Physica A 398 (2014) 56–64

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Universality in the distance between two teams in a football tournament

Roberto da Silva *, Silvio R. Dahmen

Instituto de Física, Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves 9500, Caixa Postal 15051, 91501-970 Porto Alegre RS, Brazil

HIGHLIGHTS

- We studied the statistics related to the soccer scoring process.
- We elaborated an agent-based model which is able to reproduce soccer statistics.
- We analyzed the fluctuations of the distance between two teams in soccer leagues.

ARTICLE INFO

Article history:
Received 8 December 2013
Available online 18 December 2013

Keywords:
Football statistics
Stochastic process
Fitting and universality

ABSTRACT

Is football (soccer) a universal sport? Beyond the question of geographical distribution, where the answer is most certainly yes, when looked at from a mathematical viewpoint the scoring process during a match can be thought of, in a first approximation, as being modeled by a Poisson distribution. Recently, it was shown that the scoring of real tournaments can be reproduced by means of an agent-based model (da Silva et al. (2013) [24]) based on two simple hypotheses: (i) the ability of a team to win a match is given by the rate of a Poisson distribution that governs its scoring during a match; and (ii) such ability evolves over time according to results of previous matches. In this article we are interested in the question of whether the time series represented by the scores of teams have universal properties. For this purpose we define a distance between two teams as the square root of the sum of squares of the score differences between teams over all rounds in a double-round-robin-system and study how this distance evolves over time. Our results suggest a universal distance distribution of tournaments of different major leagues which is better characterized by an exponentially modified Gaussian (EMG). This result is corroborated by our agent-based model.

1. Introduction

In recent years few topics in theoretical physics have experienced an increase in interest like complex systems have. This is in part due to their range of applicability and intrinsically multidisciplinary flavor [1,2]. The topics studied range from more traditional ones like phase transitions and critical phenomena in equilibrium and non-equilibrium systems [3–6] to more eclectic applications such as econophysics [7,8] and sociophysics/evolutionary game theory [9] among others.

Recently, an interesting application has drawn the attention of many physicists: sport statistics (see for example [10–16]), and more particularly football (see for instance [17–25]), a sport whose international federation, FIFA, can boast to having more than 200 confederation members.

Several questions can be formulated, the scoring process being perhaps the most natural one. Several authors [20–23] have claimed that the probability of a team to score n goals per match should follow a Poisson distribution $(\lambda^n/n!)e^{-\lambda}$, where

* Corresponding author. Tel.: +55 51 8196 0903.
E-mail addresses: rdasilva@if.ufrgs.br, rdasilva10e5@gmail.com (R. da Silva).

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.12.008

λ is the average number of goals of the team, or modifications/extensions of this distribution. Since a team usually has to try many shots before actually scoring a goal, this idea is quite reasonable. But how close does this come to the scores of real tournaments?

In order to answer this question, we recently proposed an agent-based model whose goal was to incorporate some refinements which could account for the time evolution of teams during a championship [24]. The main idea behind our model is quite simple: n teams playing in a tournament according to a double-round-robin-system (henceforth DRRS): each team plays the other $n-1$ teams twice during the whole tournament, i.e. team A plays B in the first half of the competition and B plays A in the second half. For the purpose of modeling, no distinction is made between home and away matches (the "home court advantage" could be easily inserted into the model, although it was checked to be irrelevant for the statistics). In a game between team i and team j, the probability that i beats j (represented by $i \succ j$) in the tth round is given by

$$\Pr(i \succ j, t) = \left[1 - r_{draw}^{(i,j)}(t)\right] \cdot \frac{\varphi_t^{(i)}}{(\varphi_t^{(i)} + \varphi_t^{(j)})} \tag{1}$$

where $\varphi_t^{(i)}$, $i = 1, \ldots, n$ is the potential of the ith team in round t. The most sophisticated version of the model we shall use in this paper adopts the so-called **prescription III** (for other prescriptions see e.g. [24]), where as initial condition one considers $\varphi_0^{(i)}$ = average number of goals of ith team classified in a real tournament. In this expression $r_{draw}^{(i,j)}(t)$ represents the probability of a draw between i and j. This can be calculated based on the teams' scores during a game according to a Poisson distribution whose rates are their potentials $\varphi_t^{(i)}$ and $\varphi_t^{(j)}$. So, given their rates, the draw probability is the probability that both teams score the same number of goals ($n_i = n_j = n$), which is calculated by

