Physica A 448 (2016) 132–143

Contents lists available at ScienceDirect

Physica A

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

Universal and non-universal properties of recurrence intervals of rare events

Xiaojun Zhao [a], Pengjian Shang [b], Aijing Lin [b,*]

[a] School of Economics and Management, Beijing Jiaotong University, Beijing, 100044, China

[b] Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing, 100044, China

HIGHLIGHTS

- Recurrence intervals of long-range persistent records obey stretched exponential.
- Recurrence intervals of long-range anti-persistent records obey exponential.
- Recurrence intervals of long-range anti-persistent records are uncorrelated.
- Non-universal properties of recurrence intervals of short-range autocorrelated records exist.

ARTICLE INFO

Article history:
Received 22 August 2015
Received in revised form 27 October 2015
Available online 29 December 2015

Keywords:
Recurrence interval
Rare event
Long-range correlation
Stretched exponential
ARFIMA

ABSTRACT

This paper is devoted to the statistical analysis on the recurrence intervals of rare events that are defined above a given threshold. The memory property of original records is found to have significant effects on the distribution and the correlation structure of recurrence intervals, so that some universal and non-universal properties arise. (i) For long-range persistent records by the ARFIMA processes, where large values are likely to follow large values, the recurrence intervals yield stretched exponential distributions and further show long-range persistence. (ii) For long-range anti-persistent records by the ARFIMA processes, the recurrence intervals obey exponential distributions, also absence of autocorrelation. (iii) For short-range autocorrelated records, the distribution and the correlation structure of recurrence intervals both depend on the parameters of the model and the threshold of rare events.

1. Introduction

Nowadays, it has become a multi-disciplinary topic of the rare events, e.g., the earthquakes in geophysics [1–3], the floods and droughts in hydrology [4,5], the hurricanes in climate [6], the stock crashes in finance [7,8], the criminal activities in society [9], and the traffic jams [10]. The study on rare events is of particular importance for both scientific researches and practical applications [11–15]. Rare events, typically low in occurrence possibility while extremely large (or extremely small) in amplitude, usually have the potential to cause giant effects. Unfortunately, it is rather difficult to understand the mechanism behind these behaviors. In each of these systems, too many factors nonlinearly interact with each other, leading their outputs to be erratic, unstable, or chaotic. The data-based techniques concerning time series analysis make it possible for us to gain insight into the statistical behaviors that emerge from the aggregate performances of the factors. In statistics, a

* Corresponding author.
E-mail address: ajlin@bjtu.edu.cn (A. Lin).

http://dx.doi.org/10.1016/j.physa.2015.12.082

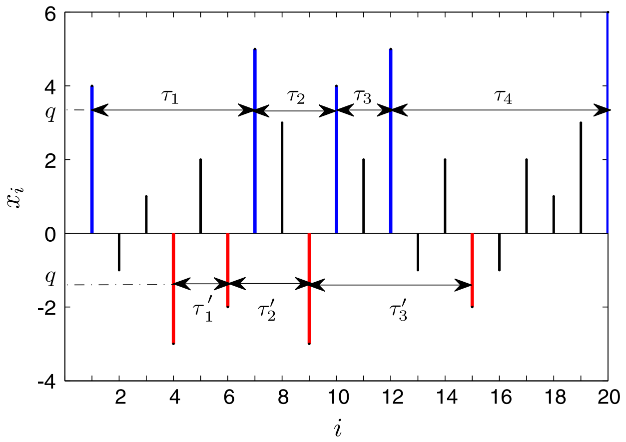

Fig. 1. A set of records with length 20. When the threshold q is set 3.2 and rare events are defined as the records above q (blue), we get 4 recurrence intervals: $\tau_1 = 6$, $\tau_2 = 3$, $\tau_3 = 2$, and $\tau_4 = 8$. When the threshold q is set -1.5 and rare events are defined as the records below q (red), we get 3 recurrence intervals: $\tau'_1 = 2$, $\tau'_2 = 3$, and $\tau'_3 = 6$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

record x_i, $i \in \{1, 2, \ldots, N\}$ is called a rare event if its value is above (or below) a given threshold q. To predict the occurrence of rare events, it is crucial to analyze the time intervals between successive rare events, i.e. the recurrence intervals or return intervals [16–25] (see Fig. 1).

