that Leeds concedes 0) must be taken to therefore find the probability that Leeds wins or draws given that they concede no goals.

The following probabilities, for this attacking combination, of Leeds winning, drawing or losing given that they concede 1, 2, 3 or 4 goals can be found in Appendix B, but to understand how these results the possibility that Leeds concedes 4 goals will be used to demonstrate the thought process behind the retrieval of these numbers.

First, each player is given a letter (A, B, C) the letter is given depending on their position in the combination, Sterling is an LW so he gets the letter A, Ronaldo is an ST so he gets the letter B and

Salah is an RW so he gets the letter C. This format is used for all the players and all the combinations found in Appendix B. Additionally each player's respective probability of scoring and assisting x number of goals per 90 minutes retrieved from their Poisson distribution is labelled as well. For example, the probability that Sterling has a G/A of 3 is found in cell A4 and is 0.023, this format is used throughout the entire IA and can all be found in Appendix B. To calculate the probability that Leeds scores strictly less than 4 goals, the probabilities that were calculated previously that this combination scores exactly 0, 1, 2, and 3 goals can be used which when summed amount to 0.796 (see Appendix B). Next, the probability that exactly 4 goals must be calculated, to do this the different ways 4 goals can be scored must be considered. For example, one of the attacking players could have a G/A of 4 and the rest could have a G/A of 0, another possibility is that 2 players have a G/A of 1 and the remaining player has a G/A of 2, it is also possible that two players have a G/A of 2 and the remaining one has a G/A of 0 or that one player has a G/A of 1, another has a G/A of 3 and the last one has a G/A of 0. Then for each way 4 goals can be scored the players will shuffle around to obtain the following expression, which will be expressed using the spreadsheet cells above:

$$= (A_5 \times B_1 \times C_1) + (A_1 \times B_5 \times C_1) + (A_1 \times B_1 \times C_5) + (A_2 \times B_2 \times C_3) + (A_2 \times B_3 \times C_2) + (A_3 \times B_2 \times C_2) + (A_3 \times B_1 \times C_3) + (A_3 \times B_3 \times C_1) + (A_1 \times B_3 \times C_3) + (A_4 \times B_2 \times C_1) + (A_4 \times B_1 \times C_2) + (A_1 \times B_4 \times C_2) + (A_1 \times B_2 \times C_4) + (A_2 \times B_1 \times C_4) + (A_2 \times B_4 \times C_1)$$

After inputting this formula in the spreadsheet the number obtained will be 0.075. Then using the complementary events rule the probability that this combination will score more than 4 goals is 0.03 when rounded to 2 decimal places. These steps are then repeated to find all the other probabilities for each combination which can be found in Appendix B. After finding all the probabilities the following probability tree can be constructed:

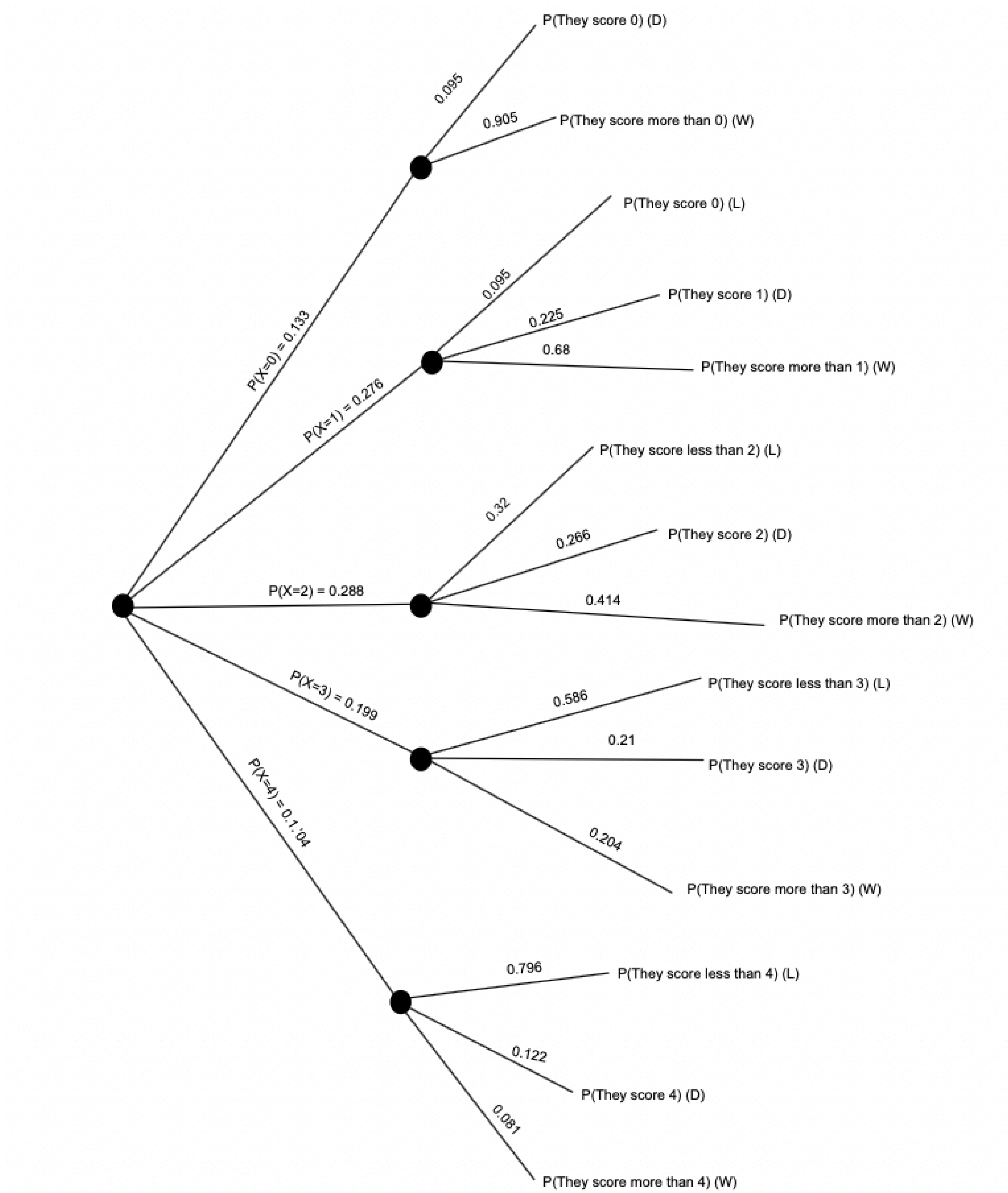


Figure 2: probability tree representing the possible match outcomes and their probabilities of combination 8. Sterling - Ronaldo - Salah

The probability tree can be used to calculate the probability of winning: which can be done by finding the probability of victory after having conceded 0, 1, 2, 3 or 4 goals, and then summing the probabilities to give one final probability of victory. This has been carried out in the following operation:

$(0.133 \times 0.905) + (0.276 \times 0.68) + (0.288 \times 0.414) + (0.199 \times 0.204) + (0.104 \times (0.081)$

Which yields a 0.476 probability of victory. After applying the same concept the final probability of Leeds drawing with this combination is 0.206. Then the probability of Leeds losing would be 0.318. Should Leed purchase this combination of players, their expected points per game would be 1.634 when rounded to 3 decimal places which is higher than their previous point-per-game ratio of 1, derived from Table 1. The rest of the probability trees can be found in Appendix B.

