Point Processes Modeling of Time Series Exhibiting Power-Law Statistics

B. Kaulakys, M. Alaburda and V. Gontis

Institute of Theoretical Physics and Astronomy of Vilnius University, A. Gostauto 12, LT-01108 Vilnius, Lithuania

Abstract. We consider stochastic point processes generating time series exhibiting power laws of spectrum and distribution density (*Phys. Rev. E* **71**, 051105 (2005)) and apply them for modeling the trading activity in the financial markets and for the frequencies of word occurrences in the language.

Keywords: point processes, 1/f noise, stochastic differential equations, power-law distributions
PACS: 05.40. .a, 72.70. +m, 89.75.Da

INTRODUCTION

Recently, we proposed [1] and generalized [2] the stochastic point process models generating a variety of monofractal and multifractal time series exhibiting power laws of the spectrum $S(f)$ and of the distribution $P(I)$ of the signal intensity and applied them for the analysis of the financial systems [3]. These models can generate $1/f$ noise with a very large Hooge parameter. They may be used as the theoretical framework for understanding huge fluctuations (see, e.g., [4]) and as well for description of a large variety of observable statistics, i.e., jointly power spectral density (PSD), $S(f) \sim 1/f^\beta$, signal probability distribution function (PDF), $P(I) \sim 1/I^\lambda$, with different slopes, different distributions, $P(\tau)$, of the interevent time τ and different multifractality.

Here we will present the extensions and generalizations of the point process models for the Poissonian-like processes with slowly diffusing mean interevent time [5]. We will adjust the parameters of the generalized model to the empirical data of the trading activity in the financial markets [6] and to the frequencies of the word occurrences in the language, reproducing the PDF and PSD.

THE MODEL

We investigate stochastic time series as a sequence of events which occur at discrete times $t_1, t_2, ... t_k, ...$ and can be considered as identical point events. Such point process equivalently is defined by the set of stochastic interevent times $\tau_k = t_{k+1} - t_k$. Let us consider the flow of events as the Poissonian-like process driven by the multiplicative stochastic equation. We define the stochastic rate $n = 1/\tau$ of event flow by continuous stochastic differential equation

$$\mathrm{d}\tau = \left[\gamma - \frac{m}{2}\sigma^2 \left(\frac{\tau}{\tau_0}\right)^m\right] \tau^{2\mu-2}\mathrm{d}t + \sigma\tau^{\mu-1/2}\mathrm{d}W, \qquad (1)$$

535

where W is a standard Wiener process, σ denotes the standard deviation of the white noise, $\gamma \ll 1$ is a coefficient of the nonlinear damping and μ defines the power of noise multiplicativity. The diffusion of τ is restricted from the side of high values by an additional term $-\frac{m}{2}\sigma^2 \left(\frac{\tau}{\tau_0}\right)^m \tau^{2\mu-2}$, which produces the exponential diffusion reversion. m and τ_0 are the power and value of the diffusion reversion, respectively. The associated Fokker-Plank equation with the zero flow gives the simple stationary PDF

$$P(\tau) \sim \tau^{\alpha+1} \exp\left[-\left(\frac{\tau}{\tau_0}\right)^m\right] \tag{2}$$

with $\alpha = 2(\gamma_\sigma - \mu)$ and $\gamma_\sigma = \gamma/\sigma^2$. Eq. (1) describes continuous stochastic variable τ, defines rate $n = 1/\tau$ with stationary distribution and PSD $S(f)$ [5, 6],

$$P(n) \sim \frac{1}{n^\lambda} \exp\left\{-\left(\frac{n_0}{n}\right)^m\right\}, \quad \lambda = 1 + \alpha, \tag{3}$$

$$S(f) \sim \frac{1}{f^\beta}, \quad \beta = 1 + \frac{\alpha}{3-2\mu}. \tag{4}$$

Here we define the fractal point process driven by the stochastic differential equation (1), i.e., we assume $\tau(t)$ as slowly diffusing mean interevent time of the Poissonian-like process with the stochastic rate n. Within this assumption the conditional probability of interevent time τ_p in the Poissonian-like process with the stochastic rate $1/\tau$ is

$$\varphi(\tau_p|\tau) = \frac{1}{\tau}\exp\left[-\frac{\tau_p}{\tau}\right]. \tag{5}$$

