

Memory Kernel in the Expertise of Chess Players

Ana L. Schaigorodsky,^{1,*} Juan I. Perotti,^{2,†} and Orlando V. Billoni^{1,‡}

¹*Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba and Instituto de Física Enrique Gaviola (IFEG-CONICET), Ciudad Universitaria, 5000 Córdoba, Argentina*

²*IMT Institute for Advanced Studies Lucca, Piazza San Francesco 19, I-55100, Lucca, Italy*

(Dated: April 27, 2015)

In this work we investigate a mechanism for the emergence of long-range time correlations observed in a chronologically ordered database of chess games. We analyze a modified Yule-Simon preferential growth process proposed by Cattuto et al., which includes memory effects by means of a probabilistic kernel. According to the Hurst exponent of different constructed time series from the record of games, artificially generated databases from the model exhibit similar long-range correlations. In addition, the inter-event time frequency distribution is well reproduced by the model for realistic parameter values. In particular, we find the inter-event time distribution properties to be correlated with the expertise of the chess players through the memory kernel extension. Our work provides new information about the strategies implemented by players with different levels of expertise, showing an interesting example of how popularities and long-range correlations build together during a collective learning process.

PACS numbers: 89.75.-k, 05.45.Tp, 02.50.-r

Keywords: Long-range correlations, Time series, Inter-event time

I. INTRODUCTION

Human behavior is generally complex. Considerable amount of resources are, and have been, dedicated to characterize and understand how humans behave as individuals, or collectively¹⁻⁷. The struggle of humans when playing well defined games according to some set of rules, provides a convenient experimental setup to understand human behavior and decision making process⁸⁻¹⁹. This is particularly convenient from the physics' point of view, because under the rules of a game, the huge number of degrees of freedom the systems have are highly constrained and more easily quantifiable. Moreover, very well documented records of games are constantly created and updated in electronic format, facilitating the statistical analysis.

For instance, recent studies have tried to understand the statistics of wins and losses in baseball teams¹³, the final standing in basketball leagues¹⁴, the distribution of career longevity in baseball¹⁵, the football goal distribution¹⁶, and face to face game rank distributions¹⁷. Useful parallels can be established between the statistical patterns of game-based human behavior, and well defined theories of physical processes. For example, the time evolution of the results table during a season can be interpreted as a random walk¹⁸, and long-range correlations have been found in the score evolution of the game of cricket¹⁹. The game of chess, which is viewed as a symbol of intellectual prowess, is not the exception^{9,20,21}. Since the skill level of chess players can be correctly and quantitatively identified²², chess has contributed to the scientific understanding of expertise¹². In addition, there is a big world-wide community of chess players, which makes the game a benchmark for studying decision making processes¹¹ and population level learning¹⁰. In particular, the study of long-range correlations in systems

where Zipf's law is present is also important, since it has been observed in literary corpora^{23,24} and because systems which exhibit Zipf's law need a certain degree of coherence²⁵ for its emergence.

In the game of chess each possible move sequence²⁶ can be mapped as one directed path in a corresponding game tree. Here, the root node is the initial position of the game, the moves are represented by the edges, and there is a one-to-one correspondence between move sequences and vertices. The game depth d is the topological distance, or shortest path length, between the root and the node reached after the last move. Exploring chess databases, Blasius and Tönjes⁹ observed that the pooled distribution of chess opening weights follows a Zipf law with universal exponent. This is a remarkable result^{27,28} where the Zipf law holds over six orders of magnitude. They also established that the popularity distribution follows a power law in which the exponent depends on the depth d of the game tree. They explained these findings in terms of an analytical treatment of a multiplicative process. In a recent paper²⁰, we study the dynamics of the game tree growth. More specifically, we find that the emerging Zipf and Heaps laws can be explained in terms of nested Yule-Simon preferential growth processes. Later on, in further studies, we find that empirically generated temporal series using a chronologically ordered chess database exhibit long-range memory effects²¹. These memory effects cannot be explained with the mechanism based in a multiplicative process, nor the pure Yule-Simon process. In this sense, the mechanism of the evolution of the game tree popularity distribution needs new ingredients in order to reproduce the observed long-range correlations, establishing the problem we address in this work. We tackle this problem by making use of a result introduced by Cattuto et al.²⁹, who modifies the Yule-Simon process by incorporating a probabilis-

tic memory kernel. This model has the advantage that while memory is introduced, the long-tailed frequency distribution is preserved. For simplicity, we refer to this model as the Cattuto's model (CM). We generate artificial databases from the CM and detect long-range correlations under plausible values of the model parameters. In particular, the Hurst exponent³⁰ obtained from the CM depends on the length of the analyzed time series, in agreement with the behavior observed for the empirical chess database. In addition, the inter-event time distribution of the most popular openings is well reproduced by CM. The parameters of the memory kernel extension in CM are correlated with the expertise of the players showing that strategies for choosing a opening line is related to the level of expertise.

