

# Draw prediction for practical soccer betting strategies

*Łukasz Łoziuk<sup>1</sup>, Urszula Libal<sup>2</sup>*

**Abstract:** *Paper describes a systematized soccer betting approach and presents methods of betting market analysis. Two algorithms, characterized by increased expected rate of draw result, are proposed. The algorithms use historical data to estimate probabilities of draw in future matches. Tests on real betting market were performed and gave positive results.*

**Keywords:** *betting, soccer, forecasting, prediction, statistical analysis*

## 1. Introduction

Game algorithm for betting is a systematic approach to gambling, that aims at combating financial policy of casino or bookmaker and profits development. Efficient algorithm should increase the chances of victory in order to develop a long-term profit. Effective betting systems, based on a statistical analysis, use rare circumstances favorable to the player. The course of the match, which bookmaker offers to players, is calculated basing on the estimated probability of a particular outcome of the match. Bookmaker always strives to offer a course to players that will give a negative value of expected profit. Bookmaker's profit is:

$$I = \nu W \quad (1)$$

where  $W$  is the total amount of bets placed,  $\nu$  is a vig [8] is described by the equation:

$$\nu = 100 \cdot \left( \frac{pq}{p+q} \right) \quad (2)$$

where  $p$  and  $q$  are offered match courses. The amount  $\nu$  can be called a commission, charged by the bookmaker. Betting outlets strongly compete with each other, offering sometimes overvalued matches. Overvalued game makes a bet profitable.

## 2. Background and related work

The first work on the statistical modeling of data from football leagues are mainly focused on the distribution of the number of goals scored in the match. Moroney [6] and Reep [7] agreed that that the outcome of the game is largely random, and predicting the end result on the basis of historical data is not possible. Hill [5] applied a simple comparative test of final positions of teams in the league with the predictions of experts, which proved to be strongly correlated. Much more complex analysis is carried out by Fahrmeir [8], who used his techniques designed for the analysis of time-dependent comparisons to data from the German football league. Fahrmeir noted the following dichotomy: it is not a problem to quite accurately predict which team will be successful, but is much more difficult to develop a model that would be precise enough to take advantage of long-term prediction. Fahrmeir [8] examined the correlation between the play as the host and the match final score. Owen [4] applied dynamic linear models taking into account the change in potential teams to predict their results, and noted an increase in the accuracy of predictions.

---

<sup>1</sup> Wrocław University of Technology, Department of Computer Science, Wrocław, lukasz.loziuk@pwr.edu.pl

<sup>2</sup> Wrocław University of Technology, Department of Control Systems and Mechatronics, Wrocław, urszula.libal@pwr.edu.pl

### 3. Data

Every football match is a rich source of information. There are saved not only results of games, but also time to get goals, names of the shooters, team position in a league, date, time of game, etc. Individual team's condition in a selected game is also determined by a number of external factors: the emergence of new players in the team, or release of team's manager. Although this information is also available, is less formal, and its quality is subjective.

The data from more than thirty leagues and ten football seasons were subjected to statistical analysis and complete search. The total number of teams was over 500. The teams come from 16 countries, especially from Europe: Portugal, Spain, France, Italy, Greece, Hungary, Czech Republic, Romania, Belgium, Holland, Germany, Austria, Poland, Denmark, Scotland and USA. The data contained match results for the years 2002-2010. The differences of simulated probability distributions for given bookmaking algorithms were searched. The data and algorithms that were profitable during the simulation, were subjected to detailed statistical analysis. There are two types of differences for probability distribution of matches with draw result. These effects are most evident in the leagues in which the number of draw results is very high, eg. in the second division of French and Italian League (more than 30% of draw results).

### 4. Probability distribution of draw

#### 4.1. Draw predictions based on historical data

Statistical analysis showed that the probability of draw match result is not the same throughout the entire football season. In the middle of the football season number of draws is higher than in other periods. In the case of the second division of Italian League, the seasonal difference – between the average probability of draw match result for the whole season and for the matches selected only from the middle of the football season – is four percentage points, which is illustrated in Fig. 1. The 4% difference for draw probability informs about increased value of the expected profit from the bet in the middle of the football season in comparison to other periods.