Then the long time distribution $P_k(\tau_p)$ of interevent time τ_p in k-space [2] has the integral form

$$P_k(\tau_p) = C \int_0^\infty \exp\left[-\frac{\tau_p}{\tau}\right] \tau^{\alpha-1} \exp\left[-\left(\frac{\tau}{\tau_0}\right)^m\right] d\tau, \tag{6}$$

with C defined from the normalization, $\int_0^\infty P_k(\tau_p)d\tau_p = 1$. The distributions of interevent time τ_p have their explicit forms for the integer values of power m. For $m = 1$ and for $m > 1$ they are expressed by the modified Bessel function [5, 6] and in terms of the hypergeometric functions, respectively.

DISCUSSION

We consider two applications of the proposed model. The frequencies of the word occurrences in the language depend on the content and diffuse in the text. We consider the flow of words in the text as the steps of discrete events, i.e., one word is the unit of the interval. Then the number of other words in between of the two successive occurrences of the same noun measures the interevent interval τ_p of the point process defined for the sequence of selected noun. One can easily calculate the sequence of all selected word occurrence intervals τ_p and so define the realization of the point
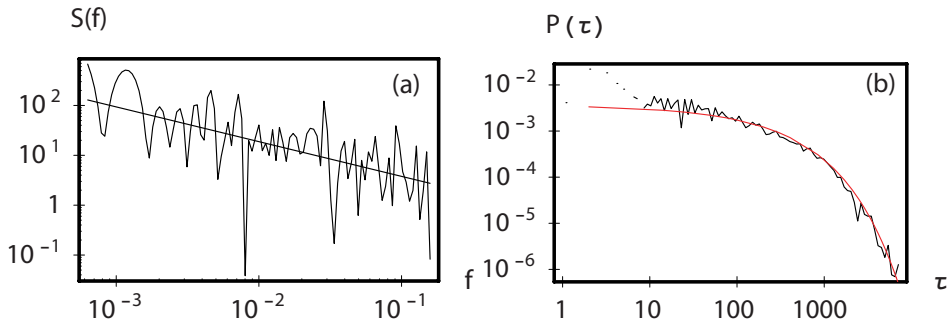
FIGURE 1. (a) The power spectral density $S(f)$ of the point process for word "eye" occurrences in the novels of Jack London. The straight line approximates the power-law with the exponent $\beta = 0.7$. (b) The interevent interval τ distribution of the word "eye" occurrences calculated from the histogram in the same novels. The smooth line represents the integral formula (6) with $m = 2$, $\tau_0 = 800$ and $\alpha = -0.3$.

process. Here we demonstrate the statistics of the word "eye" in the selected novels of Jack London, over 1.2 mln. words totally. First of all, we demonstrate that the point process, defined in such a way, has long memory as the exponent of PSD $\beta = 0.7$ [Fig. 1 (a)]. With the assumption of pure multiplicative process with $\mu = 1$ from the relation $\beta = 2\gamma_\sigma - 1 = 1 + \alpha$ one defines the parameter $\alpha = -0.3$ related with the probability distribution functions. The histogram of τ_p distribution coincides with theoretical PDF defined by its integral form (6) when $m = 2$ [Fig. 1 (b)]. The exponent λ of the power-law distribution for number of the word "eye" occurrences in the 1000 words length pieces of text is $\lambda = 4.4$. As we will see later, the presented example of the word statistics resembles the statistical properties of trading activity in the financial markets.

In the case of the financial market we consider every trade of the selected stock as a point event, i.e. the sequence of all trades for the stock composes the stochastic point process, described by the set of time intervals between the successive trades. The power spectral density of the trade sequence serves as a measure of the long range memory property of trading activity. An example of the spectrum for the stock CVX trade sequence traded in the period of two years on NYSE, Fig.2 (a), reveals the structure of the power spectral density in a wide range of frequencies and shows that the real markets exhibit two power laws with the exponents $\beta_1 = 0.8$ and $\beta_2 = 0.3$. In our recent works [5, 6] we have proposed the model adjustment, introducing a new form of the modulating stochastic differential equation instead of Eq. (1),

$$d\tau = \left[\gamma - \frac{m}{2}\sigma^2\tau^m\right]\frac{1}{(\varepsilon + \tau)^2}dt + \sigma\frac{\sqrt{\tau}}{\varepsilon + \tau}dW, \qquad (7)$$

