

ESTIMATING THE PROBABILITY OF A SHOT RESULTING IN A GOAL: THE EFFECTS OF DISTANCE, ANGLE AND SPACE.

Richard Pollard, Jake Ensum and Samuel Taylor
Statistics Department, California Polytechnic State University
E-mail: rpollard@calpoly.edu

UNITED STATES OF AMERICA

ABSTRACT

Pollard, R.; Ensum, J. and Taylor, S. (2004). Estimating the probability of a shot resulting in a goal: The effects of distance, angle and space. **International Journal of Soccer and Science**, 2(1), 50-55. It takes on average about 10 shots to score one goal. However, the probability of scoring varies greatly from shot to shot. Two previous studies, using data recorded from the World Cups of 1986 and 2002, have investigated the effect of the location of the shot on the chances of scoring, together with the influence of other factors. The data from these two analyses are combined into a single sample. Because of their different characteristics and smaller sample sizes, penalties, headed shots and shots direct from a free-kick are excluded, leaving a total of 1096 kicked shots which produced 110 goals. A logistic regression analysis shows that three factors each have a highly significant influence ($p < .001$) on the chances of a shot producing a goal. These factors are: 1. The distance from goal (each extra yard from goal decreases the odds of scoring by 15%); 2. The angle from the goalpost (each extra degree away from the goal decreases the odds of scoring by 2%); 3. Space from the nearest opponent at the time of the shot (if more than one metre of space, the odds of scoring are more than doubled). The logistic regression equation enables the probability of scoring to be estimated from any given situation. This can be illustrated by probability contours drawn on diagrams of the field of play. The low probability of scoring from long range emphasizes the importance of getting the ball into locations of higher scoring probability. For the striker, finding space and shooting first time are important attributes.

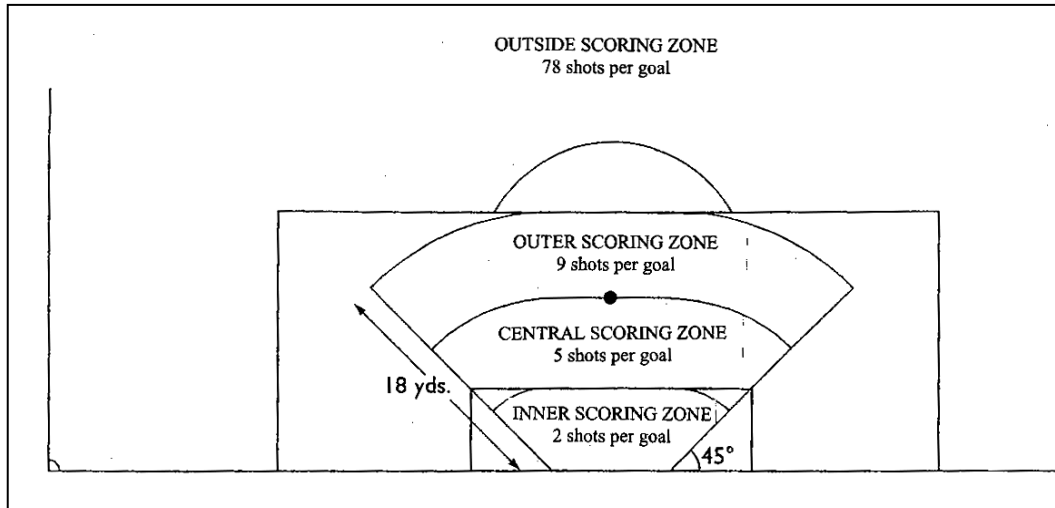
INTRODUCTION

In professional soccer, it takes on average about 10 shots to score one goal. However the probability of scoring varies greatly from shot to shot. The purpose of this paper is to quantify the chances of scoring in relation to the location of the shot and to investigate the possible influence of other factors.