II. THEORETICAL BACKGROUND

A. Models

One of the first models able to explain the emergence of Zipf's law was introduced by Yule³¹, which was devised to explain the emergence of power laws in the sizes distribution of biological genera. Later on, Simon³² introduced a similar, but less general variation of the model³³, which fits more naturally in the context of Zipf's law. It is known as the Yule-Simon model (YSM), and different variations of the model re-emerged in the literature several times. Its most recent variant, known as preferential attachment, became one of the more important ideas at the beginning of complex networks theory³⁴. Cattuto et al.²⁹ introduced a variant of the YSM, which includes memory effects by incorporating a kernel, while at the same time preserving the long tail frequency distribution exhibited by the YSM. The YSM model applied to chess game generation is as follows. We begin with an initial state of n_0 games. At each temporal step t there are two options: i) to introduce a new game with probability p or ii) to copy an already existing game with probability $\bar{p} = 1 - p$. In the latter case we have to determine which of the previous games is being copied. The probability of choosing a game that has already occurred k times is $\bar{p}k\pi(k, t)$, where $\pi(k, t)$ is the fraction of games with frequency k at time t . This means that copying a certain game does not depend on how far back in time the game took place. For this reason, the process does not exhibit long-range memory effects. For Cattuto's model (CM), we follow the same procedure, except that now, the probability of copying a game that occurred at time $t - i$ is no longer constant, but has the functional form:

$$Q(i) = \frac{C(t)}{\tau + i}. \quad (1)$$

In Eq. (1), τ is a time scale in which recently added games have comparable probabilities and can be considered as the memory kernel extension. $C(t)$ is a logarithmic normalization factor. The probability distribution density of

the game frequencies that results from this process is²⁹:

$$P(k) = \frac{p}{(n_0 + pt)(K\alpha)k} \left[\frac{\ln(A/k)}{K} \right]^{\frac{1}{\alpha} - 1}, \quad (2)$$

where $\alpha = \bar{p}$, $K = \frac{1-\alpha}{\alpha\Omega}$, $A = e^{Kt^\alpha}$ and Ω is a fit parameter.

B. Correlations

In order to study long-range correlations of a chronologically ordered chess database we map the database to a time series assigning a real number to each game. It is known that the assignation rule used in the map has a direct effect in the degree of persistence observed in the series³⁵. In other words, long-range correlations are affected both by the intrinsic properties of the database and by the mapping code. Taking this into account we use several assignation rules to study the persistence of memory effects in the database. One of these rules, which is introduced in the analysis of literary corpora²³ and already employed in a chess database²¹, is the popularity assignation rule (PAR). In PAR, each element of the time series, $X(t)$, corresponds to the popularity at depth d of the t -th game in the database. The popularity is equal to the total number of games in the database that have the same sequence of moves up to depth d . In this work, we introduce three new assignation rules: the ELO assignation rule (EAR), the Gaussian assignation rule (GAR), and the uniform assignation rule (UAR). EAR makes use of the information of the players' ELO available in the chess database, where the ELO is a standard measure used in chess to estimate the skill of chess players. In EAR, we assign to each game the average ELO corresponding to all games with the same moves up to depth d , where, for each game, we took the ELO corresponding to the most skillful player of the two. Finally, GAR and UAR are random assignation rules. More specifically, a random number taken from certain probability distribution function, e.g. Gaussian (GAR) or uniform (UAR), is assigned to each different game. It is expected that all of these assignations do not introduce spurious correlations.

There exists a wide variety of techniques to detect long-range correlations in time series, but not all suitable to analyze all kinds of series, especially if they are non-stationary or exhibit underlying trends. Peng et. al.³⁶ introduced the Detrended Fluctuation Analysis (DFA), a useful technique for detecting long-range correlations in time series with non-stationarities. In the DFA method the cumulated series $Y(t)$ is segmented into intervals of size n . At each segment the cumulated series is fitted to a polynomial $y_n(t)$ and the fluctuation function is obtained

$$F(n) = \sqrt{\frac{1}{N} \sum_{i=1}^N [y(i) - y_n(i)]^2}, \quad (3)$$

where N is the total number of data points. A log-log plot of $F(n)$ is expected to be linear. If the slope is less

than unity, it corresponds to the Hurst exponent. When $H = 0.5$ the cumulated time series, $Y(t)$, resembles a memoryless random walker. On the other hand, for $H > 0.5$ ($H < 0.5$) a random walker with persistent (anti-persistent) long-range correlations or memory effects.

III. RESULTS

A. Inter-event Time Distribution: Fitting Model Parameters

We analyze the *ChessDB* dataset, which contains around 1.4 million chess games played between the years 1998 and 2007³⁷. Our first goal is to find the best choices for the models' parameters p for the YSM, and p and τ for the CM-, in order to reproduce the different statistical patterns observed in the chess database. The value of the parameter p can be directly measured from the database, using the formula

