Physica A 390 (2011) 120–130

Contents lists available at ScienceDirect

Physica A

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

Order–disorder transition in conflicting dynamics leading to rank–frequency generalized beta distributions

R. Alvarez-Martinez [a,c,*], G. Martinez-Mekler [b,c,**], G. Cocho [a,c]

[a] Instituto de Física, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Apartado Postal 01000, D.F., México, Mexico

[b] Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Apartado Postal 48-3, 62251, Cuernavaca, Morelos, Mexico

[c] Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Apartado Postal 01000, D.F. México, Mexico

ARTICLE INFO

Article history:
Received 31 December 2009
Received in revised form 21 July 2010
Available online 7 August 2010

Keywords:
Universality
Rank–frequency distributions
Stochastic processes
Order–disorder transitions
Expansion–modification
Rare events

ABSTRACT

The behavior of rank-ordered distributions of phenomena present in a variety of fields such as biology, sociology, linguistics, finance and geophysics has been a matter of intense research. Often power laws have been encountered; however, their validity tends to hold mainly for an intermediate range of rank values. In a recent publication (Martínez-Mekler et al., 2009 [7]), a generalization of the functional form of the beta distribution has been shown to give excellent fits for many systems of very diverse nature, valid for the whole range of rank values, regardless of whether or not a power law behavior has been previously suggested. Here we give some insight on the significance of the two free parameters which appear as exponents in the functional form, by looking into discrete probabilistic branching processes with conflicting dynamics. We analyze a variety of realizations of these so-called expansion–modification models first introduced by Wentian Li (1989) [10]. We focus our attention on an order–disorder transition we encounter as we vary the modification probability p. We characterize this transition by means of the fitting parameters. Our numerical studies show that one of the fitting exponents is related to the presence of long-range correlations exhibited by power spectrum scale invariance, while the other registers the effect of disordering elements leading to a breakdown of these properties. In the absence of long-range correlations, this parameter is sensitive to the occurrence of unlikely events. We also introduce an approximate calculation scheme that relates this dynamics to multinomial multiplicative processes. A better understanding through these models of the meaning of the generalized beta-fitting exponents may contribute to their potential for identifying and characterizing universality classes.

1. Introduction

A property X of a given system is said to follow a power law if the value of that property is proportional to X^{α}; if we think of a stochastic process, we may refer to the power law distribution of obtaining a certain value X. One of the first empirical power law determinations was carried out by Zipf [1], who studied the frequency with which words are encountered in texts. Zipf reported that by ordering the words decreasingly according to their usage in the text, he obtained a rank–frequency distribution that followed a power law with exponent −1. Recently, the search for power law behaviors has proliferated, particularly in the context of complex networks [2]. However, since in most cases the power law breaks down at

* Corresponding author at: Instituto de Física, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Apartado Postal 01000, D.F., México, Mexico.

** Corresponding author.

E-mail addresses: ralvarez@fisica.unam.mx (R. Alvarez-Martinez), mekler@fis.unam.mx (G. Martinez-Mekler), cocho@fisica.unam.mx (G. Cocho).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2010.07.037