5. Final findings

After having run the other 11 combinations through the "algorithm" the probability of winning, losing and drawing (Appendix B) has been determined and therefore the projected points-per-game of each combination has been calculated. All of which are higher than Leeds's original points-per-game of 1.009. The order is listed below (all values are rounded to 3 decimal places):

7. Son - Ronaldo - Salah: 1.752 points-per-game

10. Son - Vardy - Salah: 1.710 points-per-game

2. Son - Kane - Salah: 1.708 points-per-game

8. Sterling - Ronaldo - Salah: 1.634 points-per-game

9. Sterling - Vardy - Salah: 1.591 points-per-game

1. Sterling - Kane - Salah: 1.589 points-per-game

6. Son - Ronaldo - Mahrez: 1.460 points-per-game

12. Son - Vardy - Mahrez: 1.413 points-per-game

3. Son - Kane - Mahrez: 1.411 points-per-game

5. Sterling - Ronaldo - Mahrez: 1.330 points-per-game

11. Sterling - Vardy - Mahrez: 1.280 points-per-game

4. Sterling - Kane - Mahrez: 1.278 points-per-game

The next aspect to consider is the price of each combination and the budget of Leeds United: Assuming that Leeds didn't purchase any player during the 22/23 summer window[7] (the summer before the 22/23 season starts), Leeds would have a budget of 151.39 million pounds. To obtain this number Transfermarkt website was visited to check the total money spent during the 22/23 summer window and assumed that they would be willing to spend as much on the three attacking players if they didn't spend it on other players. Each player investigated is assigned a value in pounds, this is the list of players and their corresponding value:

Sterling - £70 Million

Son - £75 Million

Kane - £90 Million

Vardy - £7 Million

Ronaldo - £30 Million

Salah - £90 Million

Mahrez - £35 Million

Of the 12 combinations, the top 6 had one player in common, Mohamed Salah, but due to his £90 million price tag Leeds, unfortunately, wouldn't be able to afford him in any combination. Another interesting observation is the fact that despite the £87 million discrepancy in price between Harry Kane and Jamie Vardy, Vardy has a higher mean G/A meaning that even at a cheaper price Vardy provides more "value" to Leeds than Kane. But overall considering the price of each combination the one with the highest projected points-per-game, while considering price is combination number 6. Son - Ronaldo - Mahrez at a total price of £140 million. With a 0.414 chance of victory, a 0.218 chance of drawing and a 0.368 chance of losing in a single game yielding a projected 1.460 points per game.

7 "Leeds United - Transfers 22/23." *Transfermarkt*, Axel Springer SE, www.transfermarkt.co.uk/leeds-united/transfers/verein/399/plus/?saison_id=2022&pos=Sturm&detailpos=&w_s=. Accessed 21 Sept. 2023.

6. Conclusion

Although in theory, the signing of Combination 6 would benefit Leeds in terms of points per game, it isn't always the case in real life. For one, football players are human, just because they performed well in one season doesn't mean they will be in good form for the next. Since football is a team sport there may be some problems with the chemistry of the team, as the players have to adapt to each other and the rest of the Leeds squad too. For example in the worst 4 combinations, 2 of the 3 players have been in a team before which means they may already be used to the others' playstyle and may even be more effective. As the data collected is only quantitative and not qualitative. Additionally, it may also be important to mention that football players are on contracts meaning that Leeds would have to pay extra money every year to ensure that these players stay at their club which may not be possible for such a small team. Then there's the added possibility of these players not even wanting to play at Leeds since it doesn't compete internationally and rarely wins any trophies. Finally, it's also worth mentioning that when the players listed above obtained their respective mean G/A they were (except for Vardy and maybe Kane and Son) with teams that are full of excellent players and these players may have had a part to play with these numbers. For example Sterling and Mahrez were both part of the title-winning team during the 21/22 season, filled with some of the best players of the decade such as Kevin De Bryne known for his incredible ability to assist and create opportunities. This may have caused their mean G/A to be inflated due to the quality of players they had with them. Which may then cause this result to be somewhat inaccurate. This was incredibly different from what was discussed as we considered foreign players but we also underestimated the purchasing power of Leeds United and assumed that they wouldn't be able to afford any players, additionally, we also believed that any players worth mentioning wouldn't want to move to Leeds due to its lack of international and local competitiveness.

Additionally, since the Poisson distribution assumes that each event is independent of another which is completely untrue in football, this is caused by the emotional, mental, and humane side of football, for example, after scoring a goal it could be postulated that overall player morale would ameliorate which may make them more confident and more likely to score another goal. In contrast, it could also be said that after conceding a goal team morale would worsen therefore they may also become more prone to making mistakes which may cause them to concede another goal. Additionally, it should also be mentioned that throughout a match, players may feel fatigue, and there may also be changes in team dynamics which can then influence the performance of a player. Therefore it can be said that because the Poisson distribution assumes independence it cannot provide a completely reliable result, in addition to this since it is also assumed that each game is independent this methodology cannot consider whether or not a team or player is having a run of good or bad performance. Another factor hindering the reliability of this exploratory analysis is the fact that it only considers one team, even though it is assumed that each game is independent and that each game will be equally difficult this is not true, even if only the offensive side of the team was to be considered. For example, should Leeds play against a team that has a stronger defence it will be relatively harder for the attacking players to score, but the probability of conceding will remain the same, ceteris paribus. Therefore it is not completely realistic to assume that each game is independent and to only analyse 1 team, to make the study more reliable it would be better to use the same methodology to analyse the defensive and offensive side of each team in the league and to compare them individually.

All in all, despite all these limitations from a purely mathematical standpoint and with all the assumptions made previously, the best combination of attacking players is Heung Min Son, Cristiano Ronaldo and Riyad Mahrez.

Bibliography

1. “Introduction to Poisson Distribution - Probability & Statistics.” *YouTube*, YouTube, 16 June 2019, https://www.youtube.com/watch?v=m0o-585xwW0&pp=ygUUcG9pc3NvbiBkaXN0cmlidXRpb24%3D. Accessed 21 Sept. 2023.
2. Khan, Sal, director. *Poisson Process 1 | Probability and Statistics | Khan Academy*. *YouTube*, YouTube, 1 Mar. 2009, https://www.youtube.com/watch?v=3z-M6sbGIZ0&pp=ygUacG9pc3NvbiBkaXN0cmlidXRpb24ga2hhbiA%3D. Accessed 21 Sept. 2023.
3. Khan, Sal, director. *Poisson Process 2 | Probability and Statistics | Khan Academy*. *YouTube*, YouTube, 1 Mar. 2009, https://www.youtube.com/watch?v=Jkr4FSrNEVY&pp=ygUacG9pc3NvbiBkaXN0cmlidXRpb24ga2hhbiA%3D. Accessed 21 Sept. 2023.
4. “Leeds United - Transfers 22/23.” *Transfermarkt*, Axel Springer SE, www.transfermarkt.co.uk/leeds-united/transfers/verein/399/plus/?saison_id=2022&pos=Sturm&detailpos=&w_s=. Accessed 21 Sept. 2023.
5. “Transfermarkt” *Transfermarkt*, Axel Springer SE, www.transfermarkt.com . Accessed 21 Sept. 2023.
6. “Official Website.” *Leeds United*, Leeds United, www.leedsunited.com/. Accessed 21 Sept. 2023.
7. “Premier League Football News, Fixtures, Scores & Results.” *Premier League*, www.premierleague.com/home. Accessed 21 Sept. 2023.
8. Desmond, Rhys. “4-3-3 vs. 4-1-4-1: Tactical Flexibility.” *TheMastermindSite*, TMS Football and Coaching, 30 Nov. 2020, themastermindsite.com/2019/04/10/4-3-3-vs-4-1-4-1-tactical-flexibility/.

Appendix

Appendix A

Right Wingers:

MOHAMED SALAH	RW £90 million	Adjusted Poisson
G/A per game	Number of occurrences	N/A
0	11	0.358
1	17	0.369
2	3	0.19
3	3	0.065
4	1	0.017
Average G/A per game	1.03	N/A

Riyad Mahrez	RW £35 Million	Adjusted Poisson
G/A per game	N of occurrences	N/A
0	17	0.577
1	6	0.329
2	5	0.094
Average G/A per game	0.571	N/A

Strikers:

Jamie Vardy	ST £7 Million	Adjusted Poisson
G/A per game	N of occurrences	N/A
0	14	0.523
1	5	0.356
2	6	0.121
Avg 0.68		N/A

Cristiano Ronaldo	ST £30 Million	Adjusted Poisson
G/A per game	N of occurrences	N/A
0	17	0.499
1	7	0.35
2	4	0.122
3	2	0.029
Average G/A per game	0.7	N/A

Harry Kane	ST £90 Million	Adjusted Poisson
G/A per game	N of occurrences	N/A
0	20	0.525
1	9	0.355
2	8	0.12
Average G/A per game	0.677	N/A

Left Wingers:

Hueng Min Son	LW £75 Million	Adjusted Poisson
G/A per game	N of occurrences	N/A
0	14	0.429
1	14	0.368
2	5	0.158
3	2	0.045
Average G/A per game	0.857	N/A