$$p = N_d(t_{\text{total}})/N_{\text{total}}, \quad (4)$$

where N_{total} is the total number of games in the database, and $N_d(t_{\text{total}})$ is the number of nodes in the game tree at depth d , after t_{total} games –i.e. all games– have been considered. Following the ideas used in the analysis of earthquakes^{38,39}, in Fig. 1 we show the frequency distribution of inter-event times corresponding to the fraction of the most popular openings in the database. More specifically, we compute the frequency k of all openings, then we filter all openings with frequency lower than $(k_{\text{max}} - k_{\text{min}})/3$. The resulting subset corresponds to the 40% of the chess games in the database, and the 30% of the games generated with the CM. For the database, we considered the set of openings corresponding to game depth $d = 4$. Both, CM and YSM yield to a good agreement with the chess database. However, in order to find a good fit in the YSM, the value of p has to be over-estimated as compared to the mean value measured in the real database. For the sake of comparison, in Fig. 1 we show the inter-event time distribution for the YSM, when the value of p is set to the one measured in the database. This curve clearly deviates from the measured to the database. While the measured value of p in the chess database is $p = 0.005$, the value that yields the best fit of the inter-event time distribution with YSM is $p = 0.1$. On the other hand, using CM we obtain a good fit with the right value of p when the other parameter, defining the characteristic time scale of the memory kernel (see Eq. (1)) is $\tau = 34$. A correct model for the inter-event has to take into account a memory of the previous played games as it is evidenced in the time series associated to the real databases. The fact that CM model works better suggest that the memory kernel could be the necessary mechanism to describe the long-range correlations observed in the chess database²¹.

In order to test the existence of memory directly from the inter-event time distribution of the database, we fitted a generalized gamma distribution

$$f(x) = \frac{C}{x^{1-\gamma}} e^{-\frac{\delta}{a}x}, \quad (5)$$

typically used to describe the inter-event times observed between earthquakes, which are known to be strongly correlated events^{38,39}. We find that a good fit is obtained restricting the value of γ to $\gamma = 1$. The value of δ is also close to 1; in this case $\delta = 0.969$ (see Inset in Fig. 1). This means that the inter-event time of the database is similar to an homogeneous Poisson process ($\delta \simeq 1$). Besides the fact that the exponential is slightly stretched, memory effects cannot be derived directly from the fittings parameters. Is quite direct to see that a Poisson process is a good approximation for the inter-event time of the YSM, and also in the CM. Therefore this is the reason of why both YSM and CM can be well fitted to the inter-event times of the database. In the case of the YSM, the probability P_t that an already existing game is repeated between time t and $t + \Delta t$ is $P_t = ((1 - p)N_g/N)\Delta t$, where N_g is number of games of the game to be repeated, and N is the total number of games; both quantities specified at time t . After a transient time N_g/N is expected to be nearly constant⁴⁰ in the YSM process, and then $P_t = \Delta t/\tau_p$, where $\tau_p = N/(N_g(1 - p))$, also becomes constant. Since the fitting of Eq.(5) in the database turns out to be $\gamma = 1$ and $\delta = 0.969$, then $a \sim \tau_p$. The fraction of the most popular games used for building the inter-event time of the database is ~ 0.4 and so $\tau_p \sim 2.5$ ($p \ll 1$). Considering that the values of τ_p from the fitting is $\tau_p = 2.689$, the inter-event time in the database can be considered as a nearly homogeneous Poisson process. In CM, the memory kernel imposes a restriction to the games that can be replicated, behaving like a window that looks at the most recent games, however if in this window the fraction of the games under consideration remains constant during the entire process, then it is similar to the YSM and therefore this is also a Poisson process. The difference is that now τ_p depends on both p and τ . The dependence on τ of the relaxation time τ_p is in particular important because it allows to fit the CM to the database with a realistic value of p .

Finally, we warn the reader that in the chess database the number of different games grows in time according to the Heaps' law²⁰, and not linearly as in Eq. (4). More specifically, it holds $N_d \sim t^{\lambda_d}$ for some $0 < \lambda_d < 1$ which depends on the game depth, d , considered. Therefore, Eq. (4) is just an approximation that we choose to work with in order to keep things simple since in the CM model p is kept constant. Moreover, the approach to the database as a Poisson process is also affected by the time dependency of p since τ_p should also depend on t and the Poisson process is at least slightly inhomogeneous.

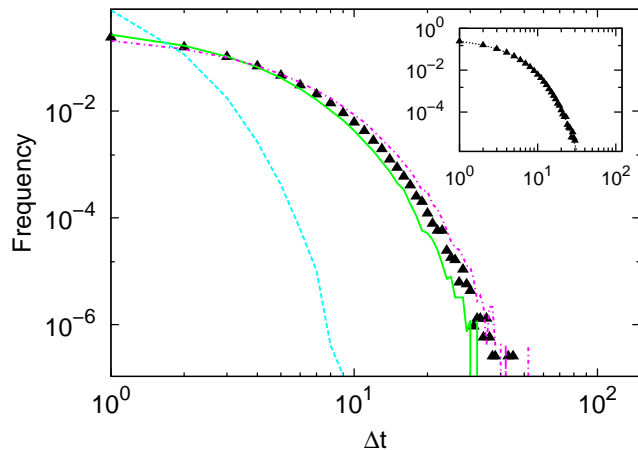


FIG. 1. (Color online) Frequency distribution of inter-event times measured in the chess database (full triangles), generated with CM model for $p = 0.005$ and $\tau = 34$ (green full line), and generated with YSM model, for parameters $p = 0.1$ (magenta point-dashed line) and $p = 0.005$ (cyan dashed line). Inset: Frequency distribution of inter-event times measured in the chess database (black triangles) and a fit using Eq. (5) for $C = 0.4265$, $\delta = 0.969$, $a = 2.689$ and $\gamma = 1$.