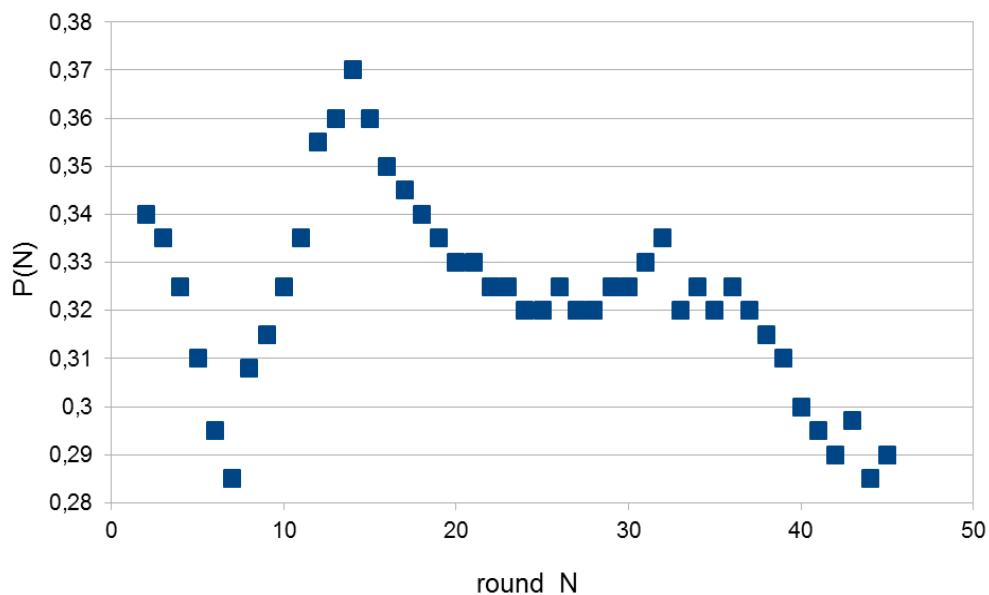


Fig. 1. Probability distribution of draw outcomes in Italian League Series B 2002-2012

#### 4.2. Team condition impact on probability of draw

Statistical analysis showed that there is a correlation between the score of the match, and the results obtained by the team in previous games. The correlation is associated with the condition of a sport team, which can be determined on the basis of several earlier matches. Studies have shown that there is increased probability of the draw result (for about two percentage points) if the result of the previous game played by the same team also was ended with the draw. This important property was used in betting algorithms design, presented in point 5.

#### 4.3. Optimization

For optimization of betting algorithms, one must specify matches  $M$  that are influenced by both the burden of a team condition  $M_F$  and the burden of the time dependence of the occurrence of draw results  $M_T$ . Matches  $M$  are the common part of the overlapping sets  $M_F$  and  $M_T$ , and is the intersection of:

$$M = M_F \cap M_T \quad (3)$$

Possible draw score of the match in the second division of French and Italian League is approximately six percentage points higher than in other divisions and leagues, what is illustrated in Fig. 2.

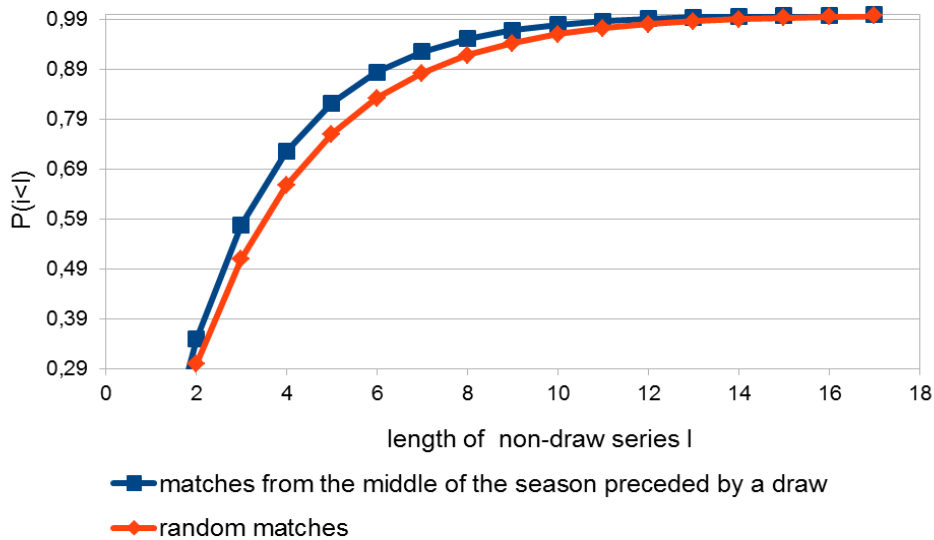


Fig. 2. Probability distributions of non-draw outcomes of a football team.

### 5. Betting algorithms

On the basis of differences in probability distributions, there were formulated two algorithms to improve the value of the expected profit from betting game. The algorithms make bets on the draw results of matches from the middle football season, but played only by the teams with previous matches ended in a draw.

