Physics Letters A 380 (2016) 1517–1525

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Statistical signatures of structural organization: The case of long memory in renewal processes

Sarah E. Marzen [a], James P. Crutchfield [b,*]

[a] *Department of Physics, University of California at Berkeley, Berkeley, CA 94720-5800, United States*
[b] *Complexity Sciences Center and Department of Physics, University of California at Davis, One Shields Avenue, Davis, CA 95616, United States*

ARTICLE INFO

Article history:
Received 7 December 2015
Received in revised form 28 February 2016
Accepted 29 February 2016
Available online 2 March 2016
Communicated by C.R. Doering

Keywords:
Fractal renewal process
Statistical complexity
Excess entropy
Power-law scaling
1/f noise
Zipf's law

ABSTRACT

Identifying and quantifying memory are often critical steps in developing a mechanistic understanding of stochastic processes. These are particularly challenging and necessary when exploring processes that exhibit long-range correlations. The most common signatures employed rely on second-order temporal statistics and lead, for example, to identifying long memory in processes with power-law autocorrelation function and Hurst exponent greater than 1/2. However, most stochastic processes hide their memory in higher-order temporal correlations. Information measures—specifically, divergences in the mutual information between a process' past and future (excess entropy) and minimal predictive memory stored in a process' causal states (statistical complexity)—provide a different way to identify long memory in processes with higher-order temporal correlations. However, there are no ergodic stationary processes with infinite excess entropy for which information measures have been compared to autocorrelation functions and Hurst exponents. Here, we show that fractal renewal processes—those with interevent distribution tails $\propto t^{-\alpha}$—exhibit long memory via a phase transition at $\alpha = 1$. Excess entropy diverges only there and statistical complexity diverges there and for all $\alpha < 1$. When these processes do have power-law autocorrelation function and Hurst exponent greater than 1/2, they do not have divergent excess entropy. This analysis breaks the intuitive association between these different quantifications of memory. We hope that the methods used here, based on causal states, provide some guide as to how to construct and analyze other long memory processes.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Many time series of interest have "short memory", meaning (loosely speaking) that knowledge of the past confers exponentially diminishing returns for predicting the future. However, many other time series of interest—those with "long memory"—exhibit intrinsic timescales that grow without bound as the amount of available data increases [1–6]. Examples include the hydrological data first studied by Hurst [7] and modeled by Mandelbrot [8] and many others, e.g., see Refs. [9,10].

These are qualitatively different processes that demand qualitatively different generative models. In other words, signatures of long memory imply a kind of structural organization of the underlying process that differs from one with short memory. This is the *inverse problem* of long memory: Which statistical signatures identify, uniquely or not, which intrinsic organizations? Sharp answers are critical to successful empirical analysis and often provide necessary first steps in predictive theory building. The complementary *forward problem*, an open question, is to identify the kinds of memoryful process structure that lead to one or another statistical signature. Answering this question requires defining statistical signatures that quantify memory in stochastic processes.

Many existing quantifications of long memory are based on second-order statistics; e.g., on using the autocorrelation function, power spectrum, or Hurst exponent. These approaches have had notable successes in analyzing hydrological data [7,9], music [4], spin systems [2], astrophysical flicker noise [6], language [11,12], natural scenery [13,14], communication system error clustering [15], financial time series, and many other seemingly complex phenomena [5,16].

However, there are at least two reasons to look to other statistics besides the Hurst exponent. First, second-order statistics alone can be misleading, as most stochastic processes seem to hide information about their temporal dependencies in higher-order statistics [17,18]. Second, as suggested in Ref. [19], our determination of whether or not a process has long memory ideally should be invariant under invertible transformations of one's measurement

* Corresponding author.
E-mail addresses: smarzen@berkeley.edu (S.E. Marzen), chaos@ucdavis.edu (J.P. Crutchfield).

http://dx.doi.org/10.1016/j.physleta.2016.02.052
0375-9601/© 2016 Elsevier B.V. All rights reserved.

values. The challenge is not only to find a new statistic that addresses these two concerns, but to find a statistic that is also easy to operationalize.