B. Popularity Distribution: Testing Predictions

After fitting the model parameters, we test the model predictions. Fig. 2 shows a plot of the distribution of popularities obtained with CM and YSM using the same parameter values obtained from fitting the inter-event time distribution in Fig. 1. First, we see how the distribution of popularities for YSM model not only yields a power law but the slope in a log-log plot is very close to 2, which is exactly what is expected in this process for small values of p ³². Secondly, the distribution obtained from the CM model shows a gentle curvature, and is very well fitted by the theoretical expression of Eq. (2). Finally, the distributions for the database and CM model are much closer in slope than to that of YSM model. Notice however, the right slope can be obtained for the YSM model as well, if an appropriately small value of p is used (not shown), but at the cost of ruining the fitting in Fig. 1. In other words, CM is able to fit Fig. 1 and 2, simultaneously, for the value of p given by Eq. (4). On the other hand, this is not the case for the YSM.

In summary, the extra parameter τ of CM that measures the kernel extension provides an extra degree of freedom which is needed to fit the inter-event time curve with a plausible value of p . The value of τ is later validated by the correct prediction of the popularity distribution.

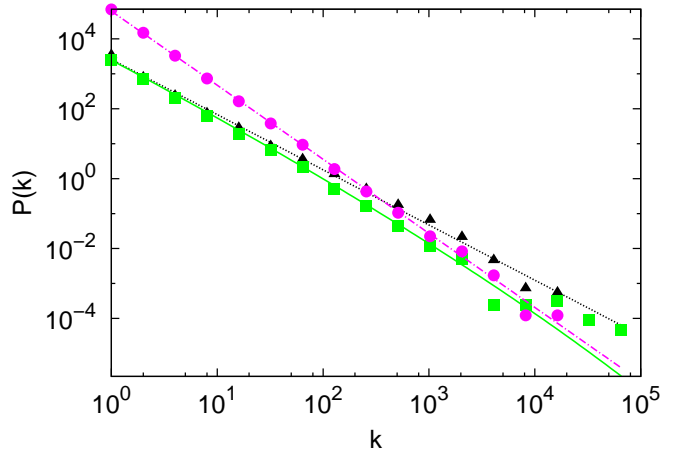


FIG. 2. (Color online) Log-log plot of the distribution of popularities measured in the database (black triangles) and linear fit $S_d(k)$ with slope $\alpha_d = -1.53$ (dotted black line); generated with CM model, $p = 0.005$ and $\tau = 34$ (green squares) and fit $P_{CM}(k)$ corresponding to Eq. (2) with parameter $\Omega = 1.65$ (full green line); and generated with YSM model, $p = 0.1$ (magenta circles) and linear fit $P_{YSM}(k)$ with slope $\alpha_{YSM} = -2.12$ (magenta point dashed line).

C. Hurst Exponent: Long-range Correlations Analysis

The CM model is not only able to simultaneously fit the inter-event time distribution and the popularity distribution. It also yields to the existence of long-range correlations, or memory effects, which we compare against the long-range correlations exhibited by the chess database. In order to analyze the existence of long-range correlations, we measure the Hurst exponent of the time series obtained from the models and data, computed using the different assignation rules: PAR, EAR, GAR and UAR (see section II). To calculate the Hurst exponent we use a linear DFA method.

Fig. 3 shows the Hurst exponent as function of the length of the analyzed time series, for the PAR. Consistently with a lack of long-range time correlations, or memory effects, the Hurst exponent is close to 0.5 in the artificial database generated with the YSM, for both $p = 0.1$ and $p = 0.005$ (this last one is not shown). In particular, the result has not significant size effects. In contrast, the time series generated with the CM exhibit both, long-range memory and size effects, although H does not grow as rapidly as it does for the empirical chess database. The value of the Hurst exponent for the whole chess database is $H = 0.69$. Using CM model we were able to obtain this value for a generated database of the same size, using the same parameters obtained from the fit of the inter-event time, i.e. $p = 0.005$ and $\tau = 34$. However, the tendency is different between the database and the CM. For the database, the Hurst exponent be-

comes large, even for short sequences of games. In contrast, the Hurst exponent grows steadily in the CM. We ignore the source of such difference, but it constitutes a problem that can be studied in future works.

It is known that large fluctuations in the values of the time series, $X(t)$, might lead to spurious long-range memory effects, i.e., values of H significantly different from 0.5⁴¹. Since the popularity distribution is long tailed—in both models and in the empirical database—the PAR assignation rule leads to large fluctuations in the values of $X(t)$. In order to test the influence of these fluctuations we repeated the calculations of H using time shuffled series $X_{\text{shuff}}(t)$ (Figure not shown). The Hurst exponents obtained after the shuffling are close to 0.5, and thus, the large fluctuations in $X(t)$ are not the reasons for which $H > 0.5$. Conversely, we can check if the condition $H > 0.5$ persist when there are no large fluctuations in the values, $X(t)$, of the time series. For that purpose, we used the others assignation rules, EAR, UAR and GAR, as the corresponding time series values have a well defined variance. In Fig. 4 we show the Hurst exponent as a function of the size of the analyzed time series for three different assignation rules; EAR at the top, GAR in the middle and UAR at the bottom panel. The results shown in the figure, confirm that long-range correlations—i.e. a Hurst exponent $H > 0.5$ —, is also obtained for this other assignation rules. We remark that these assignation rules are not expected to introduce spurious correlations. In particular, the GAR and UAR are completely random. However, as the reader can see, different assignation rules lead to different values of the Hurst exponent; both, in the model and in the database. The higher values of the H are observed in the database with the EAR rule, and at the same time, the larger discrepancy with the CM. This is expected, as the ELO's are artificially introduced into the CM, while for the chess database is an intrinsic part of the data. In fact, this is consistent with the observation that the best agreement between the CM model and the database is obtained for the UAR and GAR assignation rules. We leave open the problem of understanding the correlations existing between the ELO's and the sequence of games in the empirical chess database.

