Physica A 407 (2014) 252–275

Contents lists available at ScienceDirect

Physica A

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

On the frequencies of patterns of rises and falls

J.M. Luck*

Institut de Physique Théorique, URA 2306 of CNRS, CEA Saclay, 91191 Gif-sur-Yvette cedex, France

HIGHLIGHTS

- We study a null model for frequencies of patterns of rises and falls in data series.
- The probability of a long pattern typically decays exponentially with its length.
- The associated decay rate is interpreted as the embedding entropy of the pattern.
- The entropy is evaluated exactly for all families of periodic patterns.
- The entropy of random patterns obey non-trivial multifractal statistics.

ARTICLE INFO

Article history:
Received 14 January 2014
Received in revised form 27 March 2014
Available online 13 April 2014

Keywords:
Data series
Patterns
Rises and falls
Entropy
Multifractals
Permutations

ABSTRACT

We investigate the probability of observing a given pattern of n rises and falls in a random stationary data series. The data are modeled as a sequence of $n+1$ independent and identically distributed random numbers. This probabilistic approach has a combinatorial equivalent, where the data are modeled by a random permutation on $n+1$ objects. The probability of observing a long pattern of rises and falls decays exponentially with its length n in general. The associated decay rate α is interpreted as the embedding entropy of the pattern. This rate is evaluated exactly for all periodic patterns. In the most general case, it is expressed in terms of a determinant of generalized hyperbolic or trigonometric functions. Alternating patterns have the smallest rate $\alpha_{\min} = \ln(\pi/2) = 0.451582\ldots$, while other examples lead to arbitrarily large rates. The probabilities of observing uniformly chosen random patterns are demonstrated to obey multifractal statistics. The typical value $\alpha_0 = 0.806361\ldots$ of the rate plays the role of a Lyapunov exponent. A wide range of examples of patterns, either deterministic or random, is also investigated.

1. Introduction

Consider a data series, such as e.g. the daily temperature at a given weather station over one year. The most obvious features of such a data series are its rises and falls. Physics and other branches of science provide plenty of examples of datasets where the statistics of geometrical features, such as maxima and minima, or rises and falls, is of central interest. One example from statistical physics is provided by energy landscapes, which are ubiquitously present in theoretical studies of systems ranging from glasses to proteins [1].

In this work we investigate the probability of observing a given pattern of n rises and falls in a random stationary data series. This question has hardly been addressed so far in the physics literature, in strong contrast with the statistics of extreme values, which has recently attracted a lot of attention in many areas, including random walks, disordered systems, growth processes and random matrices [2–7].

* Tel.: +33 1 69 08 72 67; fax: +33 1 69 08 81 20.
E-mail address: jean-marc.luck@cea.fr.

http://dx.doi.org/10.1016/j.physa.2014.04.010

The setting of the present work is meant to provide a null model, to which real data could be compared. A first attempt has been made recently in this direction, with the analysis of microarray time series data in genetics [8]. We model the data series as a sequence of $n + 1$ i.i.d. (independent and identically distributed) random numbers x_i drawn from a continuous distribution. As these random numbers will only occur in inequalities, their distribution can be chosen to be uniform on the unit interval. This probabilistic approach is exposed in Section 3. An equivalent combinatorial approach (see Section 4) is obtained by coarse-graining the random numbers according to the permutation which brings them to an increasing order. We are thus led to model the data as a uniformly chosen random permutation on $n + 1$ objects. This line of thought dates back to the pioneering study of alternating permutations by André [9,10], and it has since then been addressed regularly in the mathematical literature [11–28] (this list of references is not meant to be exhaustive). To close, let us mention that the combinatorial approach to our problem pertains to the more general topic of patterns in permutations, which has been for long an active area of discrete mathematics [29–31].

2. Summary of results

Our goal is to provide a comprehensive and self-contained exposition of the calculation of the frequencies of patterns of rises and falls in a random stationary data series. We aim at using a language accessible to a broad readership in statistical physics. Let us give the detailed setup of this paper and summarize our findings.

