Viewpoint

Power laws in chess

Sergei Maslov

Department of Condensed Matter Physics and Materials Science, Brookhaven National Laboratory, Upton, NY 11973, USA

The popularity of various chess openings follows a power law distribution, but the exponent depends on the depth of the opening sequence.

Subject Areas: **Statistical Mechanics, Interdisciplinary Physics**

A Viewpoint on:

Zipf's Law in the Popularity Distribution of Chess Openings
Bernd Blasius and Ralf Tönjes
Phys. Rev. Lett. **103**, 218701 (2009) – Published November 16, 2009

A confession: I don't play chess. I just didn't inherit the gene—one that seems to be quite common among Russians—for this game. But even I have heard from my chess-playing friends about famous opening sequences, from Alekhine's defense, to the Zilbermints gambit, and of course about e2-e4 (the most popular first move in which the King's pawn moves forward two spaces). So I was not the slightest bit surprised to learn that the distribution of popularity of opening moves in chess has a long power-law tail, as reported in *Physical Review Letters* by Bernd Blasius and Ralf Tönjes at the University of Potsdam, Germany [1]. However, some surprises lurk in the results.

Finding power laws has now become *de rigueur* when analyzing popularity distributions. Long tails have been reported for the frequency of word usage in many languages [2], the number of citations of scientific papers [3], the number of visits (hits) to individual websites in a given time interval [4], and many more. In all these cases (including this new one related to chess) the exponent of the distribution is close to -2. That is, the number of entries (chess openings, words, papers, or websites) with popularity P approximately scales as P^{-2}. The semi-universal value of this exponent was first noticed by Zipf [2] when he saw the same statistics apply to such different objects as words ranked by their popularity and cities ranked by their population.

Many scientists proposed simple (and not so simple) models aimed at explaining the origins of this scaling. For city-size distribution, the celebrity list of modelers includes Paul Krugman [5]—the Nobel Prize winning economist and New York Times columnist. Even though the laws of population dynamics responsible for city formation and subsequent growth appear to have very little in common with rules dictating preferences in chess openings, all inverse quadratic power-law distributions became collectively known as "Zipf's law." There is indeed something special about the distribution

$P^{-\alpha}$, with α=2, since it separates the region with a well-defined average ($\alpha > 2$) from that where the average formally diverges and thus depends on the upper cutoff ($\alpha \leq 2$). Nevertheless, the quest for the universal "first principles" explanation of Zipf's law remains elusive.

Apart from establishing yet another example of Zipf's law, the work of Blasius and Tönjes goes a long way towards elucidating its origins in the special case of sequential games or, more generally, any composite multistep decision processes (e.g., complex business or political strategies). These processes are best visualized as decision trees with multiple choices at each level (Fig. 1). The number of possible paths on such trees grows exponentially with the number of steps. As a result, even in the simplest cases the exhaustive enumeration very soon becomes impossible.

The first important observation made by the authors is that if one concentrates on the distribution of popularity of opening sequences limited to the first d steps of the game, it will also be described by a power law, yet with a nonuniversal exponent α_d that linearly grows with the number of moves. The universal Zipf distribution with α=2 is recovered only after these d-dependent distributions are all merged together. The rationale for such a merger leading to double counting is poorly explained in the paper. Nonuniversal power-law exponents often implicate multiplicative random walks [6–8] and this case is no exception. Other examples of power laws generated by multiplicative random walks include wealth of individuals [9], stock prices and their drawdowns (deviations down from the maximum) [10], gene family sizes expanding by gene duplication [11], and many others.

One way to calculate the popularity of a particular opening sequence σ is to notice that every sequential move i of the sequence reduces the number of games in the database that open with these moves by a factor $0 < r_i \leq 1$. These factors are the same as branching ra-
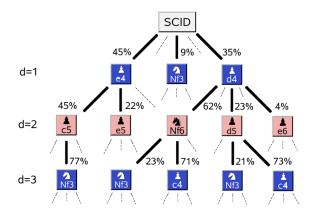
FIG. 1: Chess openings can be described as decision trees, showing each move and associated branching ratios. This diagram shows the three most popular first ($d = 1$) half-moves in the 1.5-million-game ScidBase chess database [12] and their branching ratios. For example, in 45% of the games, white starts with e4 (King's pawn to fourth row, in algebraic chess notation), 35% start with d4 (Queen's pawn to fourth row), etc. Each of these moves then have branching ratios to the second half-move by black ($d = 2$). Blasius and Tönjes find that for all games up to $d = 40$, the opening sequence popularity follows a Zipf power law with universal exponent nearly equal to -2, but for small values of d, the exponent is nonuniversal and depends linearly on d. (Adapted from Ref. [1].)

tios illustrated in Fig. 1. If the total number of recorded games is N (which is ~ 1.5 million in the professional chess database ScidBase [12] used in this work) then the number of openings starting with a particular sequence of d moves is given by $N(\sigma) = N r_1 r_2 \ldots r_d$. The value $N(\sigma) = 1$ serves as an absorbing lower boundary for this multiplicative process. When a random walker hits such an absorbing boundary it stops moving altogether. In the case of chess opening, once the diversity is down to just one realization, a unique move will be selected at each subsequent time step and $N(\sigma)$ will remain at 1 until the end of the game. In the case of standard (additive) random walks, a boundary generally gives rise to an exponential Boltzmann distribution. For multiplicative random walks after the logarithmic change of variables [6, 7], this exponential distribution becomes a power law.