D. The Expertise of the Chess Players

In order to analyze the expertise of chess players let us first look at the inter-event time distribution. To implement this analysis we divided the chess database in three disjoint ELO intervals which contain nearly the same number of games, and fitted the corresponding inter-event time distributions using the CM. In Fig. 5 we show these results, which indicate how the expertise of the players is correlated with the extent of the memory kernel parameterized by τ . For each ELO interval, the probability of adding a new game was measured in the empirical database, resulting in: ELO range [1 – 2199], $p = 0.01$; ELO range [2200 – 2399], $p = 0.009$; and ELO

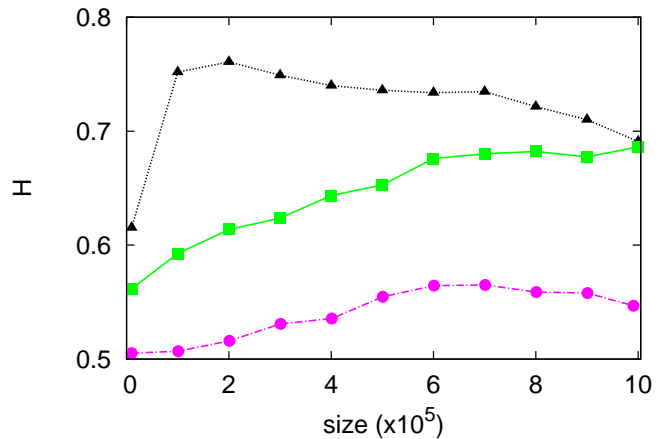


FIG. 3. (Color online) Hurst exponent obtained by the DFA method as a function of the length of the time series using the PAR assignation rule in the database (dotted black line with triangles), generated with CM model, $p = 0.005$ and $\tau = 34$ (full green line with squares), and generated with YSM, $p = 0.1$ (point dashed magenta line with circles).

range [2400 – 2851], $p = 0.007$. In consequence, we fitted the distribution for the three ELO intervals by fixing the corresponding value of p and varying the parameter τ . We were able to fit the inter-event times curves for the two higher ELO intervals using the corresponding values of p , however we were not able to achieve this for the lower ELO interval. A good fit was found for a memory extension with parameter $\tau = 10$, but at the cost of increasing the value of the probability of introducing a new opening to $p = 0.03$. The values of τ for which the best fits are achieved, increase with the ELO. Using Eq.(5) we fitted the inter-event times curves corresponding to the three ELO intervals (fittings no shown). While the exponent is approximately $\gamma \sim 1$ for all the cases, for the chess database, the parameter δ varies slightly when analyzing inexperienced and experienced players. However, we have to keep in mind that the function of Eq.(5) is extremely sensitive to variations of δ (Table I). As expected, the exponent δ is very close to unity for the YSM, and so for this model we fitted the inter-event time using a simplified version of the gamma function: $f(x) = Ce^{-x/a}$. As mentioned before, in the CM the probability of copying a game that occurred at time $t - i$ is given by the probabilistic kernel. Therefore, a more detailed analysis can be performed. More specifically, in Fig. 6 we show the histogram of all times i resulting from the process of generating the artificial database with CM. In terms of the memory kernel, Fig. 5 and 6 can be interpreted as follows. We see that for more experienced players we find a more extended, and slow decaying, memory kernel than for players with lower ELO ratings. Is more probable for inexperienced players to repeat more recent openings than players with higher ELO ratings. In contrast, at

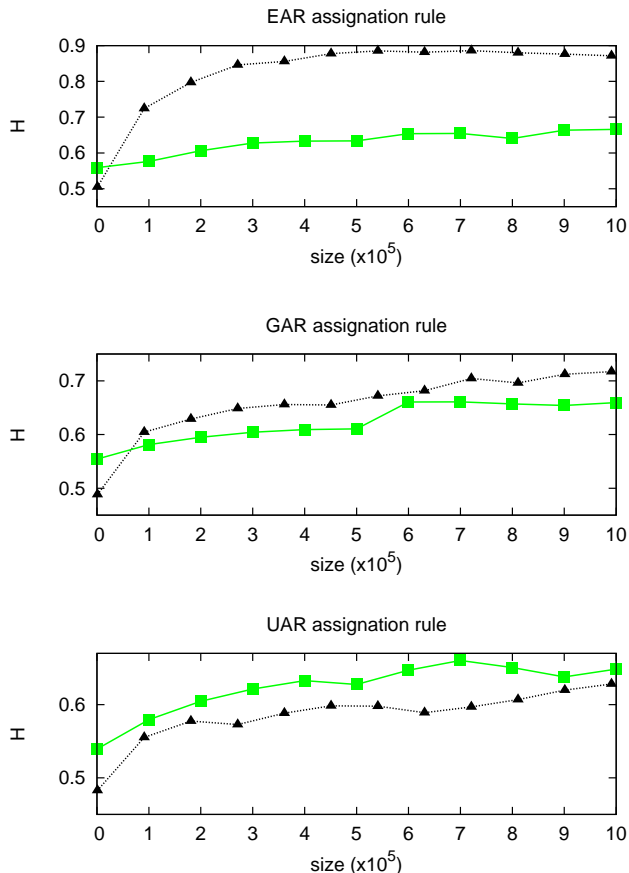


FIG. 4. (Color online) Hurst exponent obtained by the DFA method as a function of the length of the time series using the EAR (top panel), GAR (middle panel) and UAR (bottom panel) assignment rules in the database (dotted black line with triangles) and generated with CM model, $p = 0.005$ and $\tau = 34$ (full green line with squares).