References [20–22] suggested a process might be said to have long memory when the mutual information between its past and future (excess entropy) diverges, and Ref. [20] suggested that long memory is associated with divergent statistical complexity with the effective memory architecture given by a process' ϵ-machine. By construction, these statistics are invariant under invertible transformations of the data; and with sufficiently clever entropy estimation techniques, these statistics are also calculable directly from time series data.

Unfortunately, there is a paucity of concrete examples upon which to build intuition as to how these higher-order statistics and the more commonly used second-order statistics relate. In part, this lack of concrete examples might owe somewhat to the fact that it is nontrivial to construct ergodic stationary processes with divergent excess entropy, though see Refs. [23,24]. (Note that the processes considered in Ref. [21] were nonergodic [25].)

To that end, we study a tractable class of processes that can have both divergent excess entropy and Hurst exponent greater than $1/2$: the *fractal renewal processes* [26–29] in which interevent intervals are drawn independently and identically (IID) from a probability distribution with tails $\propto t^{-\alpha}$. These processes are very widely used in the physical, biological, and social sciences to model diverse long-memory phenomena, ranging from current fluctuations in electronic devices and neuronal spike trains to earthquakes and astrophysical time series [30–39].

Previous studies analyzed the second-order statistics of such processes in some detail [9,40]. Here, we use techniques inspired by those in Refs. [23,24] to calculate the excess entropy and statistical complexity of fractal renewal processes for the first time. We find that fractal renewal processes have divergent excess entropy only and exactly when $\alpha = 1$ and divergent statistical complexity as $\alpha \to 1$ from above and for all $0 < \alpha < 1$. However, fractal renewal processes have power-law power spectra for all $0 < \alpha < 2$ [40] and Hurst exponents greater than $1/2$ [9]—the latter being two of the conventional second-order statistical signatures of "long memory". Thus, even for these relatively straightforward processes, the excess entropy and statistical complexity encapsulate a different notion of long memory than one gleans using only second-order statistics. These results also add fractal renewal processes to a very short list of known stationary ergodic processes with divergent excess entropy [23,24] and so, we hope, pave the way for more general comparisons between different definitions of long memory.

Section 2 briefly reviews definitions of memory in stochastic processes. Section 3 calculates informational measures of memory for fractal renewal processes. Section 4 then compares our findings to the second-order statistics calculated by Refs. [9,40] and draws out the lessons for the above application examples. We close by reflecting on structural organization associated with long memory.

2. Background

There are many definitions for a stochastic process to have long memory; Ref. [19] provides a particularly helpful survey. Consider a sequence of ℓ observations $x_0, x_1, \ldots, x_{\ell-1}$, realizations of discrete-valued random variables $X_0, X_1, \ldots, X_{\ell-1}$. For instance, if the *autocorrelation function* $\mathcal{C}(\tau)$ is asymptotically a power law multiplied by a slowly varying function $g(\tau)$, then a process can be said to have "long memory":

$$\begin{aligned}\mathcal{C}(\tau) &= \sigma^{-2}\sum_{j=0}^{\ell}(x_j-\mu)(x_{j+\tau}-\mu)\\ &\propto g(\tau)\tau^{-\gamma}\,,\end{aligned}$$

with $0 < \gamma < 1$, mean μ, and variance σ^2. Yet other definitions are based on the decay of the *spectral density*:

$$\mathcal{P}(f) = \ell^{-1}\left|\sum_{j=0}^{\ell} x_j e^{-ijf}\right|^2 .$$

The process has long memory when $\mathcal{P}(f) \propto f^{-\beta}L_1(f)$ as f approaches 0 with $0 < \beta < 1$, where $L_1(f)$ is a slowly varying function near $f = 0$. Other definitions still are based on how variances deviate from time-local linear extrapolation. Starting with the variance of partial sums $S_j = X_1 + \cdots + X_j$, one uses the *rescaled-range* statistics:

