Physica A 391 (2012) 3043–3053

Contents lists available at SciVerse ScienceDirect

Physica A

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

On the fractal characterization of Paretian Poisson processes

Iddo I. Eliazar [a,*], Igor M. Sokolov [b]

[a] *Department of Technology Management, Holon Institute of Technology, P.O.B. 305, Holon 58102, Israel*

[b] *Institut für Physik, Humboldt-Universität zu Berlin, Newtonstr. 15, D-12489 Berlin, Germany*

ARTICLE INFO

Article history:
Received 7 December 2011
Received in revised form 12 January 2012
Available online 21 January 2012

Keywords:
Gini's index
Pietra's index
Evenness ratio
Min–max ratio
Moment ratio
Power-laws

ABSTRACT

Paretian Poisson processes are Poisson processes which are defined on the positive half-line, have maximal points, and are quantified by power-law intensities. Paretian Poisson processes are elemental in statistical physics, and are the bedrock of a host of power-law statistics ranging from Pareto's law to anomalous diffusion. In this paper we establish evenness-based fractal characterizations of Paretian Poisson processes. Considering an array of socioeconomic evenness-based measures of statistical heterogeneity, we show that: amongst the realm of Poisson processes which are defined on the positive half-line, and have maximal points, Paretian Poisson processes are the unique class of 'fractal processes' exhibiting scale-invariance. The results established in this paper are diametric to previous results asserting that the scale-invariance of Poisson processes – with respect to physical randomness-based measures of statistical heterogeneity – is characterized by exponential Poissonian intensities.

1. Introduction

Data sets consisting of vast collections of real-valued points are prevalent in science. Examples include physical measures of elements and events in large physical systems, and financial measures of firms and events in large economic systems. Vast collections of real-valued points are analyzed statistically via the corresponding statistical distributions of their points. Moreover, the common statistical approach to model a given collection of real-valued points is by Independent and Identically Distributed (IID) random variables—each random variable representing a collection point, and the distribution of the random variables being the collection's statistical distribution.

A more general statistical approach to model vast collections of real-valued points is the method of *Poisson processes* [1]. Poisson processes are a highly applicable statistical methodology to model the scattering of points in general domains, and have a wide spectrum of applications ranging from insurance and finance [2] to queueing systems [3], and from fractals [4] to power-laws [5]. Collections of real-valued points often have a distinctive maximal element. Poisson processes consisting of real-valued points and having a distinctive maximal point are termed *Poissonian populations* [6]. Poissonian populations are infinite objects, and they can be ordered as monotone decreasing sequences: the largest element, the second largest element, etc.. Consequently, given an arbitrary 'cutoff level', Poissonian populations have infinitely many points below the level, and have finitely many points above the level. The points of a Poissonian population residing above any given cutoff level form a collection of IID random variables, and hence the 'Poissonian approach' indeed generalizes the 'IID approach'. We note that the introduction of a cutoff level is practically natural, as any physical measurement is indeed limited by the sensitivity level of the measuring device applied.

A fundamental question regarding the statistical analysis of a given collection of real-valued points is the measurement of its statistical heterogeneity. Various scientific disciplines developed various measures of statistical heterogeneity such as:

* Corresponding author. Tel.: +972 507 290 650.
E-mail addresses: eliazar@post.tau.ac.il (I.I. Eliazar), igor.sokolov@physik.hu-berlin.de (I.M. Sokolov).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2012.01.030

(i) standard deviation—which is widely applied in almost all scientific fields; (ii) entropy—applied in statistical physics and in information theory as a gauge of randomness [7–10]; (iii) Hirschman's, Simpson's, and Blau's indices—essentially identical gauges of concentration and diversity applied respectively in economics, biology, and sociology [11–13][1]; (iv) inverse participation ratio—applied in physics as a gauge of localization [15–18]; (v) Gini's index—applied in economics and in the social sciences as a gauge of socioeconomic inequality [19–21].

All the aforementioned measures of statistical heterogeneity are applicable in the context of data sets modeled by IID random variables, but are not applicable in the context of data sets modeled by Poissonian populations – due to the intrinsic infiniteness of the latter. However, in the case of Poissonian populations it is possible to obtain a continuum of measures of statistical heterogeneity – a measure corresponding to any given cutoff level, and quantifying the statistical heterogeneity of the finite sub-population of points residing above the level. Now, if the continuum of measures all yield the same value – invariantly of the cutoff level – then this value indeed quantifies the statistical heterogeneity of the Poissonian population considered (and is not an artifact of the sensitivity level of the measuring device applied). Thus, we arrive at the following invariance question: *For which Poissonian populations is the measurement of statistical heterogeneity invariant with respect to the level applied?* Perceiving the level applied as the underlying Poissonian scale, the aforementioned invariance characterizes the *intrinsic fractality* of Poissonian populations.