larger times, the function $Q(i)$, that is, the probability of repeating an opening that occurred i time steps back, decays more rapidly for lower ELOs. This is consistent with the difference in behavior of the Hurst exponent observed in the three ELO intervals²¹.

We also analyze the expertise of the players measuring the Hurst exponent. For this we employ the same ELO intervals used in the analysis of the inter-event times. While the UAR assignment rule yields the best agreement between the database and CM model, here, the Hurst exponent has the weakest dependence with the ELO intervals when varying the size of the database. Specially for the CM. Therefore, we restrict our analysis to the GAR, which is shown in Fig. 7. Here, to generate the time series corresponding to each ELO in the CM, we use the parameter values shown in table I. In this way, we introduce into the CM, at least to some extent, a correlation between the game statistics and the ELOs. We find that for long times, the Hurst exponent increases as

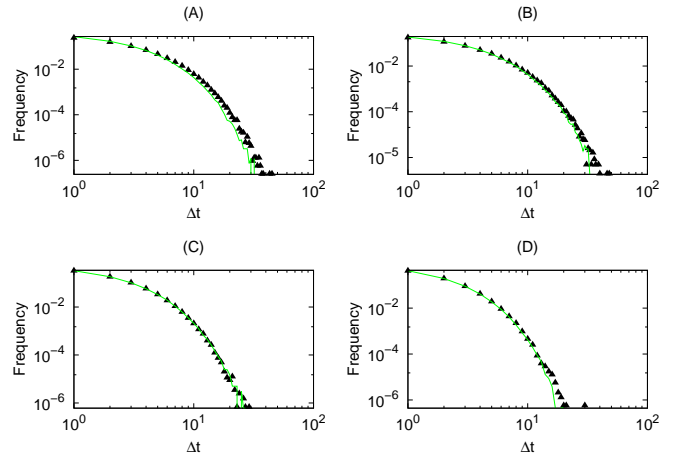


FIG. 5. (Color online) Frequency distribution of inter-event times for different ELO ranges: A) full database, B) ELO range [1 – 2199] C) ELO range [2200 – 2399] and D) ELO range [2400 – 2851]. The full black triangles correspond to the inter-event time distribution measured from the database and the full green line to that generated with CM. The parameters used in CM are: A) $\tau = 34$, $p = 0.005$; B) $\tau = 10$, $p = 0.003$; C) $\tau = 20$, $p = 0.009$; and D) $\tau = 70$ and $p = 0.007$.

TABLE I. Summary of the parameters for the best fits in model, corresponding to the inter-event time distributions in Fig. 5, for different ELO ranges and the full ELO range in the empirical chess database. τ_p is the estimated value of the corresponding Poisson process.

ELO range	p	τ	δ	a	τ_p
A) full	0.005	34	0.969	2.689	2.56
B) 1–2199	0.03	10	0.941	2.85	3.94
C) 2200–2399	0.009	20	0.977	1.978	2.65
D) 2500–2851	0.007	70	0.986	1.489	2.11
E) Yule	0.1		1	3.77	4.28

the ELO increases, specially for the chess database.

IV. DISCUSSION

From the fit of the inter-event time curve of the most popular openings in the chess database, using a generalized gamma function, we found that this curve is in fact well described with a slightly stretched exponential. values obtained from the fit of the decaying time are consistent with the value derived through direct calculation. On the other hand, the inter-event time of both CM and YSM are also well described by a Poisson process, and then a good fit of the database can be obtained using both models. However, the value of the parameter p that yields to a best fit in the YSM model is far from the measured value in the database, they are $p = 0.1$ and $p = 0.005$,

