International Journal of
Mathematics And its Applications
ISSN: 2347-1557

Int. J. Math. And Appl., **13(3)**(2025), 115–127
Available Online: http://ijmaa.in

Breaking Down the Save: A Data-Driven Approach to Penalty Shootouts

Surya Prasanna[1,*]

[1]*Inglemoor Highschool, Kenmore, Washington, United States of America*

Abstract

Penalty shootouts represent one of the most decisive and psychologically demanding moments in competitive soccer, often determining outcomes of tournaments at the highest level. This study presents a data-driven approach to optimizing a goalkeeper's decision-making process during penalty shootouts by analyzing 161 penalties from the UEFA Champions League spanning the last 20 years. Using statistical tools such as entropy, logistic regression, Chi-Square tests, and Pearson's correlation coefficient, the paper investigates patterns in ball placement, shooter's dominant foot, and dive strategies. The results highlight significant correlations between shooting foot and ball placement, revealing exploitable tendencies that enhance prediction accuracy. Further, an optimized strategy based on probabilistic modeling shows that goalkeepers can more than double their expected number of saves compared to random guessing. These findings underscore the practical value of mathematical modeling in sports, providing goalkeepers with actionable insights to improve performance in high-pressure scenarios, while also contributing to the broader dialogue on the application of statistics in real-world decision-making.

Keywords: penalty shootouts; goalkeeper optimization; probability; logistic regression; Chi-Square test; Pearson correlation; entropy.

1. Rationale

It was the 120^{th} minute of the finals in the Surf Cup San Diego Soccer tournament. The whistle blew, yet the game was tied. This entailed a penalty shootout, one of the most pressure-inducing situations in soccer. In a penalty shootout, both teams bring forth 5 shooters to shoot on the other team's goalkeeper, alternating after each shot. The team with most goals out of 5 wins the shootout and consequently wins the game. My team stepped up to shoot and missed 2 shots, putting us in a losing position. If our goalkeeper saves this next shot, we still may have a chance at winning the championship, but if he does not, we lose immediately. The other team's shooter takes his shot, and our goalkeeper dives the wrong way, and the ball hits the back of the net, leaving my team at a meager second place. This is

*Corresponding author (surya.prasanna008@gmail.com)

not an uncommon occurrence in soccer, but it is one that I would have liked to prevent. I desperately wished for weeks after that game that my goalkeeper would have saved that shot. For that reason, I authored this paper to increase a goalkeeper's chances of saving shots in a penalty shootout.

2. Aim

The aim of this math paper is to optimize a goalkeeper's chances of saving shots in a penalty shootout. To achieve this, I have taken 161 shots of data, recording every penalty shootout from the last 20 years in the Champions League. The Champions League is the highest level of soccer worldwide, with the best teams in Europe vying for its prestigious prize. Using these 161 shots, I will use probability and statistics to achieve my goal of aiding a goalkeeper in saving more shots in a penalty shootout. In this paper, I use entropy, Chi-Squared, and Pearson's correlation coefficient.

3. Introduction

Any edge gained in a penalty shootout can easily be the cause of winning a trophy, as all it takes is one extra save to change the entire scope of the shootout. Hearing many different narratives from coaches, I wondered which of the many factors affected a goalkeeper's ability to save shots the most. To ensure I could collect a full representation of statistics that could aid a goalkeeper, I tracked many things, including if the shots were a goal, where the keeper dove, where the ball went, and the shooting foot of the penalty taker. This is an excerpt of the data I collected.

Goal?	GK Direction	Ball Placement	Shooting Foot	Early Dive	Shooter Order
Yes	Left	Right	Right	Yes	1
No	Middle	Middle	Left	No	2
No	Left	Left	Right	Yes	3
Yes	Right	Left	Left	Yes	4
Yes	Left	Left	Left	No	5
...	...	...	...	...	...

**To clarify, this data was taken from the perspective of the penalty taker.

4. Definitions of Important Terms:

Stutter run-up penalty: When a player runs up to take the penalty, pauses before shooting as they wait for the goalkeeper to dive early, and then proceeds to shoot the opposite direction of the goalkeeper's dive.