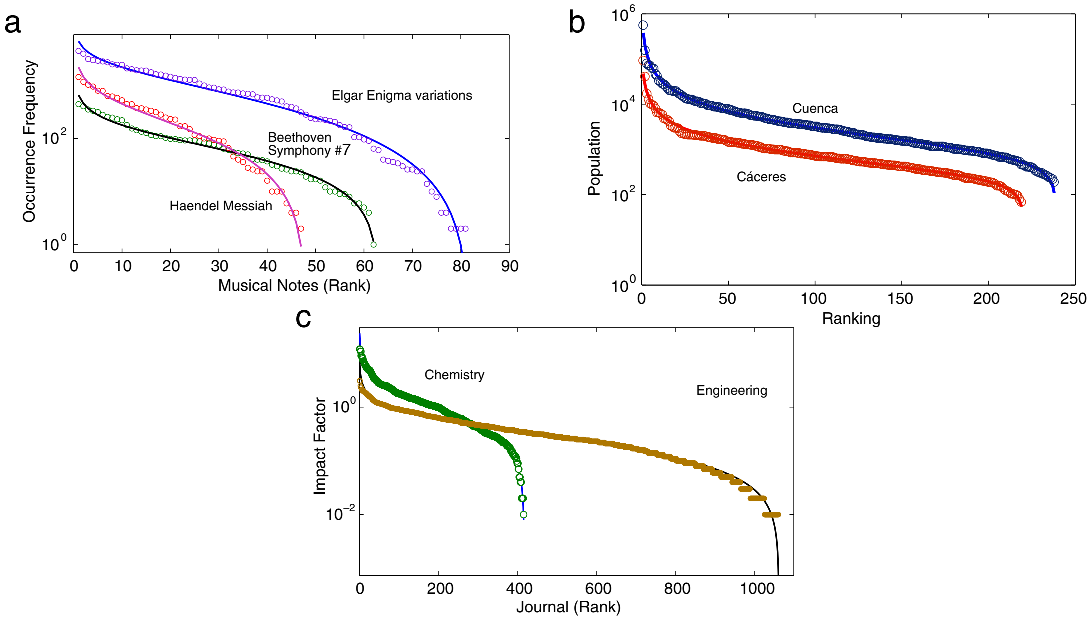

Fig. 1. Rank-ordered distributions: (a) The frequency of occurrence, ordered decreasingly, of musical notes in scores. The bold lines are the generalized beta distribution (GBD) fit with exponents and determination coefficient (a, b, R^2), for Handel's Messiah (0.55, 1.46, 0.98), Beethoven's Symphony #7 (0.39, 1.92, 0.97) and Elgar's Enigma Variations (0.39, 1.92, 0.97). (b) Population of the municipalities of the Spanish provinces of Cuenca and Cáceres with $(a, b, R^2) = (0.99, 0.50, 0.99), (0.86, 0.40, 0.99)$ respectively, the population for Cuenca is multiplied by 10 for display reasons. (c) Engineering and Chemistry journals ordered by their impact factor $(a, b, R^2) = (0.39, 0.95, 0.99), (0.51, 0.82, 0.99)$, respectively.
Source: (b): Spanish National Statistics Institute, 2009; (c): Institute for Scientific Annual Data Sets of SCI-JCR, 1974–2001.

the extremes [3], several "correction" schemes have been implemented, such as finite size scaling [4], network dilution and growth constraints [5,6], with various degrees of success. In a previous work [7], we proposed a two-parameter functional form for rank-ordered distributions which, besides encompassing both the extreme and bulk behaviors, is successfully applicable to a wide range of phenomena coming from multiple areas in both the arts and the sciences. The function we consider is a generalization of the functional expression for the beta distribution (GBD) [8] given by

$$f(r) = A\frac{(N+1-r)^b}{r^a}, \tag{1}$$

where r is the rank, which goes from 1 to the total number N of elements, A is a constant related to normalization, and a and b are fitting parameters. Fits are carried out by means of a log–log multiple linear regression. As examples, consider the frequency with which musical notes appear in musical compositions. In Fig. 1(a), fits of this GBD to Handel's Messiah, Beethoven's 7th Symphony and Elgar's Enigma Variations are shown in a semi-logarithmic plot. Notice the value close to 1 of the determination coefficient R^2. An analysis of over 1800 musical pieces [9] always gives good R^2 values, and in most cases $a < b$. In Fig. 1(b), rank-size ordered distributions related to the population in the municipalities of the Spanish provinces Cuenca and Cáceres are shown. Fits with the GBD are again impressive for the whole range of rank values. For most of the provinces in Spain, similar accuracy is achieved. Another example of social bearing is the rank-ordered journal impact factor; two instances are shown in Fig. 1(c): engineering and chemistry. Equivalent fittings hold for almost all the other disciplines, with a notable exception in medicine. In [7], we have shown that this type of behavior is recurrent in a variety of systems coming from the arts and sciences, such as rank-ordered distributions of universities, geometric motifs in abstract paintings, links in movie actor collaboration networks and genetic regulatory networks, to mention a few.

Of particular interest in our search for the significance of the GBD exponents is the case of the rank-ordered distributions of codons (nucleic acid triplets that code for amino acids) in genetic sequences. Fig. 2 shows a semi-log plot of the frequency with which the 61 possible codons (stop codons excluded) appear in the coding genetic sequences of the bacterium *Escherichia coli* and the amoeba *Entamoeba histolytica* plotted in decreasing order from the most common to the least common one. The inset corresponds to the log–log plot for *E. coli* in decreasing order: the straight line is a guide to the eye for the intermediate range of values where a power law could hold; notice the breakdown of this trend for the initial and end distribution tails. The impressive fit with the GBD holds for the genetic sequences of tens of organisms we have analyzed, coming from archea, bacteria and eukaryotes, both for amino acid and codifying codon distributions.

These findings can be related to the expansion–modification systems introduced by Wentian Li [10,11], where Boolean sequences are generated via two stochastic processes that compete with each other: one of them tends to create long-range correlations, whereas the other tends to destroy them. In the context of the neutral biomolecular evolution scheme