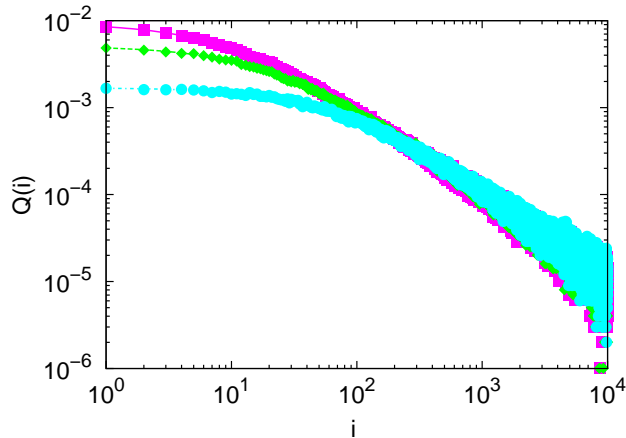


FIG. 6. (Color online) Histogram of times i (Memory kernel function $Q(i)$) up to $i = 10^3$ for ELO range [1 – 2199] (magenta squares), ELO range [2200 – 2399] (green diamonds) and ELO range [2400 – 2851] (cyan circles) and fits using $Q(i) = \tilde{C}/(i + \tilde{\tau})$ with resulting fit parameters $\tilde{C} = 0.0975$ and $\tilde{\tau} = 10.48$ for ELO range [1 – 2199] (full magenta line), $\tilde{C} = 0.1039$ and $\tilde{\tau} = 20.46$ for ELO range [2200 – 2399] (dashed green line) and $\tilde{C} = 0.1189$ and $\tilde{\tau} = 70.28$ for ELO range [2400 – 2851] (point-dashed cyan line) .

respectively. On the contrary, CM model permits a good fit with the value of p measured in the database by a convenient choice of the parameter τ which quantifies the memory kernel extension. This result indicates that the memory kernel of CM model can be used to quantify memory effects in the games of the chess database.

The popularity distribution is also well described by CM model using the parameter p and τ obtained from the fitting of the inter-event time. In this sense CM is also a better approximation than YSM model. This provides additional information on the importance of the memory kernel to model the behavior of the real chess database. Moreover, time series generated using the CM model show long-range correlation and size effects in a similar fashion as the database of real game does. All these results show that many of the main ingredients necessary to model the memory effects in the real chess database are captured by the CM model.

In the analysis of long-range correlations, we found that size effects and the degree of persistence are affected by the particular choice of the assignation rule. In a previous work, we realized that the popularity assignation rule (PAR) is adequate to discriminate the behavior of players with different levels of expertise²¹. However, when we use the PAR to compare the behavior of the Hurst exponent as a function of the length of the time series, the value of H for the full size of the database is nearly the same for the chess database and CM, but the Hurst exponent does not grow as rapidly in the case of the CM. On the other hand, the random maps of the Gaussian and Uniform Assignation Rules (GAR and UAR)

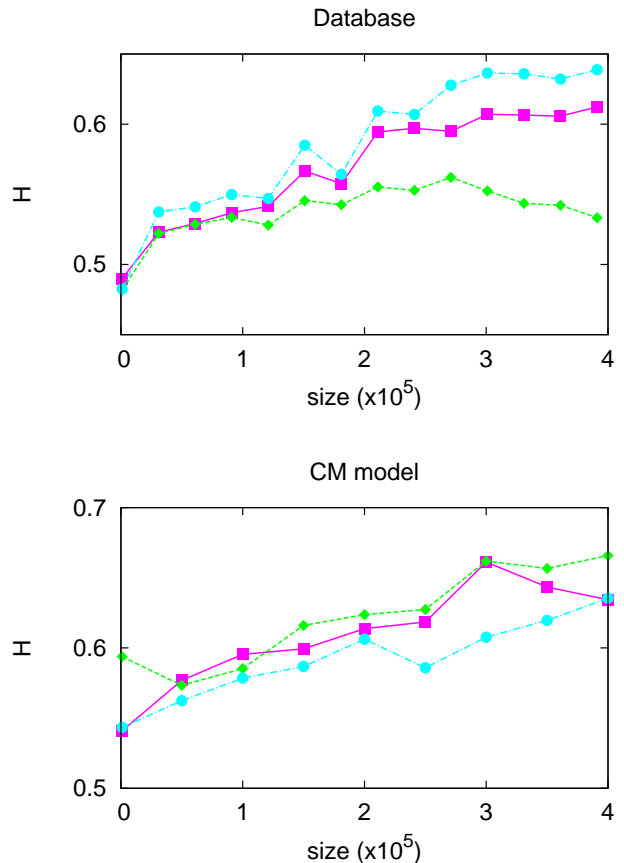


FIG. 7. (Color online) Hurst exponent obtained by the DFA method as a function of the length of the time series corresponding to the three different ELO intervals, ELO range [1 – 2199] (full red line with squares), ELO range [2200 – 2399] (dashed green line with diamonds) and ELO range [2400 – 2851] (point dashed blue line with circles), generated with the GAR assignation rule for the database (top panel) and CM model (bottom panel) using the corresponding parameters that give the best fit for the inter-event time curves.

seem to be adequate, since they give similar results for CM model and the database, though in the case of the GAR we obtain more similar results to those for the PAR. Despite this, the CM does not display distinguishable difference in the behavior of the Hurst exponent for the different levels of expertise. We remark however, that the observation of the different strategies employed by players with different levels of expertise depends on the chosen assignation rule.

These observations reinforce the use of the inter-time event frequency distribution to detect the level of expertise of the chess players in the database since this analysis does not use an assignation rule. In this sense, we find through the fittings of the inter-time frequency distribution using CM model, that the characteristic time τ increases as the expertise of the players increases. Since

τ fixes the memory kernel extension, this indicates that high level players use on average more opening lines than less expert players. This has a direct interpretation since it is expected for more expert players to have a more

extensive knowledge of opening lines. These supports previous results where the expertise of the players was detected using long-range correlations analysis.