Raheem Sterling	LW £70 Million	Adjusted Poisson
G/A per game	N of occurrences	N/A
0	17	0.533
1	8	0.337
2	4	0.107
3	1	0.023
Average G/A per game	0.633	N/A

Appendix B

7. Son - Ronaldo - Salah: 67 points

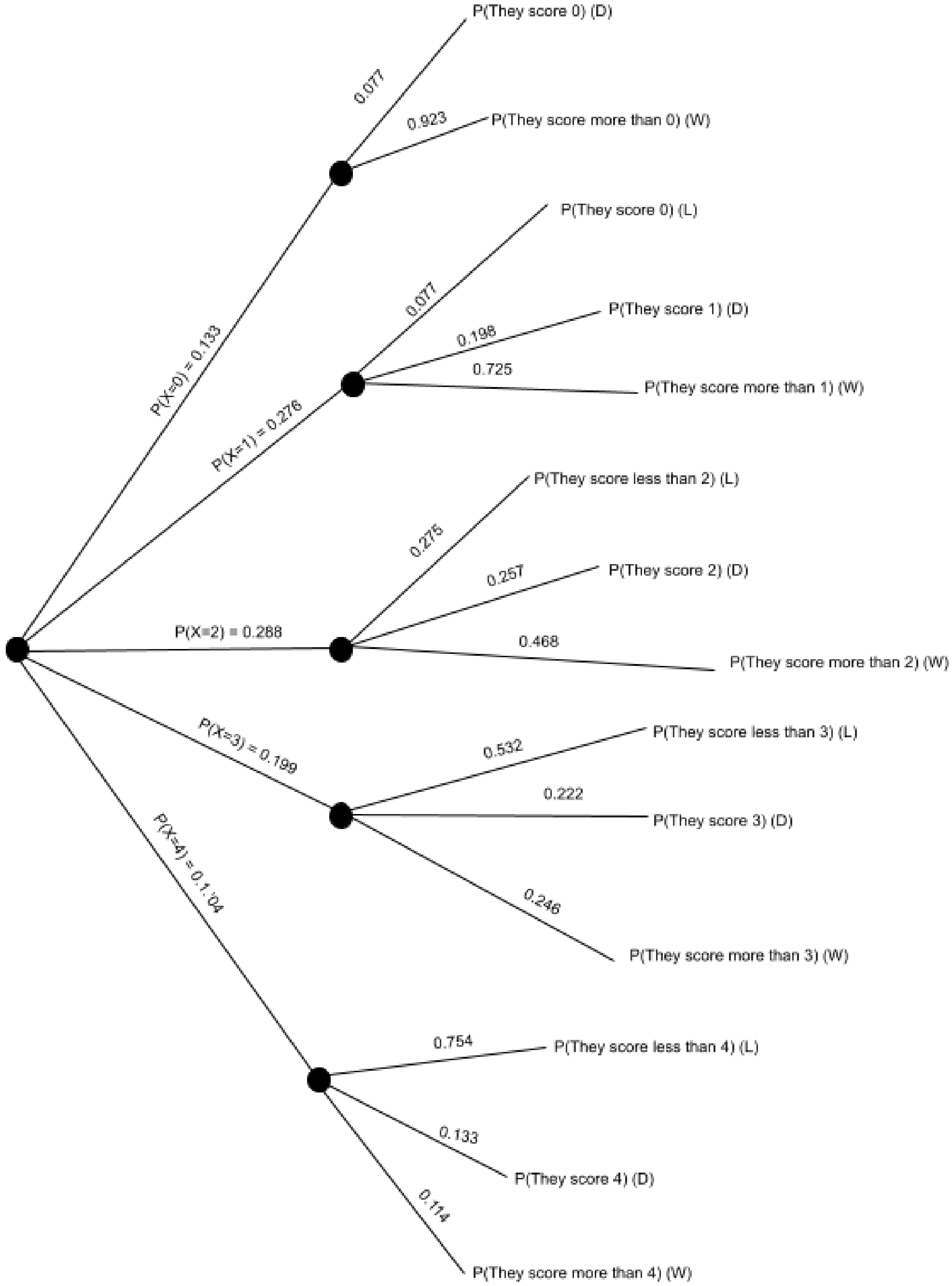
Base Probabilities:

	A	B	C	D
1	0.429	0.499	0.358	P(G/A = 0)
2	0.368	0.35	0.369	P(G/A = 1)
3	0.158	0.122	0.19	P(G/A = 2)
4	0.045	0.029	0.065	P(G/A = 3)
5	0	0	0.017	P(G/A = 4)
6	Son	Ronaldo	Salah	

The probability that this combination wins, loses or draws and projected points:

P(WIN) =>	0.5185036568
P(DRAW) =>	0.196852692
P(LOSE) =>	0.2846436512
Expected Points	1.752363662

Probability Tree:



10. Son - Vardy - Salah: 65 points

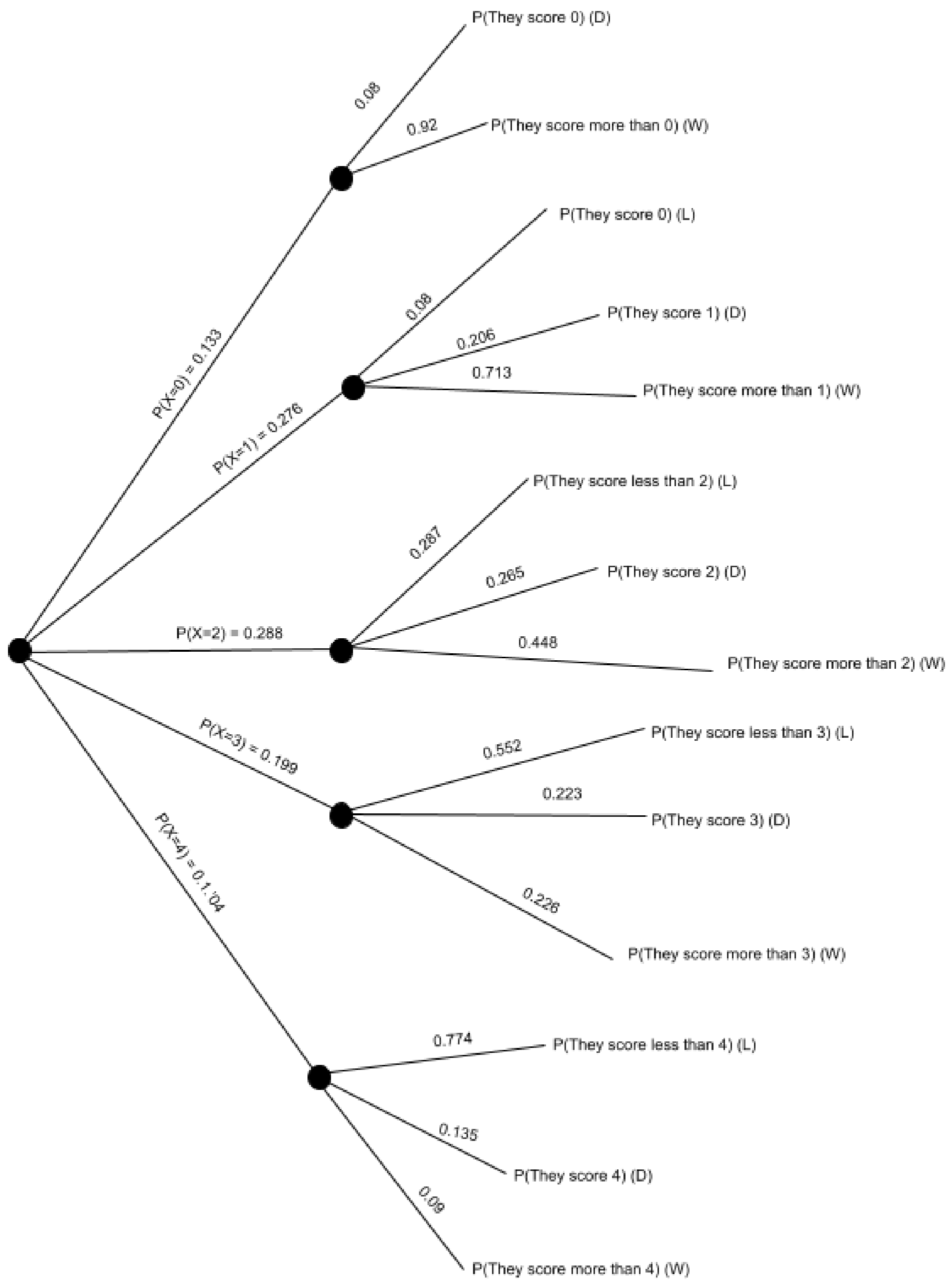
Base Probabilities:

	A	B	C	D
1	0.429	0.523	0.358	P(G/A) = 0
2	0.368	0.356	0.369	P(G/A) = 1
3	0.158	0.121	0.19	P(G/A) = 2
4	0.045	0	0.065	P(G/A) = 3
5	0	0	0.017	P(G/A) = 4
6	Son	Vardy	Salah	

The probability that this combination wins, loses or draws and projected points:

Expected Points	1.710072605
P(WIN)	0.5025751442
P(DRAW)	0.2023471723
P(LOSE)	0.2950776835

Probability Tree:



2. Son - Kane - Salah: 65 points

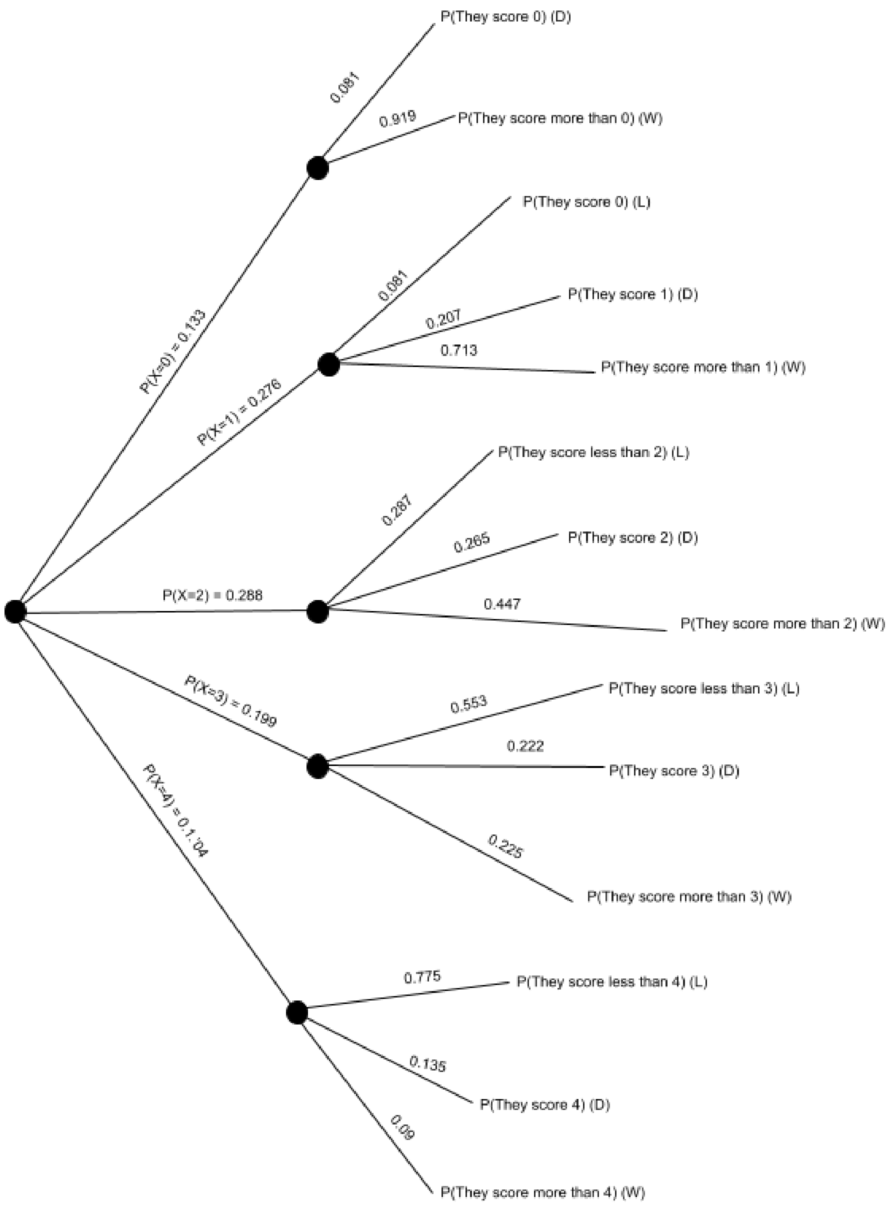
Base Probabilities:

	A	B	C	D
1	0.429	0.525	0.358	P(G/A) = 0
2	0.368	0.355	0.369	P(G/A) = 1
3	0.158	0.12	0.19	P(G/A) = 2
4	0.045	0	0.065	P(G/A) = 3
5	0	0	0.017	P(G/A) = 4
6	Son	Kane	Salah	

The probability that this combination wins, loses or draws and projected points:

Expected Points	1.708261972
P(WIN)	0.5019333756
P(DRAW)	0.2024618453
P(LOSE)	0.2956047792

Probability Tree:



8. Sterling - Ronaldo - Salah: 62 points

Base Probabilities:

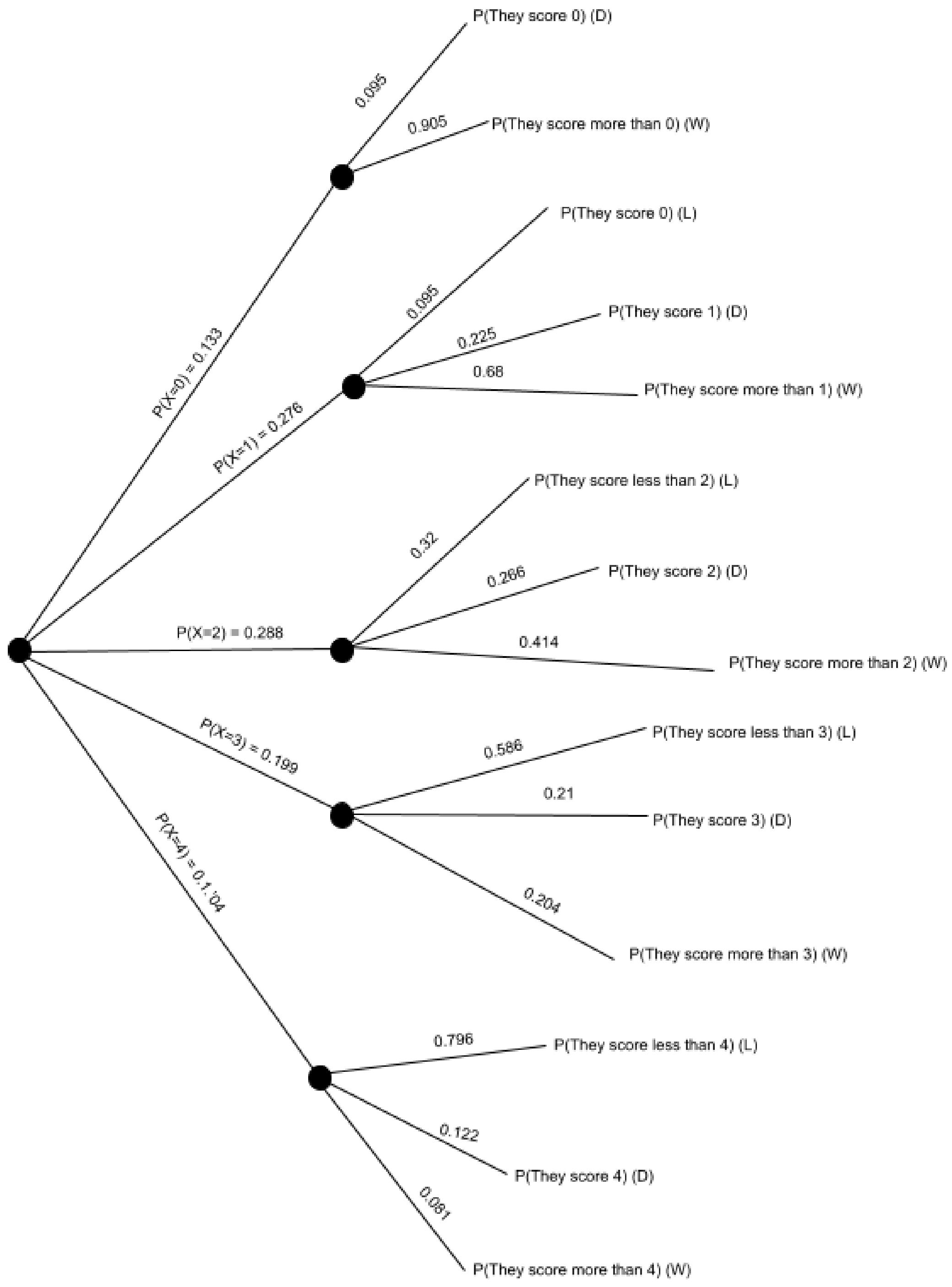
	A	B	C	D
1	0.533	0.499	0.358	P(G/A) = 0
2	0.337	0.35	0.369	P(G/A) = 1
3	0.107	0.122	0.19	P(G/A) = 2
4	0.023	0.029	0.065	P(G/A) = 3
5	0	0	0.017	P(G/A) = 4
6	Sterling	Ronaldo	Salah	