Early dive: When a goalkeeper dives before the shooter takes the penalty to reach the ball earlier (denoted by "Yes" on the Early Dive column). These dives are complete guesses. This strategy is weak against players who use employ a stuttered run-up.

Reaction dive: When a goalkeeper dives immediately as a penalty taker shoots (denoted by "No" on the Early Dive column). These dives can be guesses. These types of dives combat against players who use stuttered run-ups as that strategy relies on a goalkeeper diving early. However, this dive is slower at reaching the edges of the goal.

Ball placement: Refers to where in the goal the penalty taker shoots the ball from the point of view of the shooter

5. Mathematical Analysis

5.1 Calculating Randomness of Ball Placement:

Looking at the data, there are some initial key takeaways to be seen. These are the statistics of ball placement.

$P(\text{Left}) = \frac{74}{161} \approx 0.4596$	$P(\text{Middle}) = \frac{16}{161} \approx 0.0994$	$P(\text{Right}) = \frac{71}{161} \approx 0.4410$

With this data of where shooters tend to strike the ball, I decided to calculate entropy for this set of data. Entropy could measure the level randomness in the probabilities of where a penalty taker would place the ball. Where p_i is the probability of the i_{th} event and n is the total number of outcomes, entropy H is calculated as $H = -\sum_{i=1}^{n} p_i \log_2 (p_i)$.

To begin these calculations, I decided to find the maximum entropy, to have a number to compare my results to. Since I am using 3 probabilities, left, right, and middle, I tested the maximum entropy using $\frac{1}{3}$ as a probability. Since each probability is equal, there is no way to predict which probability will occur, meaning this maximum entropy represents complete randomness. Through substituting $\frac{1}{3}$ as a probability first, I maximized entropy and could compare something to complete randomness.

$$p_1 = p_2 = p_3 = \frac{1}{3}$$
$$H = -\left(\frac{1}{3}\log_2\left(\frac{1}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right)$$
$$H = -3 \cdot \frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right) = -\log_2\left(\frac{1}{3}\right) = \log_2 3$$
$$H = \log_2 3 \approx 1.585 \text{ bits}$$

So, with 1.585 bits as the maximum entropy for 3 probabilities established, I proceeded to do this same process with my penalty data.

$$H = -\left[P(\text{left}) \cdot \log_2 P(\text{left}) + P(\text{middle}) \cdot \log_2 P(\text{middle}) + P(\text{right}) \cdot \log_2 P(\text{right})\right]$$
$$H = -\left[0.4596 \cdot \log_2(0.4596) + 0.0994 \cdot \log_2(0.0994) + 0.4410 \cdot \log_2(0.4410)\right]$$
$$H \approx -[0.4596 \cdot (-1.119) + 0.0994 \cdot (-3.330) + 0.4410 \cdot (-1.170)]$$
$$H \approx -(-0.513 - 0.331 - 0.515)$$
$$H \approx 1.359 \text{ bits}$$

The entropy value of 1.359 bits is lower than my previously calculated maximum entropy: 1.585 , meaning that ball placement is not completely random. However, the entropy is relatively high, indicating moderate uncertainty and unpredictability in shooting direction. Because of this, I realized I couldn't only look at the basic probabilities of ball placement and had to expand to other factors.

5.2 Logistic Regression

Realizing that I had to look at more factors, I decided to look at what heightens the probability of a goal in a penalty shootout. Utilizing logistic regression, I modeled the probability of a goal based on various factors a penalty kick: the shooter's shooting foot, ball placement, goalkeeper dive direction, and the order in which the player took the shot. I loaded my data as a comma-separated value file, cleaned the values, and trained my model through the Python library, "sklearn". I then applied the function "model.coef_" to interpret how each feature changes the chance of a goal.

Feature	Coefficient
Shooting Foot	0.211275
Goalkeeper Direction	0.439279
Ball Placement	0.263445
Shooter Order	-0.052797

To interpret this data, if the number is positive, the factor results in a higher chance of a goal. According to the Champions League data, I was able to make a few broad statements. My model suggests that penalty kicks are more successful if:

1. They're taken by right-footed shooters
2. The ball is placed to the right
3. The goalkeeper dives away from the shot
4. The shooter is earlier

Because the coefficients aren't huge, these effects are not dramatic. However, they do present subtle advantages within the high-stake environment of a penalty shootout. For a goalkeeper, learning how the opponent can increase their chances of scoring is essential to understanding the optimal method of approaching a penalty shootout. The coefficients were not significant enough to provide concrete methods for the goalkeeper, but by applying a basic Logistic Regression model, I already began to see patterns within my data. I continued by examining a new factor, early dives.