$$\begin{aligned} r_{draw} &= \Pr\left[(n_i = n_j)|(\varphi_t^{(i)}, \varphi_t^{(j)})\right] \\ &= \sum_{n=0}^{\infty} \frac{(\varphi_t^{(i)}\varphi_t^{(j)})^n}{n!^2} \mathrm{e}^{-(\varphi_t^{(i)}+\varphi_t^{(j)})} \\ &= \mathrm{e}^{-(\varphi_t^{(i)}+\varphi_t^{(j)})} I_0\left(2\sqrt{\varphi_t^{(i)}\varphi_t^{(j)}}\right) \end{aligned}$$

where

$$I_\nu(z) = \left(\frac{1}{2}z\right)^\nu \sum_{k=0}^{\infty} \frac{\left(\frac{1}{4}z^2\right)^k}{k!\Gamma(\nu + k + 1)}$$

is the modified Bessel function of the first kind. In case of victory a team has its potential incremented by its own potential divided by the number of rounds in the tournament, $k = 2(n-1)$, i.e., $\varphi \to \varphi + \varphi/k$, and the team gains 3 points. If it loses, its potential is decreased through a similar prescription, $\varphi \to \varphi - \varphi/k$, losing also 3 points. In the case of a draw, the teams gain 1 point each and their potential remains unchanged.

A particular case of this algorithm considers that the abilities of teams and the draw probability do not change over time, i.e, $\varphi_t^{(i)} = \varphi$ and $r_{draw}^{(i,j)}(t) = r$ with φ and r constants. From this we get $p_{win} = \Pr(i \succ j, t) = p_{loss} = \Pr(i \prec j, t) = (1-r)/2$ and $p_{draw} = \Pr(i = j, t) = r$, henceforth called the mean-field approximation. By denoting the probability $f(x, t)$ as the probability of a team to score $x = 0, \ldots, 6(n-1)$ in round $t = 1, \ldots, 2(n-1)$ of the tournament, we have

$$f(x, t) = \frac{(1-r)}{2}\left[f(x-3, t-1) + f(x, t-1)\right] + rf(x-1, t-1). \tag{2}$$

The corresponding continuous approximation of this equation leads us to the PDE:

$$\frac{\partial f}{\partial t} = -\frac{(3-r)}{2}\frac{\partial f}{\partial x} + \frac{(1-r)}{2}\left(3\frac{\partial^2 f}{\partial x^2} - \frac{\partial^3 f}{\partial x^3}\right) \tag{3}$$

whose solution is given by

$$f(x, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \cos\left\{\frac{(r-1)}{2}k^3 t + \left[\left(\frac{3-r}{2}\right)t - x\right]k\right\} \mathrm{e}^{-\frac{3}{2}(1-r)^2 k^2 t} \mathrm{d}k. \tag{4}$$

We note that this equation depends on a unique parameter: the draw probability r which is computed directly from the championship under consideration by averaging the number of draws of all teams.

This mean-field approximation reproduces, as expected, some features of real data, such as e.g. the drift velocity $C = \frac{(3-r)}{2} \approx 1.37$, and the exponent ξ for the second moment of the score, which behaves as $\langle x^2\rangle_t \propto t^\xi$, $\xi \approx 1.80$ as compared to $1.88 < \xi < 1.96$ obtained from real data [24]. Another feature which seems intrinsic to some sports is anomalous diffusion behavior, $\sigma(t) = \left(\langle x^2\rangle_t - \langle x\rangle_t^2\right)^{1/2} = \sqrt{D}t^\beta$: as shown in Ref. [26], cricket has an exponent $\beta \approx 0.65$. Our mean-field approximation fails to show this behavior in football, but a fuller version of our agent-based model, where