Although much research has been conducted on goalscoring, comparatively little of it has been directed at the association between shots and goals. In an analysis of over 10,000 shots between 1952 and 1967, Reep and Benjamin (1968) showed that on average 10 shots were required to produce one goal, and that this figure varied very little between the data sets which were based on league games in England and on international competition. Pollard (1995) analyzed 3,931 shots from competitions worldwide between 1969 and 1991 and obtained a figure of 10.0 shots per goal. Shots were separated from inside and outside the penalty area and a large difference in scoring potential was found; an average of 6.5 shots per goal from inside and 46.1 shots per goal from outside. Further

analysis from within the penalty area gave more insight into the relative chances of scoring. A 'scoring zone' was identified as a section of the penalty area from where most goals are scored. The average number of shots per goal from three distinct regions from within this zone were derived and are shown in Figure 1, together with the very high ratio from outside this zone.

Figure 1. Scoring zone with average number of shots per goal from different locations



Following this initial attempt to quantify the importance of location on the outcome of a shot, a much more detailed study used logistic regression to produce an equation from which the exact probability of scoring could be estimated (Pollard & Reep, 1997). In addition to the precise location of the shot, the effects of other factors on the success of a shot were investigated and, where significant, quantified. The study was based on shots recorded from the 1986 World Cup. Penalties (5), shots taken direct from a free kick (43), all other kicked shots (367) and headed shots (74) were each analyzed separately because of their different characteristics in relation to the effect of factors likely to influence the chances of scoring.

The same approach was taken in a follow-up study in which 729 kicked shots recorded from the 2002 World Cup were used (Ensum, Pollard and Taylor, 2003). A similar logistic regression was carried out on this larger sample, and a broader set of possible factors affecting the outcome of the shot was investigated.

The purpose of this paper is to combine the results of these two previous studies in order to produce a more reliable equation from which the probability of scoring a goal from a kicked shot can be estimated.

DEFINITIONS AND METHODS

Shots

There is no official definition as to what constitutes a shot in soccer. The definition in this study is the one adopted by Reep in 1950 and used in the studies by Reep & Benjamin (1968), Pollard (1995) and Pollard & Reep (1997).

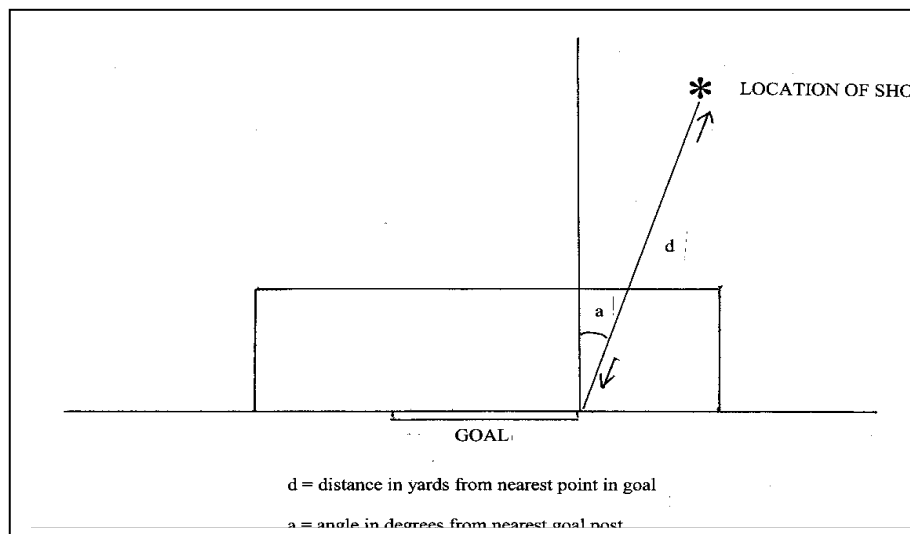
A **shot** is defined as a direct attempt to score a goal resulting in a player striking the ball towards the opponents' goal. Shots blocked by players other than the goalkeeper are counted as shots only if the goalkeeper is not in a position to save the shot, or if the player blocking the shot is inside the goal area. This treatment of blocked shots may be slightly more restrictive than other definitions, including that by Ensum et al. (2003). There may therefore be a few shots included in that study which would not have been recorded as shots using Reep's definition. There is also a gray area between crosses and shots that will always cause problems (remember Ronaldinho's goal against England in the 2002 World Cup) and a subjective decision will have to be made as to whether or not an attempt at goal was intended.