Betting function for Algorithm 1 and Algorithm 2.

```
function betting(Match, Bid):
    Bookmaker = chooseBookmaker(Match)
    bet(Match, Bid, Bookmaker)
```

### Algorithm 1.

```

I = (J-1)*2 // I - number of rounds, J-1 - number of opponent teams
// half of rounds in spring and the other half in autumn
for(i=I/4; i<I*3/4; i++): // iterations for all rounds from middle spring i=I/4
// to middle autumn I*3/4
    for(j=0; j<J; j++): // iterations for all teams
        if score(M[j][i-1]) == DRAW: // match score for team j in preceded round i-1
            betting(M[j][i], bid) // fixed bid

```

The variable  $J$  is the number of teams in league football, the variable  $M$  is the matrix of all the matches played by teams above the league. The function `chooseBookmaker(Match)` searches the bet offers from various bookmaker outlets and chooses the highest course. The function `bet(Match, Bid, Bookmaker)` make a bet with a chosen bookmaker. The second algorithm is modification of the first algorithm, and in contrast to the algorithm 1, is trying to compensate for the failures in selecting draw results by doubling the bid in every next round when making a bet for a particular team. The bet amount will increase until a draw is accurately predicted or until the maximum limit value - expressed in units of  $N$  - is not exceeded.

### Algorithm 2.

```

teams_to_bet = [] // empty table
series_without_draw = [0] * J // list of length J, J - number of teams
I = (J-1)*2 // I - number of rounds
for(i=1; i<I; i++): // first iteration for i=0 skipped to gather historical data
    for(j=0; j<J; j++): // iterations for all teams
        if score(M[j][i-1]) == DRAW:
            teams_to_bet.append( j )
            series_without_draw[ j ]=0
        else:
            series_without_draw[ j ] ++
            if series_without_draw[ j ] > N:
                teams_to_bet.remove( j )
A = length(teams_to_bet) // A - table length
for(k=0; k<A; k++): // determining bid basing on the length of series without draw
    bid = 2^ series_without_draw[ teams_to_bet[ k ] ] // variable bid
    betting(M[teams_to_bet[ k ]][ i ], bid) // betting

```

At the beginning of the season, one have to wait for the results of the first round of matches. If there is a draw, the team lands in a set of teams to bet (table `teams_to_bet`). The procedure is repeated after every round, and in addition is stored the length of a series of non-draws in `series_without_draw` parameter. If the length of the series without draws exceeds  $N$ , then the team is removed from a set of teams to bet. The amount of bid, put into play, depends on the length of the non-draw series to compensate – in case of winning – all cost gained earlier by losing.

According to the simulation, the performance of the algorithms is defined by

$$w = \frac{z}{k} - 1 \quad (4)$$

where  $z$  is the maximum player portfolio balance and  $k$  is the maximum amount that a player had to invest to be able to continue playing for the described algorithm. The performance of algorithm 1 and algorithm 2 is given in Tab. 1.

Tab. 1. Algorithm performance  $w$

|             |              |
|-------------|--------------|
| Algorithm 1 | <i>0.604</i> |
| Algorithm 2 | <i>0.801</i> |

## 6. Practical evaluation of algorithm for real funds

The algorithm 2 has been tested on real financial resources in the online bookmaker outlet. After six months, the game showed sixty percent profitability. But it should be stressed that most of bets were made in Series B of French and Italian League, because of the large ratio of draw results in the middle of seasons.

