Physica A 394 (2014) 304–311

Contents lists available at ScienceDirect

Physica A

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

Memory and long-range correlations in chess games

Ana L. Schaigorodsky [a], Juan I. Perotti [c], Orlando V. Billoni [a,b,*]

[a] Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba, Ciudad Universitaria, 5000 Córdoba, Argentina

[b] Instituto de Física Enrique Gaviola (IFEG–CONICET), Argentina

[c] Department of Biomedical Engineering and Computational Science (BECS), Aalto University, Rakentajanaukio 2, Otaniemi campus, Espoo, FI-00076 Aalto, Finland

HIGHLIGHTS

- We introduced two mapping rules for building discrete time series from a set of chess games.
- We found long range correlations in an extensive chess database.
- The extent of long range correlations depends on the level of the players.
- Depending on the level of expertise the players use different strategies.

ARTICLE INFO

Article history:
Received 2 July 2013
Received in revised form 21 August 2013
Available online 27 September 2013

Keywords:
Long-range correlations
Zipf's law
Interdisciplinary physics

ABSTRACT

In this paper we report the existence of long-range memory in the opening moves of a chronologically ordered set of chess games using an extensive chess database. We used two mapping rules to build discrete time series and analyzed them using two methods for detecting long-range correlations; rescaled range analysis and detrended fluctuation analysis. We found that long-range memory is related to the level of the players. When the database is filtered according to player levels we found differences in the persistence of the different subsets. For high level players, correlations are stronger at long time scales; whereas in intermediate and low level players they reach the maximum value at shorter time scales. This can be interpreted as a signature of the different strategies used by players with different levels of expertise. These results are robust against the assignation rules and the method employed in the analysis of the time series.

1. Introduction

Some games are complex systems which provide information about social and biological processes [1–12]. They are currently being studied because of the existence of very well documented registers of games which are useful for statistical analysis. In particular, recent works have tried to understand the statistics of wins and losses in baseball teams [6], the final standing in basketball leagues [7], the distribution of career longevity in baseball [8], the football goal distribution [9], and face to face game rank distributions [10]. The time evolution of the table during a season, for instance, can be interpreted as a random walk [11], and long-range correlations have been found in the score evolution of the game of cricket [12]. Among them, the game of chess has a main place in occidental culture because its intrinsic complexity is viewed as a symbol of intellectual prowess. Since the skill level of chess players can be correctly identified [13] chess has contributed to the

* Corresponding author at: Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba, Ciudad Universitaria, 5000 Córdoba, Argentina. Tel.: +54 351 4334051.

E-mail addresses: alschaigorodsky@gmail.com (A.L. Schaigorodsky), perotti@aalto.fi (J.I. Perotti), billoni@famaf.unc.edu.ar (O.V. Billoni).

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.09.035

scientific understanding of expertise [5]. In addition, nowadays, there is a big world-wide community of chess players which makes the game a benchmark for studying, for instance, decision making processes [4] and population level learning [3].

Exploring a chess database Blasius and Tönjes [2] observed that the pooled distribution of all opening weights follows a Zipf law with universal exponent. This is a remarkable result [14,15] since the Zipf law is followed in a range that comprises six orders of magnitude. In their work, the authors explain the results using a multiplicative process with a branching ratio distribution which resembles the real branching distribution measured in the database. Beyond the very good agreement between their empirical observations and the proposed model to explain Zipf's law, they do not provide an explanation of the evolution of the pool of games; disregarding for instance, the question of possible memory effects between different games. The development of a given chess game should depend on the expertise of the players because the first moves (game opening) are usually known in advance by high level players as a result of their theoretical background. Other aspects can influence the development of a chess game according to the level of expertise. It has been established that skillful players develop outstanding rapid object recognition abilities that differentiate them from the non-expert players [16,17]. In particular, Gobet and Simon [17] suggested that chess players, like experts in other fields, use long-term memory retrieval structures or templates in addition to chunks [18,19] in short-term memory in order to store information rapidly. Then, it can be expected that the differences in the level of expertise of chess players are reflected in the historical development of chess databases introducing correlations between games through memory effects.

Currently there is a big corpus of digitized texts which allows the study of cultural trends tracing memory effects [20], opening a new branch in science known as culturomics [21]. In particular, long-range correlations have been observed in literary corpora [22,23] where Zipf's law was also studied using extensive databases [24]. Testing signatures of memory effects is important because it has been shown that systems which exhibit Zipf's law need a certain degree of coherence [25] for its emergence. Therefore, the detection of long-range memory effects can be useful to shed light on the general mechanism behind the Zipf law.

In this work we explored the existence of long-range correlations in game sequences. To that end, time-series are constructed using a chronological ordered chess database similar to the one used by Blasius and Tönjes. In order to support the reliability of our results, we used two mapping rules to build the discrete time series. We also used different techniques to detect long-range correlations, namely the rescaled range analysis and the detrended fluctuation analysis [26].

2. Long-range correlation analysis

Rescaled range analysis (R/S) and detrended fluctuation analysis (DFA) are two useful tools for detecting long-range correlations in discrete time series. The R/S analysis was introduced by Hurst when studying the regularization problem of the Nile River [26], whereas DFA was introduced by Peng et al. [27] for detecting long-range (auto-) correlations in time series with non-stationarities. Both methods use the accumulated time series which can be thought as the displacement of a one dimensional random walker whose steps are dictated by the values of the temporal series.

As a general procedure, the discrete time series $X(t)$ is first centered and normalized,

$$\tilde{X}(t) = \frac{X(t) - \langle X(t) \rangle}{\hat{\sigma}} \tag{1}$$

where $\hat{\sigma} = \sqrt{\langle X(t)^2 \rangle - (\langle X(t) \rangle)}^2$ and $\langle \ldots \rangle$ means arithmetic averages over the complete series. Normalization is useful in order to compare different assignation rules [22]. Then the accumulated series $Y(t)$ is constructed,

$$Y(t) = \sum_{u=1}^{t} \tilde{X}(u). \tag{2}$$

Once the series $Y(t)$ is obtained, we applied the rescaled range analysis [22,26] and the detrended fluctuation analysis [28–30]. In the R/S method the time series $\tilde{X}(t)$ is divided into non-overlapping intervals of size n. The range of each interval is computed,

$$R(n) = \max[Y(1), Y(2), \ldots, Y(n)] - \min[Y(1), Y(2), \ldots, Y(n)], \tag{3}$$

and the corresponding standard deviation,

$$S(n) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - m)^2}, \tag{4}$$

where m is the average of $X(t)$ in the interval of size n. Then the average of the rescaled range $E[R(n)/S(n)]$ over all the intervals of size n is calculated. The Hurst exponent H is obtained by varying the size of the intervals and fitting the data to the expression Cn^H where H is the Hurst exponent.