The invariance question was introduced and analyzed in Ref. [6]. Surprisingly, there was one single class of Poissonian populations which turned out to be 'fractal' with respect to all but one of the aforementioned measures of statistical heterogeneity. This class of fractal Poissonian populations is characterized by an underlying exponential structure, and the one exception is Gini's index. In the case of Gini's index the class of fractal Poissonian populations is characterized by an underlying power-law structure. Poissonian populations quantified by a power-law structure form the class of *Paretian Poisson processes*—which are elemental in statistical physics [22], and are the bedrock of a host of power-law statistics ranging from Pareto's law to anomalous diffusion [5].

The distinction between Gini's index and all the other aforementioned measures of statistical heterogeneity is profound. Gini's index measures the evenness of the data set considered—as it quantitatively answers the socioeconomic question: "How equal is the distribution of wealth in a given society?". On the other hand, all the other aforementioned measures of statistical heterogeneity provide various physical gauges of dispersion, localization, and randomness. Thus, in the context of Poissonian populations, 'physical perspectives' of statistical heterogeneity yield fractality characterized by an underlying exponential structure, whereas a 'socioeconomic perspective' of statistical heterogeneity yields fractality characterized by an underlying power-law structure.

The goal of this paper is to further explore measures of statistical heterogeneity which lead to Paretian Poisson processes. In this paper we consider various evenness-based and evenness-related measures of statistical heterogeneity, and show that there is one single class of Poissonian populations which turns out to be fractal with respect to these measures—the class of Paretian Poisson processes. The results obtained in this paper further fortify and reinforce – in the context of fractality of Poissonian populations – the aforementioned dichotomy between the underlying exponential and power-law structures.

The paper is organized as follows. We begin with a review of Poissonian populations (Section 2), and a review of the measurement of statistical heterogeneity of Poissonian populations (Section 3). We then turn to explore the fractality of Poissonian populations with respect to various stochastic gauges of statistical heterogeneity: normalized random variables underlying Gini's and Pietra's indices (Section 4),[2] evenness ratio (Section 5), min–max ratio (Section 6), and moment ratios (Section 7). In what follows we shall prove that the classes of Poissonian populations that are fractal with respect to these measures of statistical heterogeneity all coincide with the class of Paretian Poisson processes.

2. Poissonian populations

Consider a collection of points $\mathcal{P}$ scattered randomly on the real interval (a, b) $(-\infty \leq a < b \leq \infty)$. The standard statistical approach to model $\mathcal{P}$ is as a collection of IID random variables. Namely, each point of $\mathcal{P}$ is considered to be a random variable which is independent of the other points, and all the random variables are considered to be governed by a common probability density function $f(x)$ $(a < x < b)$. According to the probabilistic approach, if we focus on a single point then the probability of finding it in the infinitesimal interval $(x, x + dx)$ is given by $f(x)\,dx$.

A more general statistical approach to model $\mathcal{P}$ is as a Poisson process on the interval (a, b) [1]. According to the Poisson approach $\mathcal{P}$ is governed by a Poissonian intensity function $\lambda(x)$ $(a < x < b)$. Informally, this means that: the probability that the infinitesimal interval $(x, x + dx)$ contains a point is given by $\lambda(x)\,dx$, and the probability that the infinitesimal interval $(x, x + dx)$ is empty is given by $1 - \lambda(x)\,dx$ (independently of all other infinitesimal intervals). The precise formulation of the Poisson approach is defined by the two following rules [1]: (i) the number of points residing in the sub-interval (a', b') is a Poisson-distributed random variable with mean $\int_{a'}^{b'} \lambda(x)\,dx$; (ii) the numbers of points residing in disjoint sub-intervals are independent random variables.

[1] Hirschman's index was invented by the German-born economist Albert Otto Hirschman in 1945 [11], reinvented in 1950 by the economist Orris Herfindahl, and is commonly – yet mistakenly – referred to as the "Herfindahl index" or the "Herfindahl–Hirschman index" [14].

[2] Pietra's index is measure of socioeconomic inequality, which will be described and explained hereinafter.