$$\begin{aligned}RS(\ell) &= \frac{\max_{0\le j\le \ell}(S_j - \frac{j}{\ell}S_\ell) - \min_{0\le j\le\ell}(S_j - \frac{j}{\ell}S_\ell)}{\sigma}\\ &\propto \ell^{-H}\,,\end{aligned}$$

where $H \in (0,1)$ is the *Hurst index*. Processes with $H > 1/2$ are interpreted as having long memory. Unfortunately, even these second-order statistics are not always equivalent signatures of long memory. See Ex. 5.2 of Ref. [19] for an example of a process in which the spectral density but not correlations are regularly varying.

In a search for general principles from ergodic theory, Sec. 4 of Ref. [19] proposed that we require a definition of long memory independent of invertible transformations of the data. That is, if an invertible transformation is applied pointwise to each observation X_i, we would hope that the resulting process has long memory if and only if the original process had long memory [41]. This desideratum is not always satisfied by definitions based on the above second-order statistics, though see Thm. 4.1 of Ref. [19].

Since strongly mixing processes have short memory and nonergodic processes could be said to have infinite memory [25], Ref. [19] proposed that one or another type of nonmixing property is a good candidate for long memory in ergodic stationary processes. This criterion satisfies the invariance desideratum above but can be rather difficult to evaluate.

Fortunately, the information-theoretic notions of memory we consider also satisfy the transformation-invariant desideratum and have been successfully deployed as quantifications for the "complexity" of stochastic processes [21,42]. We study two: the *excess entropy* $\mathbf{E} = I[\overleftarrow{X};\overrightarrow{X}]$, or the mutual information between a process' past $\overleftarrow{X} = \ldots X_{-3}X_{-2}X_{-1}$ and future $\overrightarrow{X} = X_0X_1X_2\ldots$ [22]; and the *statistical complexity* C_μ, or the amount of information from the past $\overleftarrow{X}$ required to predict the future $\overrightarrow{X}$ as well as possible [42]. When the excess entropy diverges, we are interested in the asymptotic rate of divergence of finite-length excess entropy estimates $\mathbf{E}(\ell) = I[\overleftarrow{X};\overrightarrow{X}^{\ell}]$ [21,22]. This asymptotic rate of divergence is also invariant to temporally local convolutions and invertible transformations of the data [21].

To more precisely define and calculate the statistical complexity and the excess entropy, we need to recall the causal states of computational mechanics. Consider clustering pasts according to an equivalence relation $\sim$ in which two pasts are equivalent when they have the same conditional probability distribution over futures: $\overleftarrow{x} \sim \overleftarrow{x}'$ if and only if $\Pr(\overrightarrow{X}\,|\,\overleftarrow{X} = \overleftarrow{x}) = \Pr(\overrightarrow{X}\,|\,\overleftarrow{X} = \overleftarrow{x}')$. The resulting clusters are *forward-time causal states* $\mathcal{S}^+$, which inherit a probability distribution from the probability distribution over pasts. The *forward-time statistical complexity* is the entropy of these causal states: $C_\mu^+ = H[\mathcal{S}^+]$. For more detail, see Refs. [43,44].

We can similarly define the *reverse-time causal states* $\mathcal{S}^-$ by clustering futures with equivalent conditional probability distributions over pasts: $\overrightarrow{x} \sim \overrightarrow{x}'$ if and only if $\Pr(\overleftarrow{X}\,|\,\overrightarrow{X} = \overrightarrow{x}) = \Pr(\overleftarrow{X}\,|\,\overrightarrow{X} = \overrightarrow{x}')$. The *reverse-time statistical complexity* is the entropy of those reverse-time causal states: $C_\mu^- = H[\mathcal{S}^-]$. Renewal processes are time-reversal invariant [45], or *causally reversible*, so

Download English Version:

https://daneshyari.com/en/article/1860736

Download Persian Version:

https://daneshyari.com/article/1860736

Daneshyari.com