Testing Correlation Between Dive Type and Outcome of a Penalty:

One of the things that piqued my interest the most was the relationship between an early dive and a reaction dive. I have heard many things regarding these two types of dives from coaches. Some of my coaches have advised never to use an early dive, while others have advocated for it voraciously. So, I tracked whether a dive was an early dive or a reaction dive when compiling my data. Furthermore, I tracked whether the dive resulted in a goal or a save. With this data, I calculated the Pearson's Correlation Coefficient to find if the type of dive used correlates with a goal or save. I calculated the r value, which ranges from -1 to 1 . Its value determined the strength and direction of the correlation between the two variables. The strength of an r value can most easily be viewed using a table:

$-1 < r < -0.5$	Strong negative correlation
$-0.5 \leq r < -0.3$	Moderate negative correlation
$-0.3 \leq r < 0$	Weak negative correlation
$r = 0$	No correlation
$0 < r \leq 0.3$	Weak positive correlation
$0.3 < r \leq 0.5$	Moderate positive correlation
$0.5 < r < 1$	Strong positive correlation

Pearson's Correlation Coefficient r is calculated with the following formula:

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \cdot \sum (Y_i - \bar{Y})^2}}$$

To calculate r, I first defined my variables using binary (0 s or 1 s). I defined X as the dive type ($1 =$ Early, $0 =$ Reaction) and Y as the outcome of the penalty ($1 =$ Goal, $0 =$ Save). I organized my data utilizing a table.

Dive Type (X)	Outcome (Y)	Count
Early (1)	Goal (1)	55
Early (1)	Save (0)	17
Reaction (0)	Goal (1)	65
Reaction (0)	Save (0)	24

From this table, I found the total number of early dives and reaction dives using addition. There were 72 early dives and 89 reaction dives. To continue to find rI calculated the mean values of X and Y. I calculated the mean value of X.

$$\bar{X} = \frac{\sum X_i}{n}$$

$$\bar{X} = \frac{(55 \cdot 1) + (17 \cdot 1) + (65 \cdot 0) + (24 \cdot 0)}{161} = \frac{72}{161} \approx 0.447$$

Next, I calculated the mean value of Y.

$$\bar{Y} = \frac{\sum Y_i}{n}$$

$$\bar{Y} = \frac{(55 \cdot 1) + (17 \cdot 0) + (65 \cdot 1) + (24 \cdot 0)}{161} = \frac{55 + 65}{161} = \frac{120}{161} \approx 0.746$$

Then, I calculated the standard deviations away from the mean by substituting my mean values into $X_i - \bar{X}$ and $Y_i - \bar{Y}$. This is most easily shown using a table.

Early (1)	$X_i - \bar{X} = 1 - 0.447 = 0.553$
Reaction (0)	$X_i - \bar{X} = 0 - 0.447 = -0.447$
Goal (1)	$Y_i - \bar{Y} = 1 - 0.746 = 0.254$
Save (0)	$Y_i - \bar{Y} = 0 - 0.746 = -0.746$

My next step in the process was to calculate the cross products for each type of data. For this, I multiplied the standard deviations away from the mean to each other using $(X_i - \bar{X})(Y_i - \bar{Y})$.