The probability that this combination wins, loses or draws and projected points:

Expected Points	1.634175495

P(WIN)	0.4760980422
P(DRAW)	0.2058813685
P(LOSE)	0.3180205893

Probability Tree:



9. Sterling - Vardy - Salah: 60 points

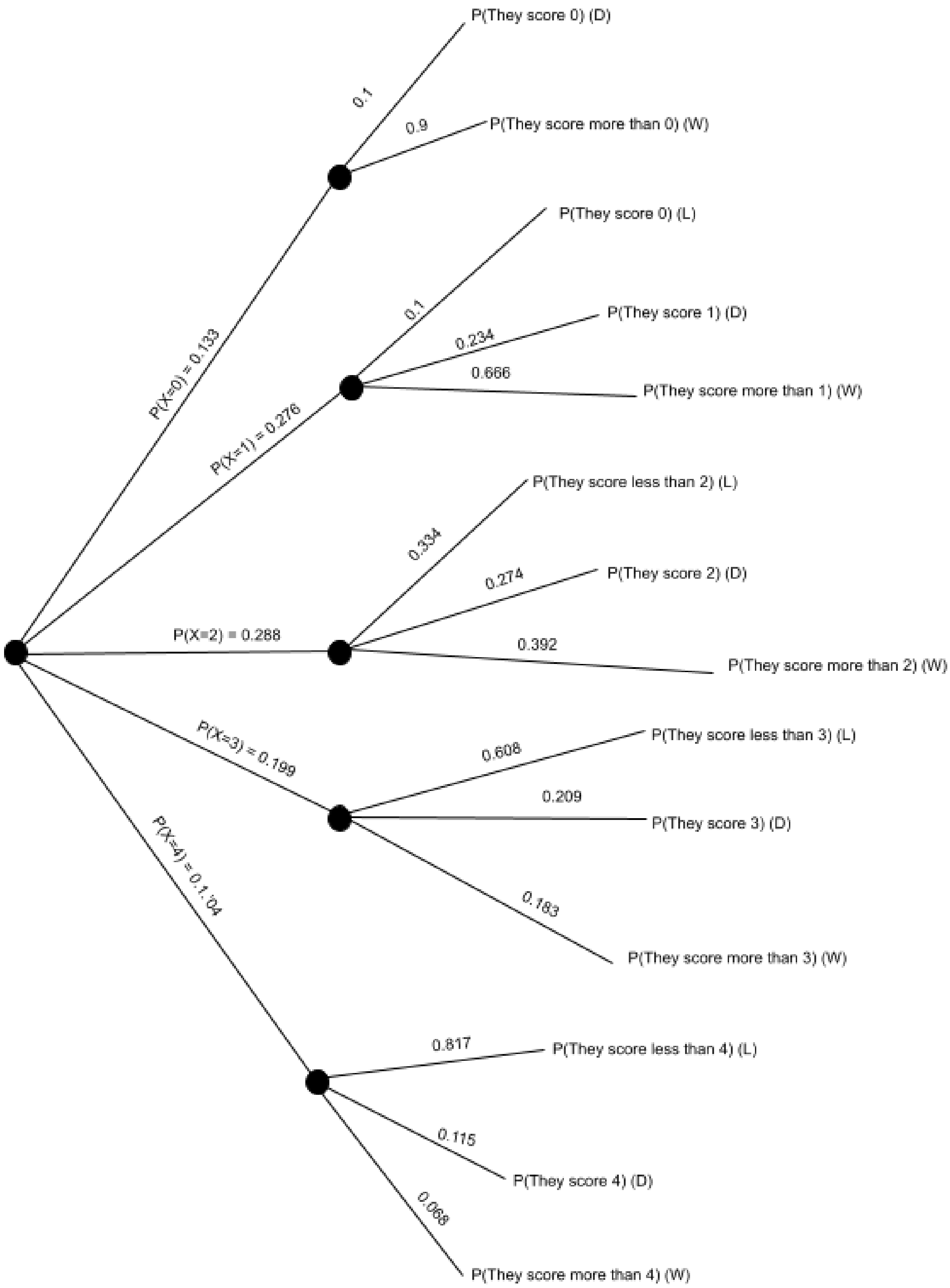
Base Probabilities:

	A	B	C	D
1	0.533	0.523	0.358	P(G/A) = 0
2	0.337	0.356	0.369	P(G/A) = 1
3	0.107	0.121	0.19	P(G/A) = 2
4	0.023	0	0.065	P(G/A) = 3
5	0	0	0.017	P(G/A) = 4
6	Sterling	Vardy	Salah	

The probability that this combination wins, loses or draws and projected points:

Expected Points	1.590750226
P(WIN)	0.4601393833
P(DRAW)	0.210332076
P(LOSE)	0.3295285407

Probability Tree:



1. Sterling - Kane - Salah: 60 points

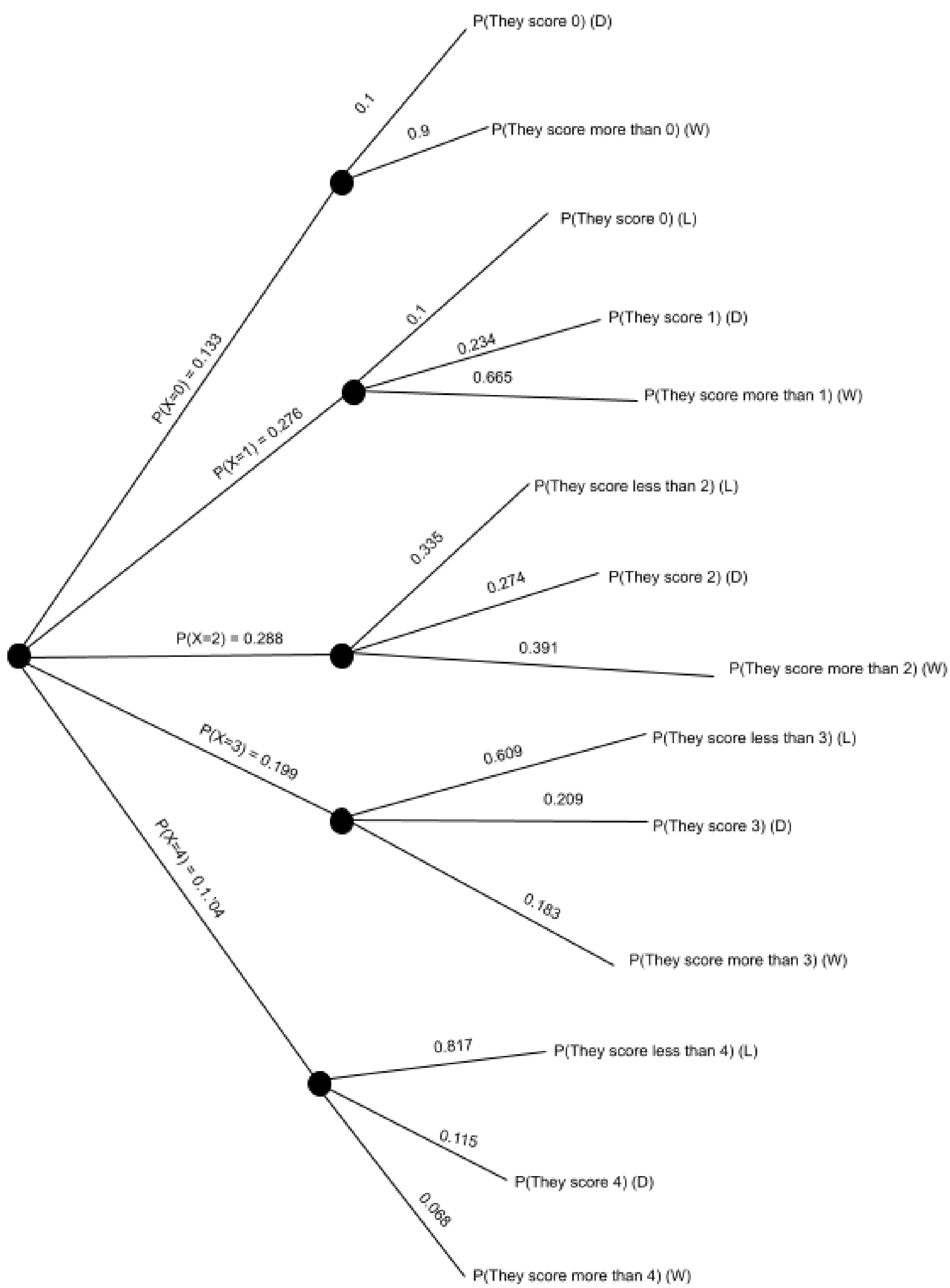
Base Probabilities:

	A	B	C	D
1	0.533	0.525	0.358	P(G/A) = 0
2	0.337	0.355	0.369	P(G/A) = 1
3	0.107	0.12	0.19	P(G/A) = 2
4	0.023	0	0.065	P(G/A) = 3
5	0	0	0.017	P(G/A) = 4
6	Sterling	Kane	Salah	

The probability that this combination wins, loses or draws and projected points:

Expected Points	1.588860193
P(WIN)	0.4594789422
P(DRAW)	0.2104233661
P(LOSE)	0.3300976917

Probability Tree:



6. Son - Ronaldo - Mahrez: 55 points

Base Probabilities:

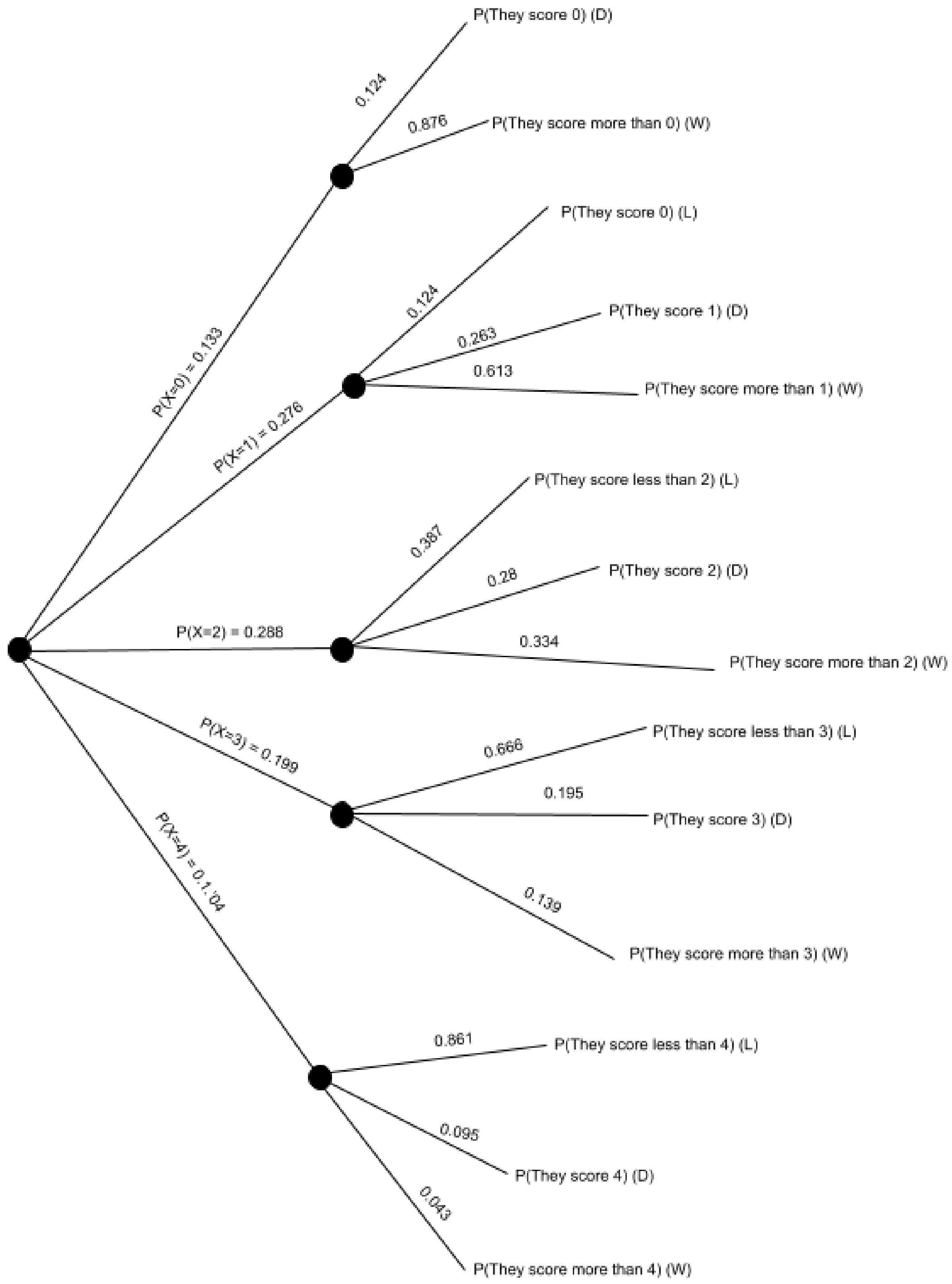
	A	B	C	D
1	0.429	0.499	0.577	P(G/A) = 0
2	0.368	0.35	0.329	P(G/A) = 1
3	0.158	0.122	0.094	P(G/A) = 2
4	0.045	0.029	0	P(G/A) = 3
5	0	0	0	P(G/A) = 4
6	Son	Ronaldo	Mahrez	

The probability that this combination wins, loses or draws and projected points:

Expected Points	1.460466313

P(WIN)	0.4140457981
P(DRAW)	0.2183289187
P(LOSE)	0.3676252831

Probability Tree:



12. Son - Vardy - Mahrez: 54 points

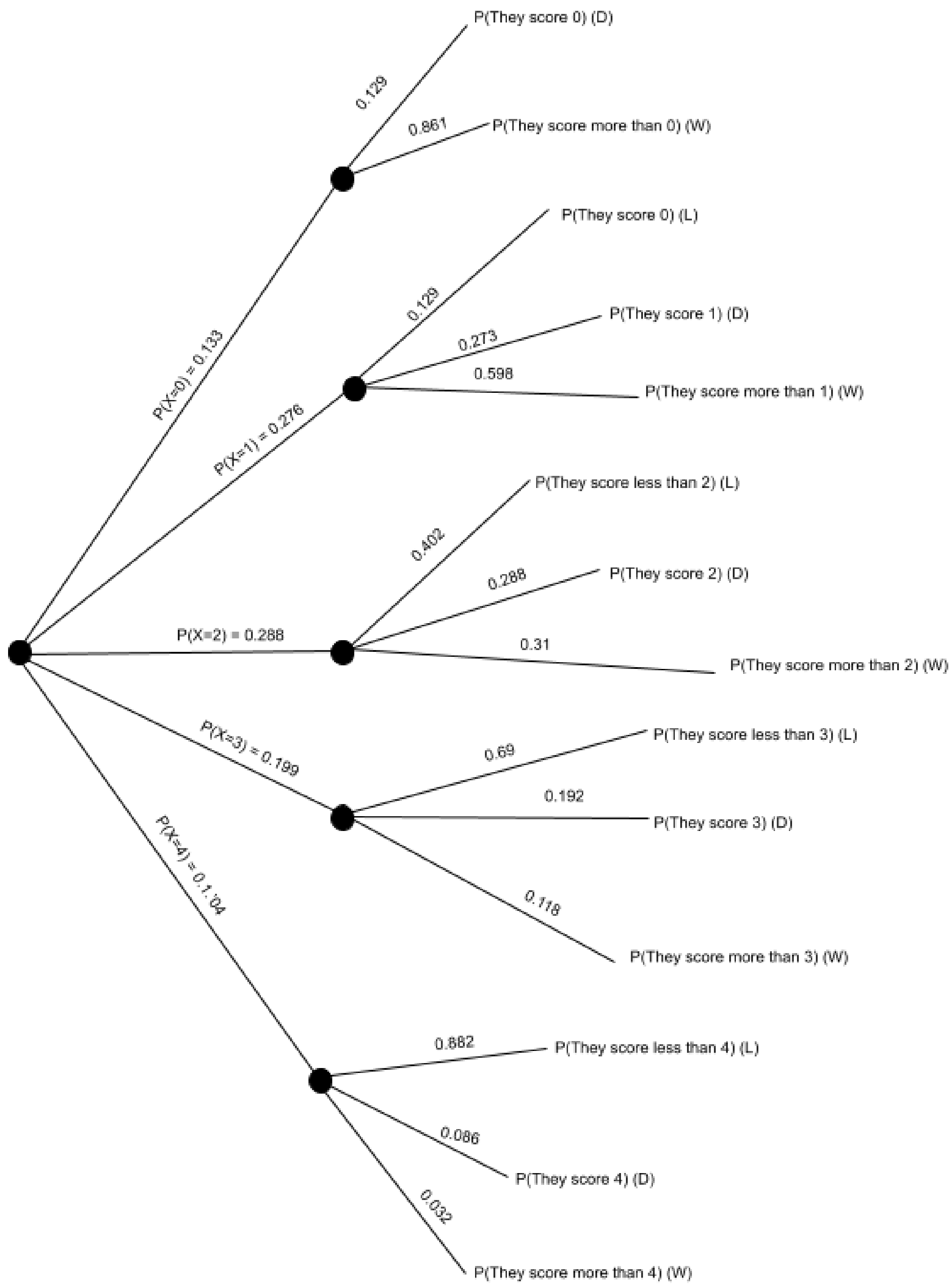
Base Probabilities:

	A	B	C	D
1	0.429	0.523	0.577	P(G/A) = 0
2	0.368	0.356	0.329	P(G/A) = 1
3	0.158	0.121	0.094	P(G/A) = 2
4	0.045	0	0	P(G/A) = 3
5	0	0	0	P(G/A) = 4
6	Son	Vardy	Mahrez	

The probability that this combination wins, loses or draws and projected points:

Expected Points	1.412723982
P(WIN)	0.396725421
P(DRAW)	0.2225477192
P(LOSE)	0.3807268598

Probability Tree:



3. Son - Kane - Mahrez: 54 points

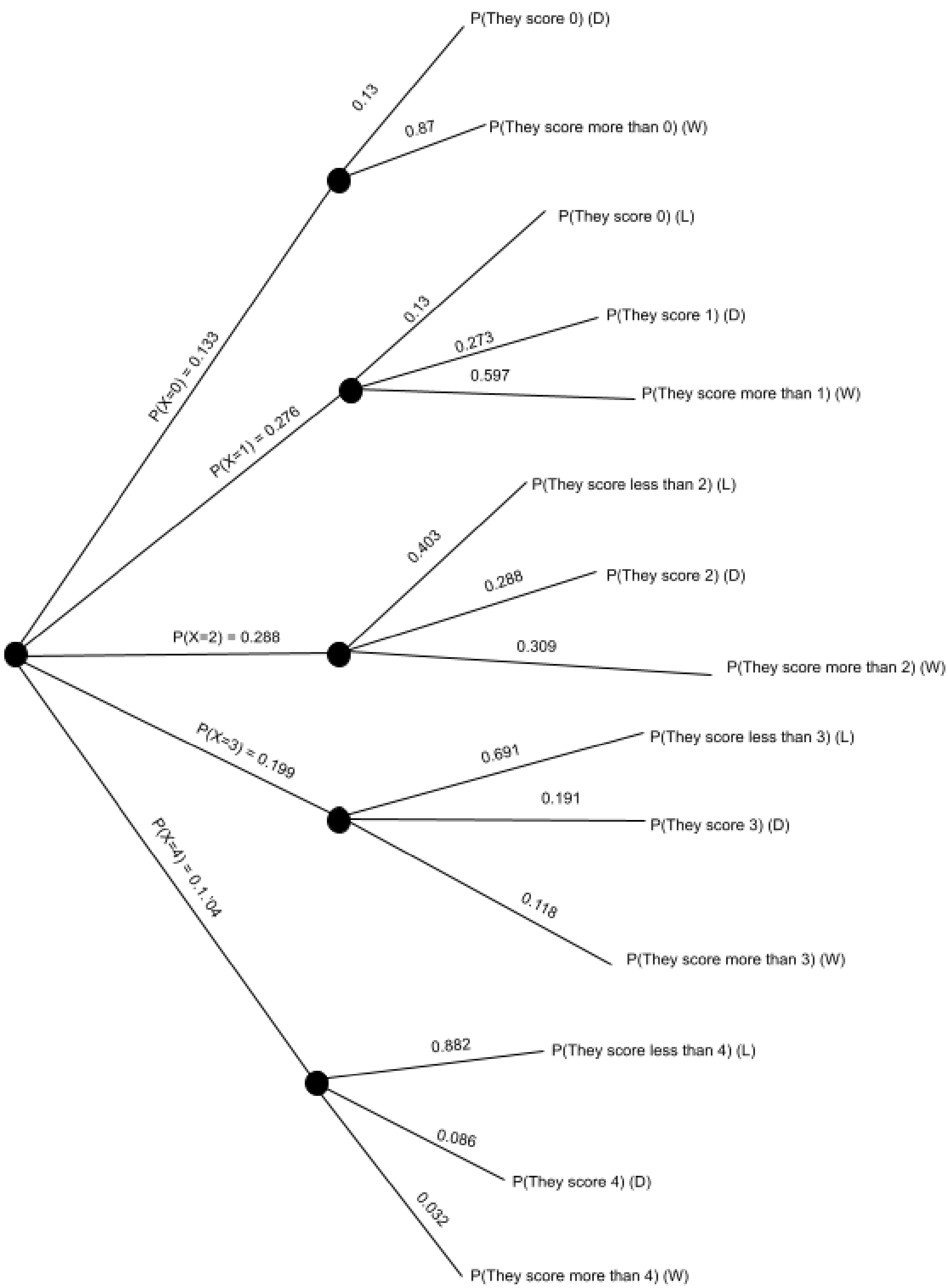
Base Probabilities:

	A	B	C	D
1	0.429	0.525	0.577	P(G/A) = 0
2	0.368	0.355	0.329	P(G/A) = 1
3	0.158	0.12	0.094	P(G/A) = 2
4	0.045	0	0	P(G/A) = 3
5	0	0	0	P(G/A) = 4
6	Son	Kane	Mahrez	

The probability that this combination wins, loses or draws and projected points:

Expected Points	1.410712947
P(WIN)	0.3960359371
P(DRAW)	0.2226051355
P(LOSE)	0.3813589274

Probability Tree:



5. Sterling - Ronaldo - Mahrez: 51 points

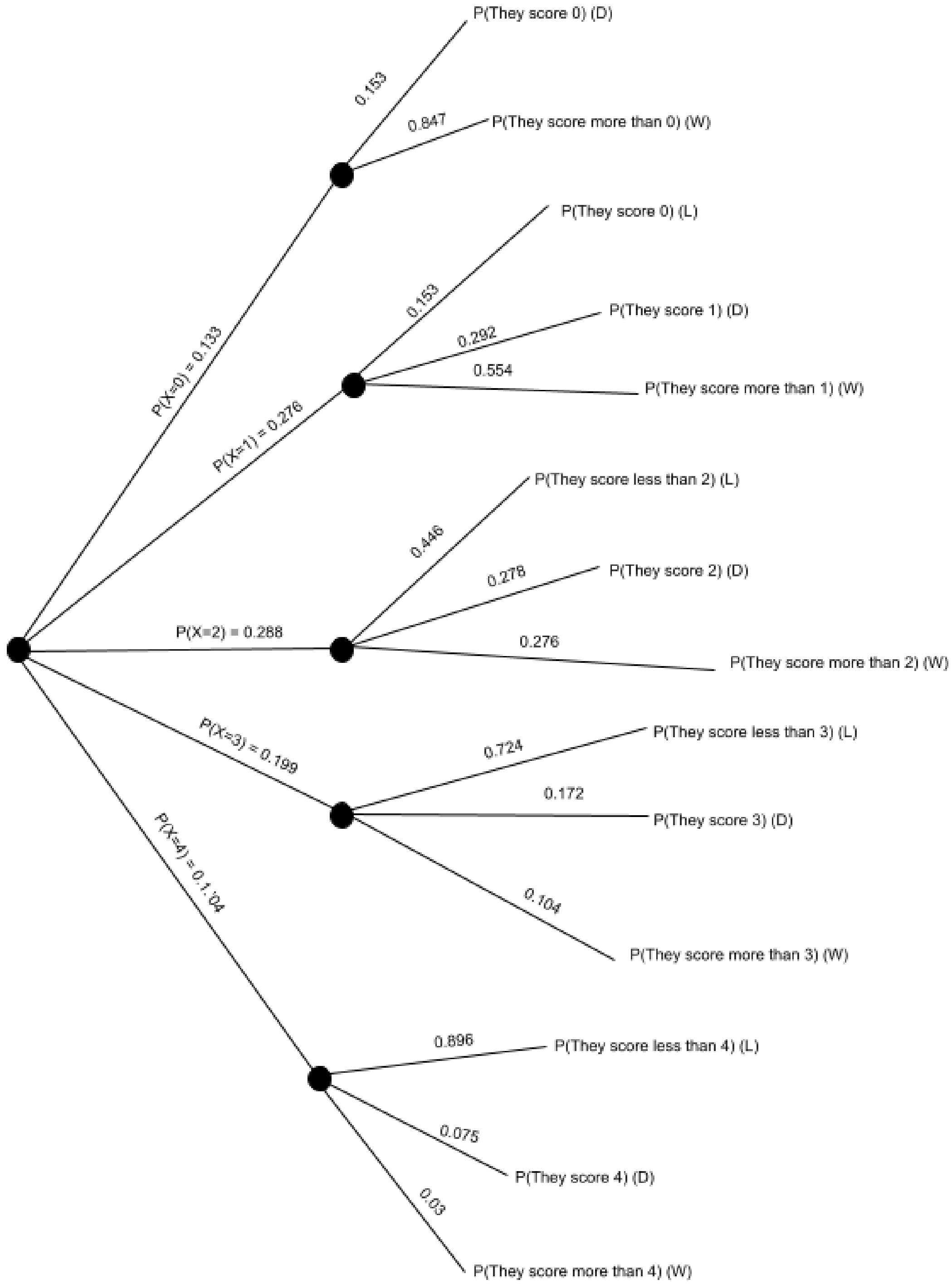
Base Probabilities:

	A	B	C	D
1	0.533	0.499	0.577	P(G/A) = 0
2	0.337	0.35	0.329	P(G/A) = 1
3	0.107	0.122	0.094	P(G/A) = 2
4	0.023	0.029	0	P(G/A) = 3
5	0	0	0	P(G/A) = 4
6	Sterling	Ronaldo	Mahrez	

The probability that this combination wins, loses or draws and projected points:

Expected Points	1.33010924
P(WIN)	0.3689908871
P(DRAW)	0.2231365784
P(LOSE)	0.4078725345

Probability Tree:



11. Sterling - Vardy - Mahrez: 49 points

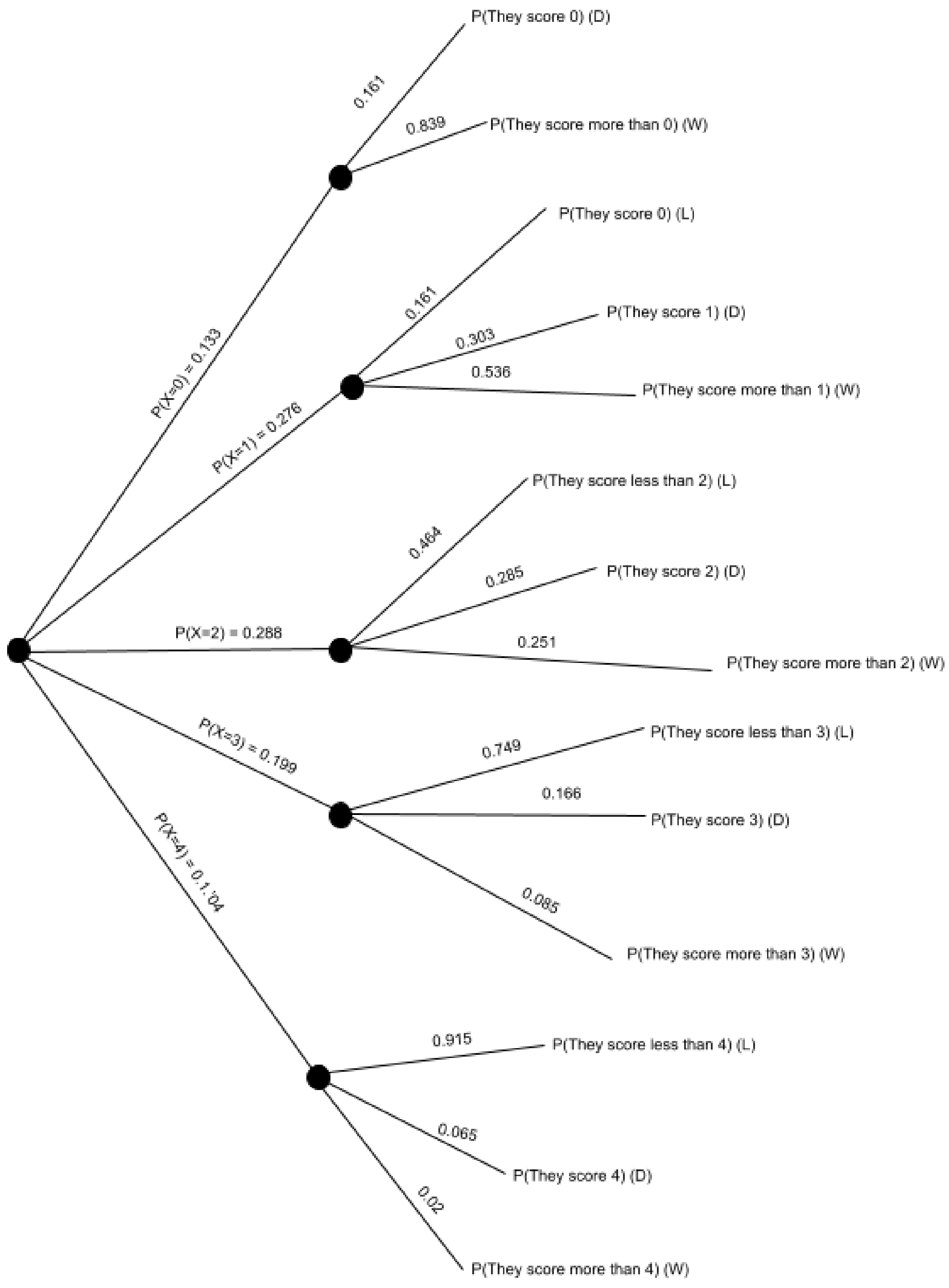
Base Probabilities:

	A	B	C	D
1	0.533	0.523	0.577	P(G/A) = 0
2	0.337	0.356	0.329	P(G/A) = 1
3	0.107	0.121	0.094	P(G/A) = 2
4	0.023	0	0	P(G/A) = 3
5	0	0	0	P(G/A) = 4
6	Sterling	Vardy	Mahrez	

The probability that this combination wins, loses or draws and projected points:

Expected Points	1.279616047
P(WIN)	0.3508977458
P(DRAW)	0.2269228099
P(LOSE)	0.4221794444

Probability Tree:



4. Sterling - Kane - Mahrez: 49 points

Base Probabilities:

	A	B	C	D
1	0.533	0.525	0.577	P(G/A) = 0
2	0.337	0.355	0.329	P(G/A) = 1
3	0.107	0.12	0.094	P(G/A) = 2
4	0.023	0	0	P(G/A) = 3
5	0	0	0	P(G/A) = 4
6	Sterling	Kane	Mahrez	

The probability that this combination wins, loses or draws and projected points:

Expected Points	1.277554979
P(WIN)	0.3502039158
P(DRAW)	0.2269432317
P(LOSE)	0.4228528525

Probability Tree: