

Research

Power Law and q-Exponential Distribution in Football Goalscoring

Župec Matjaž^{1,*}

¹ University of Maribor, Maribor, Slovenia

* Correspondence: Matjaž Župec: matjaz.zupec@student.um.si

Citation: Župec M. Power Law and q-Exponential Distribution in Football Goalscoring. Proceedings of Socratic Lectures. 2026, 14, 76-83.
<https://doi.org/10.55295/PSL.14.2026.III7>

Publisher's Note: UL ZF stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract:

In the present study, we investigate the power law distribution as a model for heavy-tailed statistical distributions. We look at the main parameters of the power law model, as well as statistical methods used to test its applicability. The method is applied to goalscoring data from the top five European leagues from 1977/78 season to 2024/25 season on a season-by-season basis. The p -value is determined, and the results are plotted as a time series. We also consider the Tsallis entropy formalism, and focus on q-exponential distribution, which is a result of maximisation of Tsallis entropy. Particular attention is given to the parameter q . It is fitted to data of top scorers and players with most goals-and-assists (GA) of multiple national leagues across multiple seasons. The obtained values of q are found to be consistent with previously reported results. We also apply both distributions to lists of top scorers from multiple clubs. Finally, we discuss possible connections between the obtained statistical behaviour and tactical or structural aspects of modern football.

Keywords: Power law; Tsallis entropy; Q-exponential distribution; Football; Heavy tail distribution

1. Introduction

As the most popular sport in the world, football has long been the subject of research by many scientists, including physicists. Numerous studies have investigated statistical and tactical aspects of the sport. In (Greenthrough et al., 2002), distributions of goals scored by home teams, away teams, and both teams were analysed. In (Velasco Stock et al., 2022) an algorithm was adapted to predict the results of football matches. More recently, (Chacoma & Billoni, 2025) attempted to characterise playing styles using data-driven methods. In this article, we focus on the statistical properties of goalscoring. Specifically, we analyse goalscoring distributions among players and fit two candidate models: the power law distribution and the q-exponential distribution. Both distributions are particularly suitable for systems with heavy-tailed statistical behaviour, where extreme events occur more frequently than predicted by standard exponential models.

1.1. Power law

A power law is often used to describe systems with large variability spanning several orders of magnitude. We should first look into the cases where the power law would be applicable. For example, human height is concentrated around a characteristic value, with deviations larger than a factor of two being practically non-existent. In such cases, a power law description is unnecessary. In contrast, quantities such as city populations span several orders of magnitude, making power law models considerably more appropriate. In Section 2, we introduce the key parameters of the power law distribution and discuss statistical methods for testing its applicability. Power laws have previously been observed in football-related data; for instance, (Da Silva et al., 2013) showed that the relationship between team rank and points follows power law behaviour. We apply the power law model to the distribution of goals scored by players in the top five European leagues between the 1977/78 and 2024/25, as well as to all-time scoring records of multiple clubs.

1.2. q-exponential distribution

An alternative approach to modelling the observed distributions is provided by the q-exponential distribution, which generalises the standard exponential distribution. Just as the exponential distribution can be obtained, the q-exponential distribution is derived by maximising the Tsallis entropy, a generalisation of the standard Boltzmann-Gibbs entropy. The formalism was first proposed by Constantino Tsallis in 1988. We apply the q-exponential distribution to goalscoring distributions in multiple national leagues, as well as to all-time scoring records of multiple clubs.

2. Material and Methods

2.1. Power law

Let x be a quantity whose distribution we are interested in. Then we can describe a continuous probability distribution $p(x)$ such that:

$$p(x)dx = \Pr(x \leq X < x + dx) = Cx^{-\alpha}dx, \quad (1)$$

where X is the observed value, C is the normalization constant and α is the scaling exponent. However, since we are working with numbers of goals, we will need a discrete distribution. Here we write:

$$p(x) = \Pr(x = X) = Cx^{-\alpha}. \quad (2)$$

We must also note that this density diverges as $x \rightarrow 0$ (Clauset et al., 2009), so our distribution cannot hold for all non-negative values of x . We will define the lower bound for the power law behaviour by $x_{\min}$. One more thing, it is sometimes more useful to define a complementary cumulative distribution function (CCDF), defined as $P(x) = \Pr(X \geq x)$.

We will now look into estimating the values of α and $x_{\min}$. We first estimate α using the maximum likelihood estimator (MLE). The MLE for the continuous case is (Clauset et al., 2009):

$$\hat{\alpha} = 1 + n \left[\sum_{i=1}^n \ln \frac{x_i}{x_{\min}} \right]^{-1}. \quad (3)$$

For the discrete case, there is no closed-form expression. There is, however, an approximation, which works similarly to Eq. (3) and can be written as (Clauset et al., 2009):

$$\hat{\alpha} \approx 1 + n \left[\sum_{i=1}^n \ln \frac{x_i}{x_{\min} - \frac{1}{2}} \right]^{-1}. \quad (4)$$

This expression is found to be accurate to 1 % when $x \geq 6$.

We also need an accurate value for $x_{\min}$. If it is set too low, the fit includes data that do not follow a power law. If it is set too high, we lose good data. According to the method presented in (Clauset et al., 2007), we compare the distance between our observed data and the best-fit power law model. We use the Kolmogorov–Smirnov (KS) statistics, which measures the maximum distance between the two distributions:

$$D = \max_{x \geq x_{\min}} |S(x) - P(x)|. \quad (5)$$

Here $S(x)$ is the empirical distribution and $P(x)$ is the best-fit model power law. The optimal value for $x_{\min}$ is the one that minimises D .

Now we can calculate both of the desired parameters, but we still need to know whether the power law hypothesis is applicable to a given dataset. For confirmation, many studies rely on visual inspection, but this can be tricky, as many models can be visually similar to the power law. A better approach would be to use a goodness of fit test. The idea is to measure the difference between the data and the best-fit model, using the KS statistics. We also measured the difference between the synthetic data sets drawn from the same model. Then we defined the p -value as the fraction of the difference between the synthetic data sets that were larger than the difference between the data and the best-fit model. If the p -value were large, i.e. close to 1, then the power law hypothesis would be considered plausible. If the p -value were small, then our data would very unlikely follow the power law. We adopted a cut-off value of 0.1, the same as (Clauset et al., 2009). Some authors took the cut-off value of 0.05, but it is argued in (Clauset et al., 2009) that this would be too lenient, as it would let through some datasets that are very unlikely to follow the power law. We also noted that the p -value defined above did not agree with the results of some experimental studies. Therefore we wished to test our hypothesis stating that a small p -value means that the null hypothesis is very unlikely to be correct. Therefore, small p -values are desired. Here, the p -value quantifies the plausibility of the power law hypothesis itself, so we desire larger values.

2.2. q -exponential distribution

The second model we consider is the q -exponential distribution, obtained by maximising the Tsallis entropy, which was proposed by Constantino Tsallis in 1988 (Tsallis, 1988):

$$S_q = k \frac{1 - \sum_{i=1}^W p_i^q}{q-1}, \quad (6)$$

where k is the Boltzmann constant, W is the total number of possible microscopic configurations, p_i are the associated probabilities and q is a real number. In the limit of $q \rightarrow 1$, this expression reduces to the Boltzmann–Gibbs entropy.

We are interested in the q -exponential distribution, written in the form (Malacarne & Mendes, 2000):

$$N(x) = N_0 (1 - (1 - q)\lambda x)^{\frac{1}{1-q}}, \quad (7)$$

where $N(x)$ is the number of players with x goals, and N_0 , λ and q are real parameters. In the limit of $q \rightarrow 1$, the q -exponential reduces to the standard exponential function. But since the parameters N_0 and λ are not directly known, we replace them with the ones we can

determine from our data. We can evaluate the equation (7) at $x = 1$ (Malacarne & Mendes, 2000):

$$N(1) = N_0(1 - (1 - q)\lambda)^{\frac{1}{1-q}}, \quad (8)$$

which allows us to express N_0 . We can also calculate the average number of goals scored by all players with at least one goal scored. We denote this quantity by m (Malacarne & Mendes, 2000):

$$m = \frac{\int_0^\infty xN(x)dx}{\int_0^\infty N(x)dx} = \frac{1}{\lambda(3-2q)}, \quad (9)$$

which can be used to express λ . Now we can rewrite equation (7) as:

$$N(x) = N(1) \left(\frac{1 - \frac{1-q}{(3-2q)m}}{1 - \frac{1-q}{(3-2q)m}x} \right)^{\frac{1}{q-1}}. \quad (10)$$

The only remaining free parameter is q , which is determined by fitting the model to the data. The key difference between the standard exponential distribution and the q -exponential distribution is that the latter decays more slowly for large values of x . Consequently, the q -exponential distribution assigns a higher probability to extreme events and is therefore better suited for describing heavy-tailed distributions, such as the distribution of goals scored. Larger values of q correspond to heavier tails and therefore to a higher probability of extreme events.

3. Results

We now apply both distributions to the collected data. The data were collected from publicly available databases provided by Transfermarkt (2025) and MNZ Maribor (2025).

3.1 Power law

We fitted the power law model to the distribution of the number of players that scored x goals in the top five European national leagues for every season from 1977/78 to 2024/25. Only the goals scored in national leagues and national and international cups were considered. The analysis was performed using the Python package *Powerlaw* (Alstott et al., 2014), which made our work easier. We analysed the p -values for individual seasons and their evolution over time. **Figure 1** shows the development of p -values. Out of 48 seasons, 24 produced distributions were consistent with a power law, while for the remaining 24 the hypothesis was rejected. More recent seasons appear to follow a power law distribution more frequently than the earlier seasons. Since 2004, only 6 seasons were found to deviate from the power law. Additionally, the most recent 7 seasons all follow the power law.

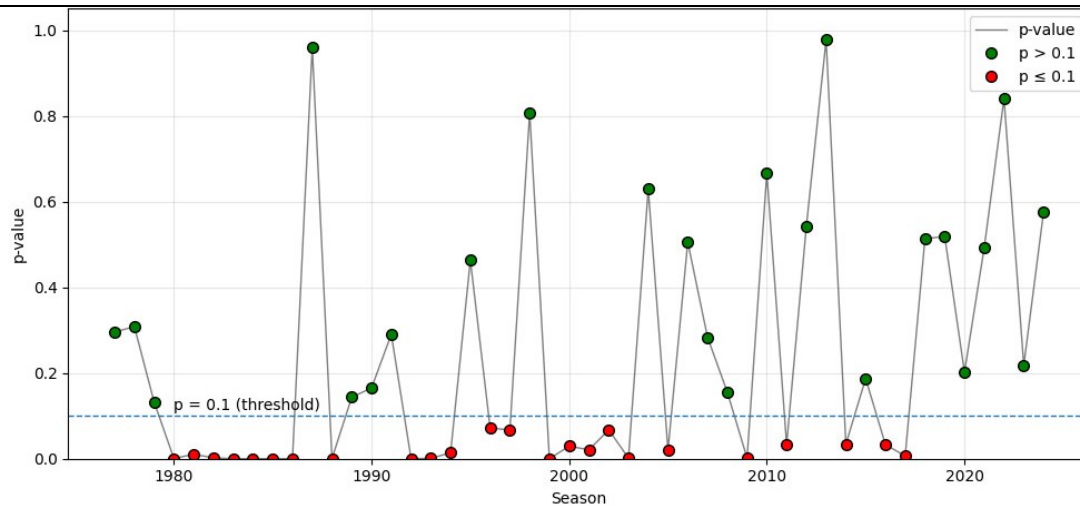


Figure 1. p -values through seasons from 1977/78 to 2024/25.

3.2 q -exponential distribution

We fitted the distribution of the number of players that scored x amount of goals in a season across multiple seasons of the English Premier League, Spanish La Liga, Slovenian Prva Liga, and the First MNZ Maribor League (the fourth tier of Slovenian football, also known as Golgeter Premium Liga) with expression (10). Only the goals scored in the leagues themselves were considered. As discussed previously, we paid special attention to parameter q . Previous studies found that $q = 1.33$ provides a good fit for multiple national leagues (Malacarne & Mendes, 2000). We tested this across multiple seasons of the aforementioned leagues (Figure 2). For all analysed leagues, the fitted values of q (also referred to as the Tsallis parameter) was close to 1.33, consistent with previous findings. The mean values were: 1.34 for the Premier League, 1.32 for LaLiga, 1.34 for Prva Liga and 1.35 for MNZ Maribor.

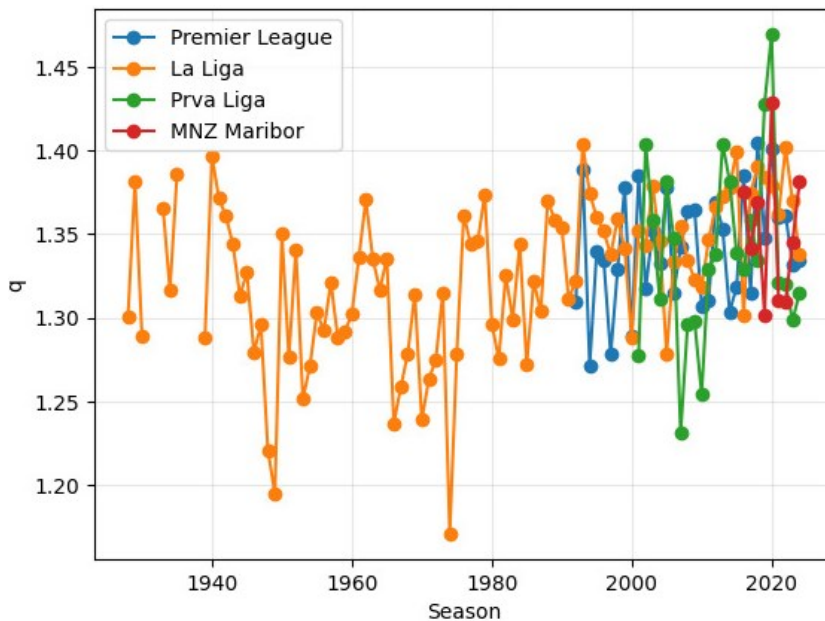


Figure 2. The values of q through seasons for Premier League, LaLiga, Prva Liga and MNZ Maribor. The two outliers from LaLiga are due to the small sample size of players during those times. The gap in LaLiga around 1940 is due to the civil war.

We also applied Eq. (10) to the distribution of goals and assists accumulated by players within a single season (further referred to as GA), as GA has become an increasingly important performance metric in modern football analysis. We then compared the fitted q -values obtained for goals alone with those obtained for GA. We also replaced MNZ Maribor with the Italian Serie A, since assist statistics are unavailable for the former league. The results are shown in **Figure 3**. At first sight, the values of q for GA are a bit smaller than for goals. The averages this time are 1.31 for the Premier League, LaLiga, and Prva Liga and 1.33 for Serie A.

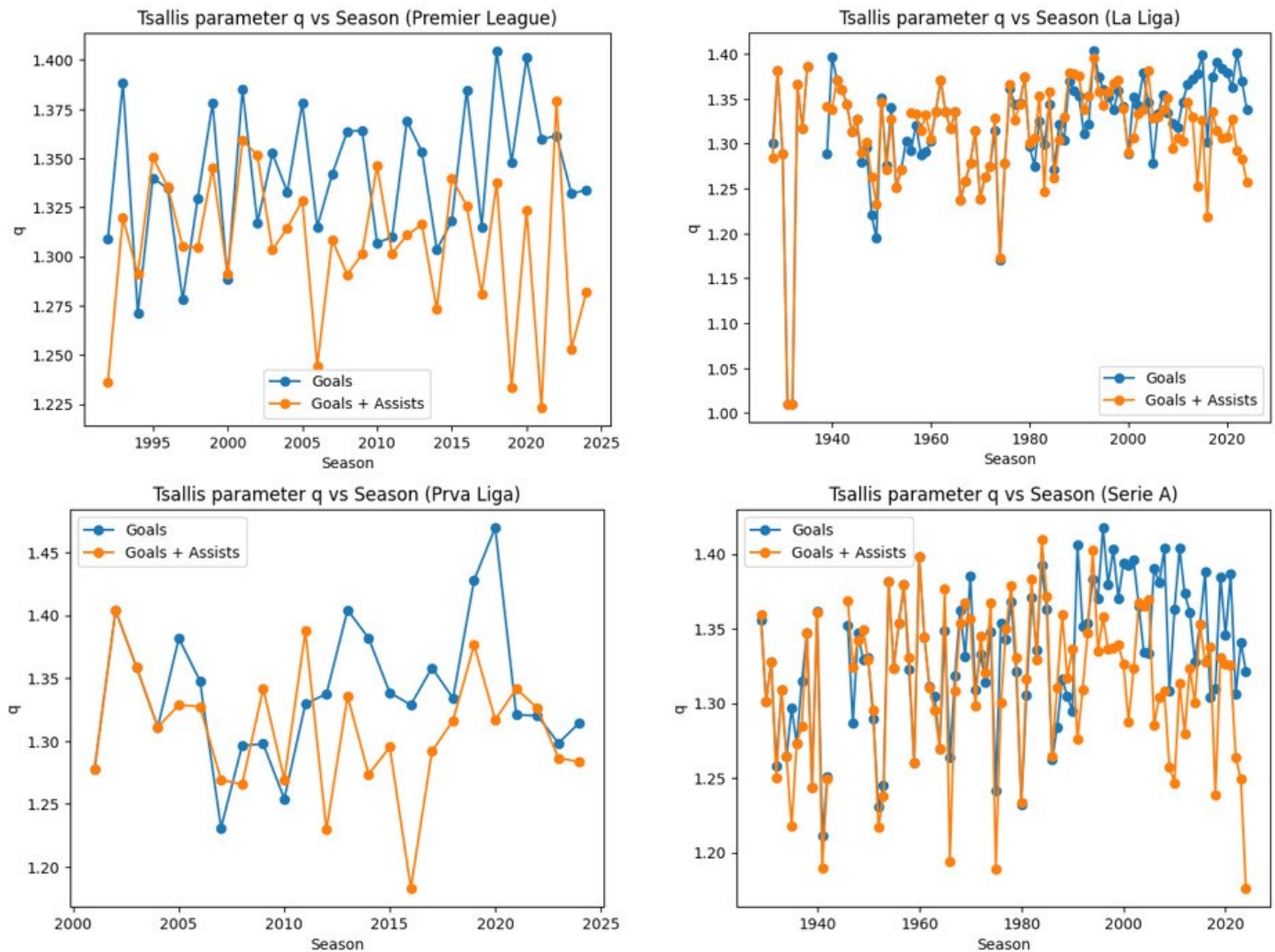


Figure 3. The values of q through seasons for Premier League, LaLiga, Prva Liga and Serie A. The two outliers from LaLiga are due to the small sample size of players during those times. The gaps in LaLiga and Serie A around 1940 are due to the civil war and the Second world war, respectively.

3.3 Scoring lists from clubs

We also fitted both models to the distributions of all-time goals scored for multiple clubs. The results are summarised in **Table 1**. The p -values indicate that only a small number of clubs can be adequately described by a power law distribution. The fitted q -values are generally larger than those obtained for seasonal goalscoring distributions.

Table 1. The values of p and q , obtained from all-time scoring lists of multiple clubs.

Club	p	q
Real Madrid	0.03	1.468
FC Barcelona	0.00	1.459
Atletico Madrid	0.04	1.449
Liverpool	0.13	1.451
Manchester United	0.01	1.449
Arsenal	0.00	1.451
Chelsea	0.09	1.446
Manchester City	0.00	1.437
Inter Milan	0.33	1.430
AC Milan	0.03	1.432
Juventus	0.02	1.437
Bayern München	0.07	1.457
Borussia Dortmund	0.04	1.437
PSG	0.28	1.437
Ajax Amsterdam	0.00	1.442
NK Maribor	0.00	1.41
NK Olimpija	0.00	1.42
NK Celje	0.00	1.37
Stoke City	0.22	1.418
Birmingham City	0.00	1.418
Exeter City	0.00	1.42
Peterborough	0.00	1.437
Al Nassr	0.56	1.37
Los Angeles Galaxy	0.00	1.435

4. Discussion

We now discuss the implications of our results. If we look back to **Figure 1**, we can see that in years after 2000, there have been considerably more seasons in which the distribution of goals scored by players can be described by a power law distribution. One possible explanation is the increasing tactical optimisation of modern football, where attacking patterns and player roles have become more structured and statistically efficient. Such optimisation may contribute to a stronger concentration of goals among a smaller subset of highly productive players, leading to distributions that more closely resemble a power law. Another possible contributing factor is the introduction of the video assistant referee (VAR) system. Interestingly, all seasons following the introduction of VAR in 2018 were found to be consistent with power law behaviour.

Turning to the q -exponential distribution, we examine the differences between the fitted values of q for goalscoring distributions (q_g) and distributions of GA (q_{ga}) (**Figure 3**). We can see that there are periods when q_g is noticeably larger than q_{ga} . One possible interpretation is that the difference between q_g and q_{ga} may be related to the overall competitive strength of a league. For Premier League, the difference became big in the late 2010s, during periods in which English clubs were particularly successful in European competitions, as well as in the late 2000s, when its clubs were also performing well in European competitions. Similar trends appear in the other analysed leagues. LaLiga has been considered as arguably the best league in the world since the late 2000s, when the difference between

both values of q appeared. For Serie A the difference appeared in the 1990s and early 2000s, during the period in which Italian clubs achieved some of their strongest European results, and for Prva Liga, the gap appeared in 2010s, during the league's most successful period in European qualification competitions.

The fitted values of q for individual clubs (**Table 1**) suggest that the most successful clubs tend to exhibit larger values of q . Real Madrid, FC Barcelona, and several other elite clubs have values around 1.45 or 1.46, while some less successful clubs have values around 1.42 or even lower. In terms of the q -exponential model, larger values of q correspond to heavier tails, indicating a broader distribution of high-scoring players.

5. Conclusions

We have presented the theoretical background for the power law and q -exponential distribution in the model for football goalscoring. The distributions were fitted to the number of goals scored by players across multiple seasons and leagues, and the evolution of the fitted parameters through time was analysed. For the top five European leagues, the power law model was found to be applicable in approximately half of the analysed seasons. Furthermore, power law behaviour appeared more frequently in recent years, which may reflect increasing tactical and statistical optimisation in modern football. The q -exponential distribution was applied to both goalscoring and goals-and-assists (GA) distributions across several national leagues. The obtained values of the Tsallis parameter q remained close to previously reported values, while the observed differences between the parameters for goals and GA may be related to the competitive strength and overall quality of a league. Both models were applied to all-time club scoring records. Only a small number of clubs produced a distribution consistent with the power law, while the fitted q -values were generally larger for more successful clubs, indicating heavy-tailed scoring distributions.

References

1. Alstott J, Bullmore E, Plenz D. powerlaw: a Python package for analysis of heavy-tailed distributions. PLOS ONE. 2014; 9: e95816. DOI: 10.1371/journal.pone.0095816.
2. Chacoma A, Billoni OV. Data-driven approach to defining football styles in major leagues. Chaos, Solitons & Fractals. 2025; 200: 116926. DOI: 10.1016/j.chaos.2025.116926.
3. Clauset A, Shalizi CR, Newman MEJ. Power-Law Distributions in Empirical Data. Siam Review. 2009; 51: 661-703. DOI: 10.1137/070710111.
4. Clauset A, Young M, Gleditsch KS. On the Frequency of Severe Terrorist Events. Journal of Conflict Resolution. 2007; 51:58-87. DOI: 10.1177/0022002706296157.
5. Da Silva S, Matsushita R, Silveira E. Hidden power law patterns in the top European football leagues. Physica A. 2013; 392: 5376-5386. DOI: 10.1016/j.physa.2013.07.008.
6. Greenthrough J, Birch PC, Chapman SC, Rowlands G. Football goal distributions and extremal statistics. Physica A. 2002; 316: 615-624. DOI: 10.1016/S0378-4371(02)01030-0.
7. Malacarne LC, Mendes RS. Regularities in football goal distributions. Physica A. 2000; 286: 391-395. DOI: 10.1016/S0378-4371(00)00363-0.
8. MNZ Maribor, 2025. *Official competition statistics*. Available from: <https://www.mnzmaribor.si>.
9. Transfermarkt. 2025. *Football statistics and market values*. Available from: <https://www.transfermarkt.com>.
10. Tsallis C. Possible Generalization of Boltzmann-Gibbs Statistics. Journal of Statistical Physics. 1988; 52: 479-487. DOI: 10.1007/BF01016429.
11. Velasco Stock E, da Silva R, Fernandes HA. A physics-based algorithm to perform predictions in football leagues. Physica A: Statistical Mechanics and its Applications. 2022; 600: 127532. DOI: 10.1016/j.physa.2022.127532.