$(X_i - \bar{X})(Y_i - \bar{Y})$	Cross-Product
Early (1), Goal (1)	$(0.553)(0.254) = 0.141$
Early (1), Save (0)	$(0.553)(-0.746) = -0.412$
Reaction (0), Goal (1)	$(-0.447)(0.254) = -0.114$
Reaction (0), Save (0)	$(-0.447)(-0.746) = 0.334$

I continued by summing the cross-products:

$$\begin{aligned}\sum (X_i - \bar{X})(Y_i - \bar{Y}) &= (55 \cdot 0.141) + (17 \cdot -0.412) + (65 \cdot -0.114) + (24 \cdot 0.334) \\ &= 7.755 - 7.004 - 7.410 + 8.016 = 1.357\end{aligned}$$

1.357 is the value of the numerator, so I proceeded to solve for the denominator of r. This is solved through this expression:

$$\sqrt{\sum (X_i - \bar{X})^2 \cdot \sum (Y_i - \bar{Y})^2}$$

To solve the denominator for X:

$$\sum (X_i - \bar{X})^2 = \left[72 \cdot (0.553)^2\right] + \left[89 \cdot (-0.447)^2\right] = 39.80$$

To solve the denominator for Y:

$$\sum (Y_i - \bar{Y})^2 = \left[120 \cdot (0.254)^2\right] + \left[41 \cdot (-0.746)^2\right] = 30.56$$

Now, to multiply the two values together:

$$\sum (X_i - \bar{X})^2 \cdot \sum (Y_i - \bar{Y})^2 = 1216.29$$

To find the square root:

$$\sqrt{\sum (X_i - \bar{X})^2 \cdot \sum (Y_i - \bar{Y})^2} = 34.875$$

Finally, with my denominator solved, I proceeded to solve for r.

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \cdot \sum (Y_i - \bar{Y})^2}}$$

$$r = \frac{1.357}{34.875}$$

$$r = 0.039$$

This r value follows in between the range of $0 < r \leq 0.25$, making this r value fall under the category of 'very weak positive correlation". Thus, there is barely any correlation between dive types and an outcome of a penalty kick. This means that a goalkeeper could use both dive types and have a similar chance of saving the shot. Because of this, I conclude that goalkeepers should stick with the style they are comfortable with, as according to this test, there isn't a correlation between a certain dive type and making a save. Instead of succumbing to coaching pressure, goalkeepers should use a strategy that works for them. Measures such as height, hand size, and time it takes to dive to the far side of the goal could affect the decision of which dive to use, but those factors are out of the scope of this paper.

5.3 Correlation Between Shooting Foot and Ball Placement

Another crucial factor I decided to look at is preferences of left-footed and right-footed shooters. I wanted to find if there was a correlation between where a player shot and whether they used their left or right foot. I determined that a Chi-Square test would be apt for this analysis. Firstly, I created my null and alternative hypotheses.

H_0 : A penalty taker's shooting foot (left or right) is independent of the direction in which the ball is shot

H_1 : A penalty taker's shooting foot is associated with the direction in which the ball is shot.

To find the probability of a player shooting a certain way on their shooting foot, I used conditional probability to find the probability of where shooters would place the ball based on their shooting foot. I used the conditional probability formula, where A represents where in the goal the shooter is striking the ball, and B is the shooter's shooting foot. To do this, I tracked whether the player went left or right and their shooting foot.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Probability Table

	Left	Right
Right Foot (RF)	$P(\text{Left}\mid RF) = \frac{59}{103}$	$P(\text{Right}\mid RF) = \frac{44}{103}$
Left Foot (LF)	$P(\text{Left}\mid LF) = \frac{15}{42}$	$P(\text{Right}\mid LF) = \frac{27}{42}$

For this test I didn't include shots to the middle, as when I performed a Chi-Squared Test, I got a value less than 5 in my expected frequency table. This rendered the test less effective, so as I proceeded in calculating my Chi-Squared value I removed the shots to the middle.

With this data I created two contingency tables to proceed with my Chi-Squared Test.

Observed Frequency Table

	Left	Right	Sum
Right Foot (RF)	59	44	103
Left Foot (LF)	15	27	42
Sum	74	71	145

Next, I created my expected frequency table, using the expected value formula.

$$\text{Expected Value} = \frac{\text{Row Total} \cdot \text{Column Total}}{\text{Total Number}}$$

6. Expected Frequency Table

	Left	Right	Sum
Right Foot (RF)	$\frac{74 \cdot 103}{145} = 52.6$	$\frac{71 \cdot 103}{145} = 50.4$	103
Left Foot (LF)	$\frac{74 \cdot 42}{145} = 21.4$	$\frac{71 \cdot 42}{145} = 20.6$	42
Sum	74	71	145

I used $x^2_{\text{calc}} = \sum \frac{(f_o - f_e)^2}{f_e}$ to test dependence, with f_o representing the observed frequency and f_e representing the expected frequency.