Consider a simple model. Assume the rare events are spontaneous, which randomly appear. The presence (or absence) of a rare event is thus considered as a Bernoulli trial; the occurrence probability of each rare event is estimated by its occurrence frequency: $P = \#\{x_i|x_i > q\}/N$, where # represents the cardinality of set and N represents the length of series. As a result, the occurrence number of rare events, described by y, would yield a binomial distribution $y \sim B(N, P)$. If $P \to 0$ (for rare events) and $N \to \infty$, the binomial distribution is approximately equivalent to the Poisson distribution, owning the only parameter $\lambda = NP$:

$$P(y = k) = \frac{\lambda^k}{k!} \exp(-\lambda), \tag{1}$$

for $k = 0, 1, \ldots, N$, where k represents the total number of rare events in the series. Of note, $N \to \infty$ and $p \to 0$ need to be satisfied: (i) For $N \to \infty$, the probability of rare events is approximate to their frequency by the law of large numbers. (ii) For $N \to \infty$ and $p \to 0$, the binomial distribution of rare events can be approximated by the Poisson distribution.

According to the property of the Poisson distribution, the distribution of recurrence intervals τ between successive rare events theoretically obeys an exponential function with the parameter $\lambda/N = P$. Due to $1/P = N/\#\{x_i|x_i > q\} \simeq \bar{\tau}$ ($\bar{\tau}$: mean recurrence interval), we obtain the distribution of recurrence intervals:

$$P(\tau) = \frac{1}{\bar{\tau}} \exp(-\tau/\bar{\tau}). \tag{2}$$

The memoryless property of exponential functions, i.e. $P(\tau \le n) = P(\tau \le n + t|\tau \ge t)$, indicates the waiting time for the next rare event cannot be inferred even if the elapsed time t ($t > 0$) is known.

Nowadays, some recent studies involving areas of geography [26,27], hydrology [28,29], climate [30,31], and finance [32] indicate that rare events do not appear randomly. Records in many systems were found to show long-range persistence [33–35]. Consequently, the clustering of rare events (also the clustering of non-rare events) make that the probabilities of having recurrence intervals well below $\bar{\tau}$ and well above $\bar{\tau}$ are strongly enhanced in the correlated records, where $\bar{\tau}$ is closely related with the threshold q [36]. Furthermore, it was revealed that the distribution of recurrence intervals for long-range persistent records by the Fourier-filtering method generally followed a stretched exponential function [30],

$$\ln[P(\tau)\bar{\tau}] \sim -(\tau/\bar{\tau})^{\gamma}, \quad 0 < \gamma \le 1. \tag{3}$$

γ is the scaling exponent of the autocorrelation function of original records: $c(s) = (x_i - \bar{x})^T (x_{i+s} - \bar{x})/[\sigma(x)^2 (N - s)] \sim s^{-\gamma}$, where $\bar{x}$, $\sigma(x)^2$ represent the mean value and variance, respectively. The fundamental basis is when considering the problem of zero-level crossings in long-range persistent data for Gaussian stationary processes, the probability of having no zero-level crossing after t time steps is bounded from above by a stretched exponential [37,38]. Moreover, it was found that the recurrence intervals were also long-range correlated, with the same scaling exponent γ of original records [17,30,39]. It is, therefore, an indication of the clustering that large recurrence intervals are more likely to be followed by large recurrence intervals while small recurrence intervals are more likely to be followed by small recurrence intervals. The scaling is important, since it allows us to extrapolate the behavior at very large q values (rare events) from the behaviors at small q values that are not rare and therefore have good statistics. Furthermore, the conditional recurrence intervals also