where a new parameter ε defines the crossover between two areas of τ diffusion with assumption $\tau_0 = 1$. The solution of Eq. (7) has to be scaled by τ_0 for other values of τ_0. The Poisonian-like point process modulated by Eq. (7) reproduces PSD, Fig.2 (a), of the empirical trade sequence in detail, including two exponents and the crossover point, Fig.2 (b). The proposed model with the same parameters reproduces the empirical PDF of τ_p, Fig.2 (c) and (d), very well. Moreover, the model describes the distribution of the
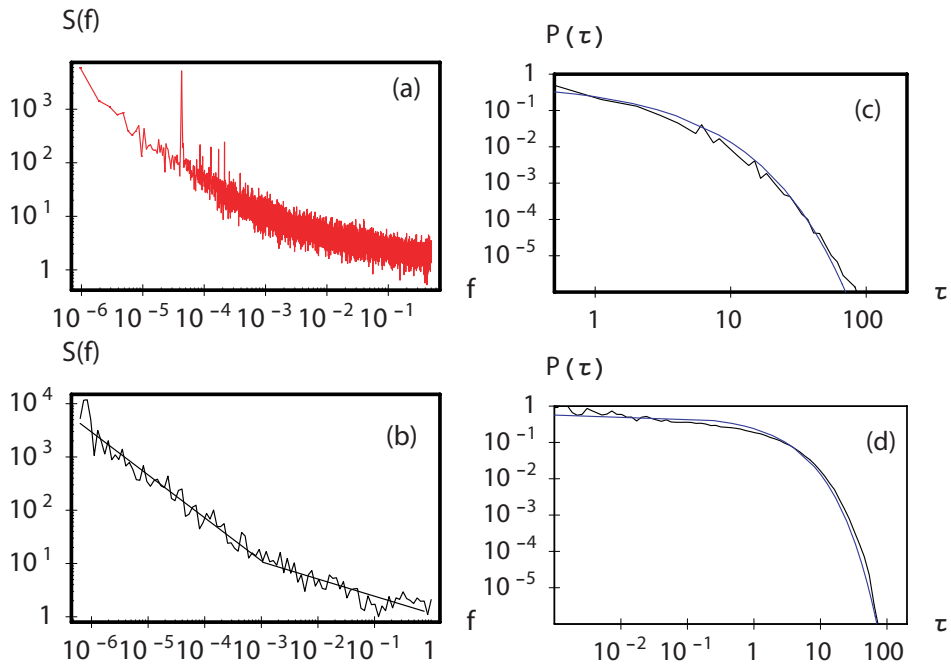
FIGURE 2. (a) The power spectral density $S(f)$ of the stock CVX trade sequence traded on the NYSE. (b) Power spectral density of the Poissonian-like process driven by Eq. (7) with the parameters $\tau_0 = 4$, $\sigma = 0.023$, $\gamma = 0.00034$, and $\varepsilon = 0.06$. Straight lines approximate power-law spectrum with exponents $\beta_1 = 0.8$ and $\beta_2 = 0.3$. (c) The empirical distribution of τ_p calculated from the histogram of CVX trades on NYSE. The smooth line represents the integral formula (6) for $m = 2$, $\tau_0 = 4$ and $\alpha = -0.3$. (d) The model distribution of τ_p calculated with same parameters as in (b), the smooth line is the same as in (c).

empirical trading activity, i.e., the number of transactions per selected time window with the power-law exponent $\lambda = 4.4$.

ACKNOWLEDGMENTS

REFERENCES

1. B. Kaulakys and T. Meškauskas, *Phys. Rev. E* **58**, 7013 (1998).
2. B. Kaulakys, V. Gontis, and M. Alaburda, *Phys. Rev. E* **71**, 051105 (2005).
3. V. Gontis and B. Kaulakys, *Physica A* **343**, 505 (2004); **344**, 128 (2004).
4. W. Kruppa, M. G. Ancona, M. W. Rendell *et al.*, *Appl. Phys. Lett.* **88**, 053120 (2006).
5. V. Gontis and B. Kaulakys, *J. Stat. Mech.* **10**, P10016 (2006).
6. V. Gontis and B. Kaulakys, *Physica A* doi:10.1016/j.physa.2007.02.012 (2007).