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
59	52.6	6.4	40.96	0.7787
44	50.4	-6.4	40.96	0.8127
15	21.4	-6.4	40.96	1.9140
27	20.6	6.4	40.96	1.9883
			Total:	5.4937

The x^2_{calc} value must be lower than the critical value to reject the alternative hypothesis. To acquire the critical value, I calculated degrees of freedom (df) through this formula using rows (r) and columns (c):

$$df = (r - 1)(c - 1)$$

The contingency table for my data set had 2 rows and 2 columns, so I substituted the values into the formula:

$$df = (2 - 1)(2 - 1) = 1$$

I used the table below to determine if my alternative hypothesis was rejected or not. I used the significance value 0.05. If the x^2_{calc} value was lower than the critical value, there would be moderate evidence that the alternative hypothesis was rejected. The critical value is in red:

Significance Level			
Degrees of Freedom	0.10	0.05	0.01
1	2.706	3.841	6.635
2	6.251	7.815	11.345

$$5.4937 > 3.841$$

The x^2_{calc} value is greater, thus, the alternative hypothesis failed to be rejected. This alludes to a correlation between a player's shooting foot and where they place a penalty. However, when the x^2_{calc} value isn't significantly greater than the critical value, another test called the Yate's Continuity Correction Test can be used. This test would lessen the x^2_{calc} value, helping prevent overestimation of correlation. If the x^2_{Yates} value is less than the critical value, the alternative hypothesis will be rejected. The equation for x^2_{Yates} is as follows:

$$x^2_{\text{Yates}} = \sum \frac{(|f_o - f_e| - 0.5)^2}{f_e}$$

I substituted my data into a table:

f_e	$\lvert f_o - f_e \rvert$	$(\lvert f_o - f_e \rvert - 0.5)$	$(\lvert f_o - f_e \rvert - 0.5)^2$	$\frac{(\lvert f_o - f_e \rvert - 0.5)^2}{f_e}$
52.6	6.4	5.9	34.81	0.6618
50.4	6.4	5.9	34.81	0.6907
21.4	6.4	5.9	34.81	1.6266
20.6	6.4	5.9	34.81	1.6898
			Total:	4.7049

I compared this to the critical value: $4.7049 > 3.841$. This result entails that there is strong evidence of correlation between a player's shooting foot and where they will shoot the ball. From this, there is strong evidence that right footed players tend to shoot left and left footed players tend to shoot right. Knowing that this correlation could help optimize a goalkeeper's chances, I decided to use these probabilities.

6.1 Optimizing a Goalkeeper's Chances with the Correlation

First, I wanted to calculate the number of saves I needed to beat with my data. This is the probability of a save:

$$P(\text{save}) = \frac{46}{161} \approx 0.286$$

To find the expected number of saves that a goalkeeper makes, I used the expected value formula:

$$E(x) = \sum xP(X = x)$$

To calculate this value, I used a table. In this table, I used the binomial probability formula where n represents the number of penalties and x indicates the number of penalties saved:

$$P_x = \binom{n}{x} p^x (1-p)^{n-x}$$

x (Number Saved)	$P(X = x) = \binom{5}{x} 0.286^x 0.714^{5-x}$	$x \cdot P(X = x)$
0	$\binom{5}{0} 0.286^0 0.714^{5-0} = 0.1856$	0
1	$\binom{5}{1} 0.286^1 0.714^{5-1} = 0.3716$	0.3716
2	$\binom{5}{2} 0.286^2 0.714^{5-2} = 0.2977$	0.5954
3	$\binom{5}{3} 0.286^3 0.714^{5-3} = 0.1193$	0.3579
4	$\binom{5}{4} 0.286^4 0.714^{5-4} = 0.0239$	0.0956
5	$\binom{5}{5} 0.286^5 0.714^{5-5} = 0.0190$	0.0950
	$E(x) = \sum xP(X = x)$	1.5155