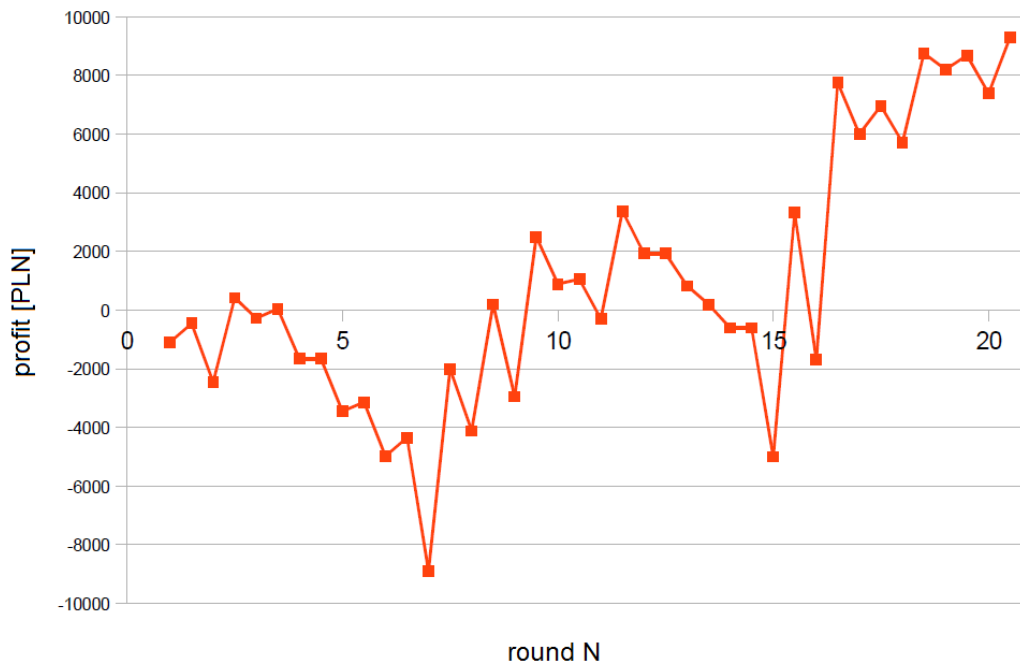


Fig. 3. Profit from betting with the use of algorithm 2 in Series B Italian League in season 2011-2012

According to the simulations carried out on historical data for the second division of the Italian League, the average rate of return from the game with the algorithm is about 25%. The detailed results are illustrated in Fig. 3.

## 7. Summary

The probability that the algorithms will bring loss for two seasons, to only bring profit in the third season, is more than 10%. Another disadvantage of the algorithms eventuates from the nature of the betting exchange. With the increase of the bets, the match rate decreases, thereby reducing the value of the expected profit from it. Thus the algorithm is not scalable.

The probability distribution of draws in football matches is close to the Bernoulli distribution [1-3]. The probability of a draw depends inter alia on the phase of football season and

team performance in previous matches. The presented algorithms select matches, in which there is an increased probability of draw. Their structure is very simple and has many weaknesses. It can be modified by specifying additional conditions for making a choice if the match should be bet. The most important conclusion drawn from the presented research is to demonstrate that it is possible to develop a long-term profit in betting, as the average value of the expected profit from the game presented algorithm is positive. But the both algorithms are not universal – they base on increased probability of draw (as it was in French and Italian League of Series B) – and if the condition is not satisfied the algorithms will end with no bets made.

## **Bibliography**

- [1] D. Moore, D. McCabe, Introduction to the Practice of Statistics, 6<sup>th</sup> ed. W. H. Freeman and Company, New York 2007, pp. 509-523.
- [2] D. Back, A. Bielecki, H. Peng, Stochastic methods in finance. Lecture Notes in Mathematics / Fondazione C.I.M.E., Firenze 2003, pp. 105-109.
- [3] W. Chin, M. Ingenoso, Risk formulae for proportional betting. In Optimal Play: Mathematical Studies of Games and Gambling. S. N. Ethier and W. R. Eadington, eds. Institute for the Study of Gambling and Commercial Gaming, Univ. of Nevada, Reno, 2007, pp. 541-548.
- [4] A. Owen, Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. IMA Journal of Management Mathematics, No. 22, 2011, pp. 99-113.
- [5] I. D. Hill, Association Football and Statistical Inference. Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 23, No. 2, 1974, pp. 203-208.
- [6] M. Moroney, Facts from Figures. Penguin Books, London 1956, p. 147.
- [7] C. Reep, B. Benjamin, Skill and chance in association football. Journal of the Royal Statistical Society, No. 17, London 1968, pp. 24-26.
- [8] L. Fahrmeir, G. Tutz, Dynamic stochastic models for time-dependent ordered paired comparison systems. Journal of the American Statistical Association, Vol. 89, No. 428, 1994, pp. 1438-1449.

## **PREDYKCJA REMISU MECZU PIŁKARSKIEGO DLA STRATEGII GRY W ZAKŁADY SPORTOWE**

Artykuł opisuje usystematyzowane podejścia do gry hazardowej w zakłady bukmacherskie. Przedstawiono systemy gry oraz praktyczne metody analizy statystycznej rynku bukmacherskiego. Rozkład występowania remisów w meczach piłkarskich jest zbliżony do rozkładu Bernoulliego. Korzystając z tej własności, szacowano prawdopodobieństwo wystąpienia remisu, które zależy między innymi od fazy, w której znajduje się sezon piłkarski oraz wyników zespołu w poprzednich meczach. Zaproponowano dwa algorytmy obstawiania remisowych wyników meczów w zakładach bukmacherskich, na podstawie przewidywania zwiększonego prawdopodobieństwa wystąpienia wyniku remisowego w meczach piłkarskich. Zaprezentowane algorytmy selekcjonują mecze, w których występuje zwiększone prawdopodobieństwo remisu. Symulacje przeprowadzono dla danych zawierających informacje o meczach z lat 2002-2010 dla ponad 500 drużyn z 16 krajów, głównie europejskich. Najbardziej wydajny algorytm został sprawdzony w kolekturach bukmacherskich przy użyciu realnych środków finansowych, przynosząc zysk.