The exponent α_d of the high-popularity tail of the distribution can be analytically derived for some special cases of the distribution of ratios r_i : $\rho(r) \sim r^\beta$ (see Eq. (6) in the paper of Blasius and Tönjes [1]). According to these calculations, α_d linearly increases with the number of moves d. This is in agreement with the actual distribution of popularity of chess openings. However, the empirically measured $\rho(r)$, shown in Fig. 3(a) of their paper, has a different profile. Surprisingly, it closely follows the parameter-free distribution $\rho(r) = 2/\pi\sqrt{1 - r^2}$. This distribution describes the density of points on a circle projected onto its diameter.

Blasius and Tönjes offer no explanation for this empir-

ical observation. Qualitatively, this profile of $\rho(r)$ makes intuitive sense. At every position of pieces on the chess board, out of many moves allowed by the rules, just one would lead to the most favorable position for the long-term outcome of the game. Such moves that maximize the gradient of "fitness" would be preferentially selected by skillful chess players (the average rating of players in the database puts them between the Candidate Master and the International Master levels). This selection would manifest itself in $\rho(r)$ increasing (and possibly even diverging) as $r \rightarrow 1$. This divergence is a direct manifestation of players' skills. Indeed, if I were to play the game of chess against other equally clueless players, the shape of $\rho(r)$ defined by our uninformed random moves would likely to be flatter than that shown in Fig. 3(a) of Ref. [1].

As a suggestion for future studies, it would be interesting to empirically verify this hunch using player ratings included in the ScidBase or other less selective chess databases. In other words, is the distribution of openings selected by the best players significantly different from that selected by the worst players? Another observation made by the authors is that the shape of $\rho(r)$ is independent of the depth of the game d. This indicates that, at least at early stages of the game, the phase space of favorable moves does not significantly depend on d.

All these empirical facts summarizing the collective knowledge of many chess players have implications for the design of chess-playing computer algorithms. Thinking about chess has a long and venerable history in computer science. One of the founding fathers of information theory, Claude Shannon, has worked on this topic. In his 1950 paper [13] he outlined two possible approaches to designing a computer program playing chess: the brute force strategy, performing the exhaustive evaluation of all possible moves and opponent's responses for as many steps as computer power would allow. The other strategy is to iteratively select a few of the most promising moves at every step and concentrate computer resources on following a smaller number of more likely paths on the decision tree of the game. The shape of $\rho(r)$ in Ref. [1] provides an empirical justification for this latter strategy that is indeed the one used by modern chess-playing programs.

During the last decade it became customary to blame all types of power laws in popularity on linear preferential attachment mechanisms first used to explain the Zipf's law by another Nobel Laureate, Herbert Simon [14]. According to these models, the high popularity of certain items is a frozen accident self-sustained by fashion. For example, an initially popular website would acquire new links at a higher rate than its less popular cousins, and as a consequence, further increase its visibility. While in certain situations fashion-driven preferential attachment is likely to be responsible for long tails of popularity distribution, it is reassuring to know that it is not the case in chess—a quintessential game of skill.

References

[1] B. Blasius and R. Tönjes, *Phys. Rev. Lett.* **103**, 218701 (2009).
[2] G. K. Zipf, *Human Behaviour and the Principle of Least Effort* (Addison-Wesley, Cambridge, MA, 1949).
[3] D. J. de S. Price, *Science* **149**, 510 (1965).
[4] M. Crovella, M. Taqqu, and A. Bestavros, in *A practical guide to heavy tails: statistical techniques and applications*, edited by R. J. Adler, R. E. Feldman, and M. S. Taqqu (Birkhäuser, Boston, 1998).
[5] P. Krugman, *J. Japanese Int. Economies* **10**, 399 (1996).
[6] M. Levy and S. Solomon, *Int. J. Mod. Phys. C* **7**, 595 (1996).
[7] D. Sornette and R. Cont, *J. Phys. I* (France) **7**, 431 (1997).
[8] M. Marsili, S. Maslov, and Y-C. Zhang, *Physica* **253A**, 403 (1998).
[9] V. Pareto, *Cours d'economie politique* (F. Rouge, Lausanne, 1896).
[10] S. Maslov and Y.-C. Zhang, *Physica*, **262A**, 232 (1999).
[11] M. Huynen and E. Van Nimwegen, *Mol. Biol. Evol.* **15**, 583 (1998).
[12] http://scid.sourceforge.net.
[13] C. E. Shannon, *Philos. Mag.* **41**, 256 (1950).
[14] H. A. Simon, *Biometrika* **42**, 425 (1955).

About the Author

Sergei Maslov

Sergei Maslov received his M.S. from the Landau Institute in Moscow, Russia, in 1992 and his Ph.D. from Stony Brook University in 1996. In 2004 he was given a Presidential Early Career Award for Scientists and Engineers. He is currently a permanent staff member at Brookhaven National Laboratory where he is working on interdisciplinary applications of statistical physics. His most recent research concentrates on topics in systems biology and complex networks.