The expected number of saves a goalkeeper is to make is 1.5155 , and I wanted to try to improve that with my knowledge of a correlation between a player's shooting foot and where they place the ball. Looking back on my probabilities:

	Left	Right
Right Foot (RF)	$P(\text{Left} \mid RF) = \frac{59}{103} \approx 0.573$	$P(\text{Right} \mid RF) = \frac{44}{103} \approx 0.427$
Left Foot (LF)	$P(\text{Left} \mid LF) = \frac{15}{42} \approx 0.357$	$P(\text{Right} \mid LF) = \frac{27}{42} \approx 0.643$

Through these probabilities, I observed that left footed players are more likely to shoot right than left, and right footed players are more likely to shoot left than right. I decided to find out how many penalties a goalkeeper could guess correctly if they dove right for left footed players and dove left for right footed players.

To start, I calculated the expected value of left footed and right footed players there would be per 5 shooters in the shootout. From my data, I found these probabilities:

$$P(LF) = \frac{47}{161} \approx 0.292 \quad P(RF) = \frac{114}{161} \approx 0.708$$

Using the expected value formula $E(x) = np$, where n is the number of penalty shooters and p represents the probability that a player is left footed, I found the expected number of left footers.

$$E(x) = np$$

$$E(x) = (5)(0.292) = 1.46$$

This gave me an expected 1.46 left footed players per 5 shooters in a penalty shootout. Repeating this with right footed players, I got an expected value of 3.54 right footed players. Using these expected values with my previous probabilities regarding where a shooter would shoot based on their preferred foot, I hoped to create a higher number of saved penalties. To do this, I needed to calculate how many correctly guessed penalties a goalkeeper could achieve per 1.46 left footed players. I continued through the usage of the expected value formula and binomial probability formula:

$$P_x = \binom{n}{x} p^x (1-p)^{n-x} \quad E(x) = \sum xP(X=x)$$

In this case, n would represent the expected number of left footed players (1.46), and p would represent the probability a left footed player shoots right (0.643). x entailed the number of left footed players shooting right. I set x to 0,1 , and 2 because after 2 , the probabilities became miniscule and insignificant to the expected value.

However, I ran into a problem. When using $\binom{n}{x}$ mathematicians typically use only positive integers. It is denoted as such:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

This is because factorials (!) don't operate without positive integers. Since my n value was 1.46, I couldn't use this formula to find the $\binom{n}{x}$ value. Looking online, I found a formula to find $\binom{n}{x}$. This way included the usage of the gamma function, a function that allows mathematicians to find the factorial of complex numbers. Gamma is noted as such:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} d^t$$

And the formula is adjusted for the gamma function as so:

$$\binom{n}{x} = \frac{\Gamma(n+1)}{\Gamma(x+1) \cdot \Gamma(n-x+1)}$$

Calculating gamma by hand is an arduous ordeal and is something I'm not able to do with my current knowledge of math. However, using python code, I wrote a method to solve for $\binom{n}{x}$.

```
def generalized_binomial_coefficient(n, k):
    return gamma(n + 1) / (gamma(k + 1) * gamma(n - k + 1))
print(generalized_binomial_coefficient(n,k))
```

Using this, I was able to find $\binom{n}{x}$ and continue with my calculations to find the expected number of correctly guessed penalties a goalkeeper would have if they dove right for left footed penalty takers. Using my code, I got the following values for $\binom{n}{x}$.

$$\binom{1.46}{0} = 1, \quad \binom{1.46}{1} = 1.46, \quad \binom{1.46}{2} = 0.336$$

x	$P(X=x) = \binom{1.46}{x} 0.643^x 0.357^{1.46-x}$	$x \cdot P(X=x)$
0	$\binom{1.46}{0} 0.643^0 0.357^{1.46-0} = 0.222$	0
1	$\binom{1.46}{1} 0.643^1 0.357^{1.46-1} = 0.585$	0.585
2	$\binom{1.46}{2} 0.643^2 0.357^{1.46-2} = 0.242$	0.484
	$E(x) = \sum xP(X=x)$	1.069

I repeated this process with right footed players, with their tendency to shoot left.