Factors relating to the shot

Each of the two studies being combined included an analysis of a number of factors that might have a possible influence on the chances of a shot resulting in a goal. The following seven factors were common to both studies:

The location of the shot was quantified by two variables, **distance** from the nearest point in goal, and the **angle** from a line perpendicular from the goal-line drawn from the nearest goalpost. This is shown in Figure 2. Shots taken from directly in front of the goal thus had an angle of zero.

Figure 2. Definition of distance and angle



The amount of **space** had by the player taking the shot was categorized as being either less than, or greater than one metre from the nearest opponent.

The number of **touches** on the ball by the player taking the shot was categorized as either one, or more than one.

The **type of play** leading up to the shot was categorized as originating either from open play or from a set play (kick off, goal kick, throw in, free kick or corner)

Whether or not the shot was immediately preceded by a **cross**

The **zone of origin** of the move leading up to the shot was categorized as being in the defensive, middle or the attacking third.

The second study (Ensum, et al. 2003) considered several other possible influential factors. The position of the goalkeeper, and the number of players between the player taking the shot and the goal were both included in the statistical analysis. Other factors which were recorded but deemed unsuitable for logistic regression included the number of passes in the movement leading to the shot, the time of the match and the state of the match (whether the team taking the shot was winning, losing or drawing).

Statistical analysis

The combined sample of 1096 kicked shots was analyzed using only the seven factors common to both studies. The scoring of a goal (yes or no) was the binary response variable. A logistic regression analysis was performed in order to investigate and quantify the effect on goalscoring of each of the seven factors which were used as the explanatory variables in the regression. The statistical package Minitab was used for the analysis.

The logistic regression equation derived from the analysis was subsequently used to estimate the probability of a shot becoming a goal under specific conditions. Only those factors found to be significant at the 5% level were incorporated into the regression equation. The probability of scoring (p) was calculated from the value (y) predicted from the logistic regression equation, using the formula:

$$p = e^y / (e^y + 1) \quad (1)$$

RESULTS

Second order interaction terms were included in the initial logistic regression model. None were found to be significant and subsequent models were based on linear combinations of subsets of the seven factors described above. The final model consisted of three factors, all of which had a highly significant effect on the probability of a shot becoming a goal ($p < 0.001$). These factors were distance, angle and space as defined above. Details of the analysis are shown in Table 1. None of the other factors were significant at the 5% level. The final logistic regression equation can therefore be written:

$$y = 0.377 - 0.159 (\text{distance}) - 0.022 (\text{angle}) + 0.799 (\text{space}) \quad (2)$$

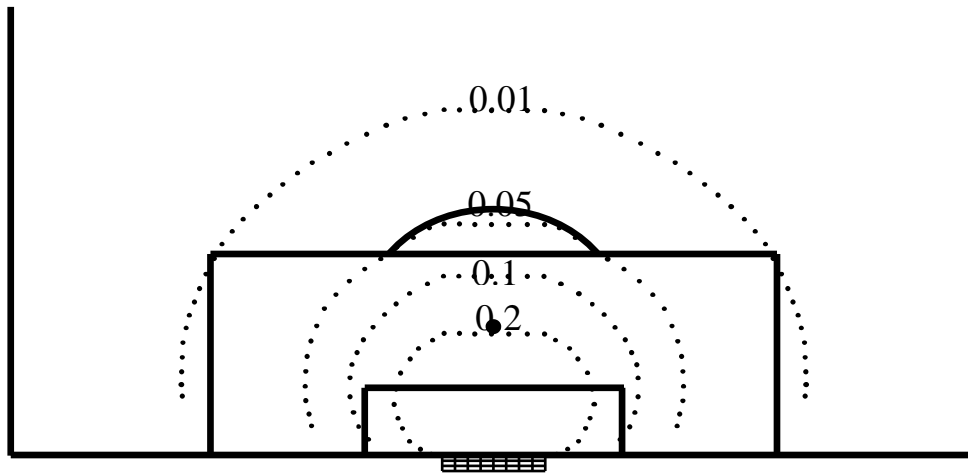
Table 1. Results of logistic regression analysis