The probabilistic and combinatorial approaches, respectively exposed in Sections 3 and 4, provide two equivalent definitions of the probability $P_n(\varepsilon_1 \dots \varepsilon_n)$ of observing a given pattern $\varepsilon_1 \dots \varepsilon_n$ of n rises and falls. The equivalence between both approaches has already been underlined in several works [8,21,23–25,27,28]. It will become clear in the following that each approach has its advantages: the probabilistic one is more suitable for analytical investigations, while the combinatorial one results in a simple recursive structure, lending itself to exact numerical calculations.

In Section 5 we show explicit results for small patterns (up to $n = 4$). We then present a heuristic analysis demonstrating that the probability of observing a pattern is essentially determined by its excursion, as long as its length is modest.

In the remainder of the paper, the emphasis is on asymptotic properties in the regime of most interest, at least from the viewpoint of statistical physics, i.e., where the length n of the pattern is large. In this regime the probability $P_n(\varepsilon_1 \dots \varepsilon_n)$ typically falls off exponentially as

$$P_n \sim e^{-\alpha n}. \tag{1}$$

The decay rate α will be our central object of interest. This quantity can be viewed as the *embedding entropy* of the binary pattern $\varepsilon_1 \dots \varepsilon_n$, i.e., the entropic cost per unit length for embedding this pattern into a sequence of i.i.d. random numbers. It is worth noticing that the above definition is entirely parameter-free. If all the 2^n patterns of length n had equal probabilities $P_n = 2^{-n}$, the rate would be constant and equal to $\alpha = \ln 2$. The observed wide range of possible values of the rate α, from $\alpha_{\min} = \ln(\pi/2) = 0.451582\dots$ to infinity, testifies the richness of the problem.

Periodic patterns are investigated in Sections 6–8. As already mentioned, the subject is an old classic of discrete mathematics. Our comprehensive approach allows us to recover many known results by more elementary means, and often to express them in simpler terms. Section 6 is a self-contained presentation of some of the beauties of the historical example of alternating patterns, for which the rate α assumes its minimal value $\alpha_{\min}$. Section 7 deals with the family of p-alternating periodic patterns, whose motif (unit cell) consists of $p - 1$ rises followed by a fall. The rate reads $\alpha = \ln z_0$, where z_0 is the smallest real positive zero of a generalized trigonometric function. In Section 8 we show how the rate α can be evaluated exactly for an arbitrary periodic pattern, with any period $p \geq 2$: z_0 is now the smallest zero of a determinant of generalized hyperbolic or trigonometric functions, whose size is at most $p/2$. Many examples are treated explicitly.

The rest of the paper covers entirely novel areas. Sections 9 and 10 serve as an intermezzo. In Section 9 we deal with examples of aperiodic patterns which are built from three classical self-similar sequences: Fibonacci, Thue–Morse, and Rudin–Shapiro. The probabilities P_n exhibit an exponential decay, characterized by a well-defined rate α, modulated by a fractal amplitude which reflects the self-similarity of the underlying sequence. Section 10 is devoted to chirping patterns, consisting mostly of rises, whereas falls are more and more scarce (or vice versa). In this case the probabilities P_n are found to decay super-exponentially. Their asymptotic form is predicted more precisely in the situation of most interest where the density of falls follows a power law.

Section 11 is devoted to the heart of the problem, namely the statistics of the probabilities P_n if patterns are chosen in various ensembles of random patterns of fixed length n. The uniform ensemble, where all patterns are considered with equal weights, is studied thoroughly. The probabilities P_n of generic patterns have the typical rate $\alpha_0 = 0.806361\dots$ The latter number can be interpreted as a Lyapunov exponent. The whole set of probabilities P_n is shown to obey multifractal statistics, with a non-trivial spectrum of multifractal dimensions $f(\alpha)$, increasing from $f(\alpha_{\min}) = 0$ to $f(\alpha_0) = 1$. Other ensembles of random patterns of fixed length n, namely the ensemble at fixed concentration c of rises and a symmetric Markovian ensemble defined by a persistence probability r, are also investigated. The probabilities P_n now generically decay according to effective typical rates $\beta(c)$ and $\gamma(r)$, which depend continuously on the ensemble parameters.

Two appendices are respectively devoted to the explicit correspondence between the combinatorial and probabilistic approaches (Appendix A) and to generalized hyperbolic and trigonometric functions (Appendix B).