-
- * alschaigorodsky@gmail.com
† juanpool@gmail.com
‡ billoni@famaf.unc.edu.ar
- ¹ E. Majorana, *Scientia* pp. 58–66 (1942).
 - ² G. K. Zipf (1949).
 - ³ J. G. March and H. A. Simon (1958).
 - ⁴ M. S. Granovetter, *American journal of sociology* pp. 1360–1380 (1973).
 - ⁵ D. J. Watts, *Nature* **445**, 489 (2007).
 - ⁶ D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, et al., *Science (New York, NY)* **323**, 721 (2009).
 - ⁷ C. Castellano, S. Fortunato, and V. Loreto, *Reviews of modern physics* **81**, 591 (2009).
 - ⁸ F. Prost, in *Turing-100*, edited by A. Voronkov (Easy-Chair, 2012), vol. 10 of *EPiC Series*, pp. 268–277, ISSN 2040-557X.
 - ⁹ B. Blasius and R. Tönjes, *Phys. Rev. Lett.* **103**, 218701 (2009).
 - ¹⁰ H. V. Ribeiro, R. S. Mendes, E. K. Lenzi, M. del Castillo-Mussot, and L. A. N. Amaral, *PLoS ONE* **8**, e54165 (2013).
 - ¹¹ M. Sigman, P. Etchemendy, D. F. Slezak, and G. A. Cecchi, *Frontiers in Neuroscience* **4**, 1 (2010), ISSN 1662-4548.
 - ¹² P. Chassy and F. Gobet, *PLoS ONE* **6** (2011).
 - ¹³ C. Sire and S. Redner, *The European Physical Journal B* **67**, 473 (2009), ISSN 1434-6028.
 - ¹⁴ Y. de Saá Guerra, J. M. González, S. S. Montesdeoca, D. R. Ruiz, A. García-Rodríguez, and J. García-Manso, *Physica A: Statistical Mechanics and its Applications* **391**, 2997 (2012), ISSN 0378-4371.
 - ¹⁵ A. M. Petersen, W.-S. Jung, and H. E. Stanley, *EPL (Europhysics Letters)* **83**, 50010 (2008).
 - ¹⁶ E. Bittner, A. Nußbaumer, W. Janke, and M. Weigel, *The European Physical Journal B* **67**, 459 (2009), ISSN 1434-6028.
 - ¹⁷ E. Ben-Naim, S. Redner, and F. Vazquez, *EPL (Europhysics Letters)* **77**, 30005 (2007).
 - ¹⁸ A. Heuer and O. Rubner, *The European Physical Journal B* **67**, 445 (2009), ISSN 1434-6028.
 - ¹⁹ H. V. Ribeiro, S. Mukherjee, and Xiao Han T. Zeng, *Phys. Rev. E* **86**, 022102 (2012).
 - ²⁰ J. I. Perotti, H. Jo, A. L. Schaigorodsky, and O. V. Billoni, *EPL (Europhysics Letters)* **104**, 48005 (2013).
 - ²¹ A. L. Schaigorodsky, J. I. Perotti, and O. V. Billoni, *Physica A: Statistical Mechanics and its Applications* **394**, 304 (2014), ISSN 0378-4371.
 - ²² M. E. Glickman, *American Chess Journal* **3**, 59 (1995).
 - ²³ M. A. Montemurro and P. A. Pury, *Fractals* **10**, 451 (2002).
 - ²⁴ E. G. Altmann and G. Cristadoro, *Proceedings of the National Academy of Sciences* **109**, 11582 (2012).
 - ²⁵ M. Cristelli, M. Batty, and L. Pietronero, *Sci. Rep.* **6**, 1 (2012).
 - ²⁶ Note1, in the bibliography game theory, for a two-players sequential game as it is the case of chess, what we here call a *move*, it is usually called a *ply*.
 - ²⁷ S. Maslov, *Physics* **2**, 97 (2009).
 - ²⁸ *Nature* **462**, 546 (2009), ISSN 0028-0836.
 - ²⁹ C. Cattuto, V. Loreto, and V. D. P. Servedio, *EPL (Europhysics Letters)* **76**, 208 (2006).
 - ³⁰ J. Beran, *Statistics for Long-Memory Processes* (Chapman-Hall, 1994), 1st ed.
 - ³¹ G. U. Yule, *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* **213**, 21 (1925).
 - ³² H. A. SIMON, *Biometrika* **42**, 425 (1955).
 - ³³ M. V. Simkin and V. P. Roychowdhury, *Physics Reports* **502**, 1 (2011).
 - ³⁴ A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999), ISSN 1095-9203.
 - ³⁵ M. Amit, Y. Shmerler, E. Eisenberb, M. Abraham, and N. Shnerb, *Fractals* **02**, 7 (1994).
 - ³⁶ C.-K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, *Nature* **356**, 168 (1992).
 - ³⁷ Note2, original data at <http://chessdb.sourceforge.net>. Filtered data at <http://juanperotti.blogspot.it/p/datasets.html>.
 - ³⁸ P. Bak, K. Christensen, L. Danon, and T. Scanlon, *Phys. Rev. Lett.* **88**, 178501 (2002).
 - ³⁹ A. Corral, *Phys. Rev. Lett.* **92**, 108501 (2004).
 - ⁴⁰ Note3, to be precise, a slowly varying quantity.
 - ⁴¹ J. Barunik and L. Kristoufek, *Physica A: Statistical Mechanics and its Applications* **389**, 3844 (2010).