Predictor	Coefficient	z	p	Odds	95% C.I.	
				ratio	Lower	Upper
Constant	0.3771	1.20	0.229			
Distance	-0.1586	-9.51	0.000	0.85	0.83	0.88
Angle	-0.0222	-3.81	0.000	0.98	0.97	0.99
Space	0.7991	3.22	0.001	2.22	1.37	3.62

From this equation it is possible to estimate the probability of a shot scoring, based on its location (quantified by distance in yards and angle in degrees) and by whether or not the person taking the shot had space (quantified by 0 if there was an opponent within one metre, and 1 if not). For example, suppose a player takes a shot 15 yards from the goal, at an angle of 20 degrees from the goal post and with no opponent with one metre. Substituting the values distance = 15, angle = 20 and space = 1 into equation (2), a value of -1.649 is obtained for y. Then this value is substituted into equation (1) to give a value of $p = 0.16$, the estimated probability of the shot scoring. It should be emphasized that this should be considered as an ‘average’ probability derived from the observed sample of 1096 shots. For a particularly good striker, this figure might be higher; for a less skillful player the figure would be lower.

Using equation (2), probability contours can be constructed showing the average probability of scoring from different locations. An example is given in Figure 3 for shots with the nearest opponent within one metre.

Figure 3. Scoring probability contours for kicked shots with the nearest opponent within one metre



The odds ratios shown in Table 1 allow the effect of the factors to be quantified. For example, the odds ratio of 0.85 for distance means that for every extra yard away from goal, the odds of scoring decrease by 15%. For angle, there is a 2% decrease for each degree, while for space, the odds of scoring more than double (odds ratio 2.22) if a player is more than one metre away from the nearest opponent. Also shown in Table 1 are 95% confidence intervals for these odds ratios.

DISCUSSION

The results from the large sample of 1096 kicked shots show that the location of the shot and the amount of space from the nearest opponent both have a highly significant influence on the outcome of the shot. The sample was the combination of data from two previous studies, the results from which were somewhat different as described in Ensum et al. (2003). For the data from the 1986 World Cup, in addition to location and space, the probability of scoring was significantly influenced by the type of play (open play having a greater influence than set play). For the data from the 2002 World Cup, once again location and space were significant, but in addition, a shot from a cross increased the chances of scoring. It should be mentioned that the number of touches on the ball did not have a significant effect on the probability of goalscoring, whereas being more than one metre away from the nearest opponent did have a beneficial effect. The implication when training a striker is to prioritize their ability to find space and shoot first time. It is important to shoot first time, even in space since this will reduce the amount of time the opposition has to anticipate the shot.

Although other factors still require further investigation, a relatively simple model for goalscoring has been established from which a number of uses can be made. These include:

1. Establishing the probability of scoring from a shot in any given situation.
2. Quantifying the effects of the factors involved in goalscoring.
3. Quantifying the effectiveness of specific tactics and strategies as outlined in Pollard & Reep (1997).
4. Quantifying team performance in a particular game as outlined in Ensum et al. (2003).
5. Quantifying the capabilities of individual strikers and goalkeepers (also discussed in Ensum et al. 2003).

BIBLIOGRAPHY

- Ensum, J., Pollard, R. & Taylor, S. (2003). Applications of logistic regression to shots at goal at association football. *Paper presented at 5th World Congress on Science and Football*, Lisbon, April.
- Pollard, R. (1995). Do long shots pay off? *Soccer Journal*, 40(3), 41-43.
- Pollard, R. & Reep, R. (1997). Measuring the effectiveness of playing strategies at soccer. *The Statistician*, 46, 541-550.
- Reep, C. & Benjamin, B. (1968). Skill and chance in association football. *Journal of the Royal Statistical Society, A*, 131, 581-585.