$$P(\text{Left} \mid RF) = \frac{59}{103} \approx 0.573$$

x	$P(X=x) = \binom{3.54}{x} 0.573^x 0.427^{3.54-x}$	$x \cdot P(X=x)$
0	$\binom{3.54}{0} 0.573^0 0.427^{3.54-0} = 0.049$	0
1	$\binom{3.54}{1} 0.573^1 0.427^{3.54-1} = 0.234$	0.234
2	$\binom{3.54}{2} 0.573^2 0.427^{3.54-2} = 0.398$	0.796
3	$\binom{3.54}{3} 0.573^3 0.427^{3.54-3} = 0.274$	0.822
4	$\binom{3.54}{4} 0.573^4 0.427^{3.54-4} = 0.050$	0.200
	$E(x) = \sum xP(X=x)$	2.052

For left footed players, I obtained an expected value of 1.069 correctly guessed penalties and for right footed players I obtained an expected value of 2.052 . Since these are independent events, I was able to add them together to find the total expected value of correctly guessed penalties for my strategy.

$$E(\text{Total}) = 1.069 + 2.052 = 3.121$$

Using the strategy of diving left for right footed players and left for right footed players gave me an expected value of 3.121 correctly guessed penalties, which is far above the 1.5155 expected saves. If a goalkeeper uses this strategy to correctly guess and save penalties, it is much more optimal than guessing without a strategy.

7. Conclusion

Through thorough calculations, I discovered a correlation between a player's shooting foot and ball placement. With this, I created a strategy which gave an expected value of 3.121 saves per 5 shots. This did help me optimize a goalkeeper's penalty saving ability, and it is something that I will be telling the goalkeeper on my team. However, I realized this strategy comes with limitations. The Chi-Squared test I did didn't include players shooting the ball to the middle, which could have changed this strategy entirely. Furthermore, even if a goalkeeper correctly guesses a penalty, it doesn't mean the goalkeeper will save the penalty. If a penalty is hit with enough pace and accuracy, there's a chance it can still go in. Also, teams can pick up on this strategy and counter it, rendering it useless.

Other factors such as the crowd, pressure, and even the type of ball can hinder a goalkeeper's ability to save a penalty. These are just a few examples, something like home field advantage can also completely change the situation for a goalkeeper psychologically. For future studies, I recommend collecting data from different competitions and tournaments to have more data points to work with. To get the most comprehensive analysis of how to optimize a goalkeeper's penalty saving ability, many metrics need to be measured. Through many correlation tests of a breadth of factors, a stronger optimization can be created. Although I created a strategy for a goalkeeper to correctly guess 3.121 penalties out of 5 in a penalty shootout, there are a myriad of factors that could prevent this strategy from working. Despite meticulous planning for penalty shootouts, the passion, tension, and nerves of soccer that keep the game alive will ensure that no two penalty shootouts are the same. This interplay between unpredictability and strategy is what makes soccer renowned for being the beautiful game.

References

[1] *Gamma Function*, Wolfram Mathworld, mathworld.wolfram.com/GammaFunction.html, (2025).

[2] *How to calculate binomial coefficients with non-integer numbers*, Stack Overflow, Prosus, (2017), stackoverflow.com/questions/47725238/how-to-calculate-binomial-coefficents-with-non-integer-numbers.

[3] Evan Pierce, *Understanding Entropy: A Comprehensive Guide with Examples*, Medium, Medium Corporation, (2023), medium.com/@mrthinger/understanding-entropy-a-comprehensive-guide-with-examples-f014cd3c6cc7.

[4] Shaun Turney, *Chi-Square (X^2) Table | Examples & Downloadable Table*, Scribbr, Learneo, (2022), www.scribbr.com/statistics/chi-square-distribution-table/.

[5] ———, *Pearson Correlation Coefficient (r) | Guide & Examples*, Scribbr, Learneo, (2022), www.scribbr.com/statistics/pearson-correlation-coefficient/.

[6] *UEFA Champions League - Penalty Shoot-out*, transfrmarket.us, www.transfermarkt.us/uefa-champions-league/elfmeterschiessen/pokalwettbewerb/CL, (2025).

[7] *Yates Correction: Definition, Examples*, Statistics How To, www.statisticshowto.com/what-is-the-yates-correction/, (2025).