Analysis of ratios of consecutive recurrence time and its connection to Thomae function

Ritam Pal and M. S. Santhanam
Department of Physics, Indian Institute of Science Education and Research, Dr. Homi Bhabha Road, Pune 411008, India.

I. INTRODUCTION

- Motivate the study of extreme events.
- Important questions to answer.
- Recent attempts to solve those.
- Why ratios in the context of times series, motivate it from RMT perspective. Also the scientific report paper.

II. MODEL AND METHODS

- time series
- exponentially correlated time series.
- extreme event and records.
- recurrence time, its ratio, record age and its ratio.

Consider a process $x(t)$, which can denote position of a particle under brownian motion or temperature of a city over time. If we sample the process with uniform time gaps then we will get a time ordered set of points $\{x_1, x_2, x_3...\}$. This is called a time series.

If $x_i \in \{x_1, x_2, x_3...\}$ are iids then the auto correlation function of the time series which is defined as $C(t_1, t_2) = \langle x(t_1)x(t_2)\rangle$ is a delta function $\delta(t_1 - t_2)$. But most real world times series have non vanishing auto correlation function. In this paper we primarily focus on time series with exponential auto correlation function (weakly correlated) of the form $e^{-\frac{t_2-t_1}{\tau}}$. To generate exponentially correlated time series we adapt a well known method []. A brief algorithm to generate a time series with correlation length τ is the following:

- 1. $x_1 = r_1$; r_1 is drawn from Gaussian distribution.
- 2. $x_{i+1} = e^{-\frac{1}{\tau}} \times x_i + \sqrt{1 - e^{-\frac{2}{\tau}})} \times r_i$; r_i is drawn from Gaussian distribution.

Extreme events – the events that exceed a predefined threshold. for our analysis purpose we define the threshold to be $q = \langle x(t)\rangle + m\sigma$. where m controls how far way the extremes are from typical value. To formally study how extreme events are distributed over times, we first define recurrence time. We define i-th recurrence time $s_i = t_{i+1} - t_i$, where t_i is the time for i-th extreme event. From consecutive recurrence time we also define ratio of recurrence time as $R_i = \frac{s_{i+1}}{s_{i+1}+s_i}$. As s_i's are integers R_i only takes rational values, and by construction it is bounded between zero and one. In this paper we will focus on studying the behavior of the distribution of R, and comment about the correlation of the time series.

simulation details: Time series of length 10^7 is generated according to the above mentioned algorithm. and all the distributions are averaged over 50 realizations.

III. RESULTS

- basic description of the distribution as and arguments that support the shape of the distribution.
- The qualitative changes in the distribution, when the threshold for the extreme events and the correlation length of the time series is varied. Arguments to support those changes.
- quantitative analysis of few peaks with varying parameters. And arguments.
- comparison with real data. What data. how it is prepared and how well does it match with the distributions with the simulated one.
- quantitative analysis of the box counting dimension of the distribution and with varying parameters. Arguments to support those.
- generalized thomae function. It's not possible to find a closed form for the distribution for all parameter value, but can be argued to a certain general form. Find the generalized thomae function that matches the distribution at at least some limit of the parameter values.

First we study the distribution of R for uncorrelated time series. As ratios are rations, which means they are countable, makes $P(R)$ discrete. For large threshold ($m = 2$) we find the distribution to be a scaled Thomae function, which is defined to be zero at irrational points and $\frac{1}{q}$ at rational point $\frac{p}{q}$. The box counting dimension of the distribution is calculated to be $\frac{3}{2}$. The larger peaks at small denominator rationals can be explained by simple numerical laws, that given a maximum size of integer, if all the possible combination of ratios are reduced to their smallest form, we find $\frac{1}{2}$ to be formed the maximum time, and similar argument goes for other small denominator fractions too. Fig 1 shows the distribution

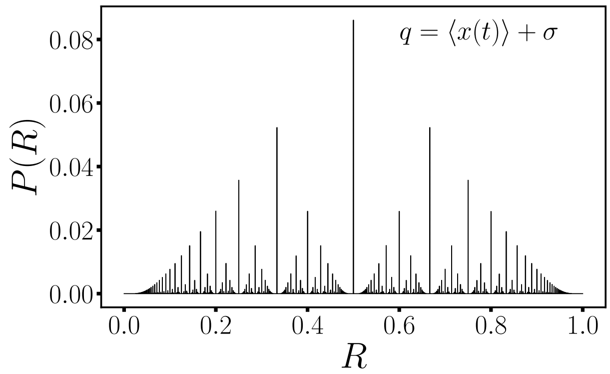

FIG. 1. The distribution of R, for uncorrelated time series, where extreme events are defined to exceed the threshold $q = \langle x(t) \rangle + \sigma$. The distribution is symmetric about $R = \frac{1}{2}$ and fractal in nature.

of R for uncorrelated time series when the threshold for extreme events is set a 1.

A generalized thomae function:

IV. DISCUSSIONS

- Proposed definition of ratio provides a novel framework to study the distribution of ratio as desecrate probability mass function which are understudied not just in the context of extreme events.

- The fractal behavior is found to be general which are applicable in the distribution of whole number ratios in various domains.

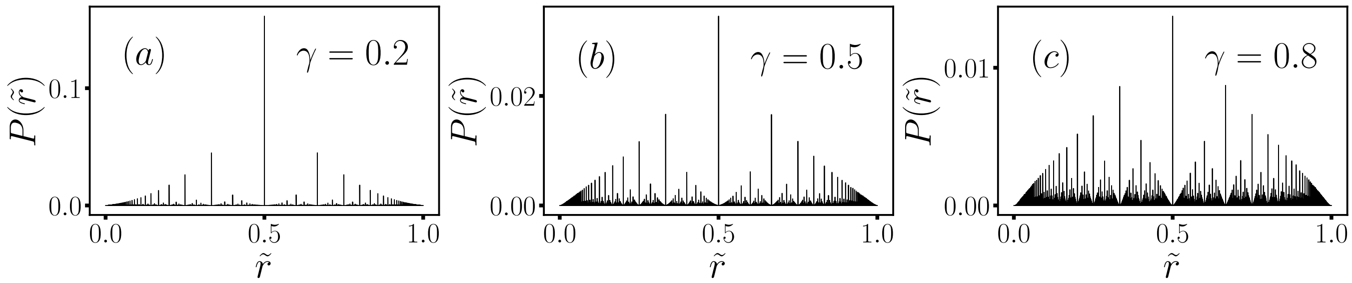

FIG. 2. Distribution of R for time series with different correlation lengths. Panels (a), (b) and (c) correspond to the distribution of of R of the time series with correlation length 1, 2 and 10 respectively for a fixed threshold $q = \langle x(t) \rangle + \sigma$. $P(R = \frac{1}{2})$, is much larger for highly correlated data, i.e. for the time series with large correlation lengths. The distribution seem too have higher box counting dimensionality for less correlated times series.

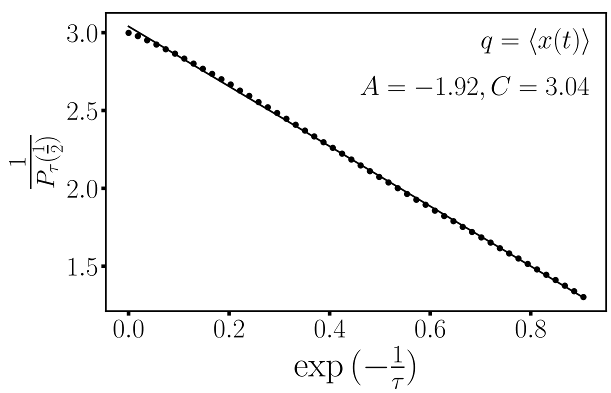

FIG. 3. For small threshold limit (i.e.Threshold for extreme events is $q = \langle x(t) \rangle)$, $P(R = \frac{1}{2})$ follows $\frac{1}{P_\tau(\frac{1}{2})} = A \exp -\frac{1}{\tau} + C$, where A and C are constants.

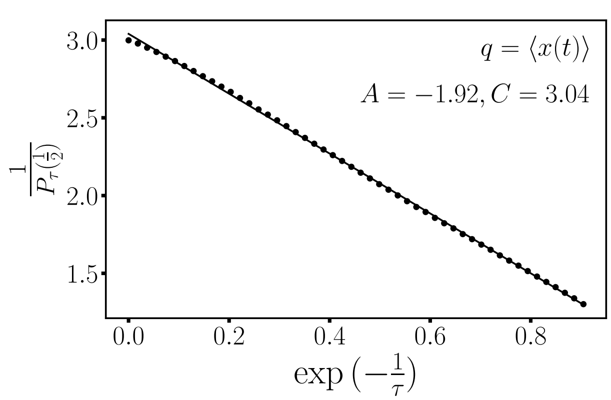

FIG. 4. For small threshold limit (i.e.Threshold for extreme events is $q = \langle x(t) \rangle)$, $P(R = \frac{1}{2})$ follows $\frac{1}{P_\tau(\frac{1}{2})} = A \exp -\frac{1}{\tau} + C$, where A and C are constants.

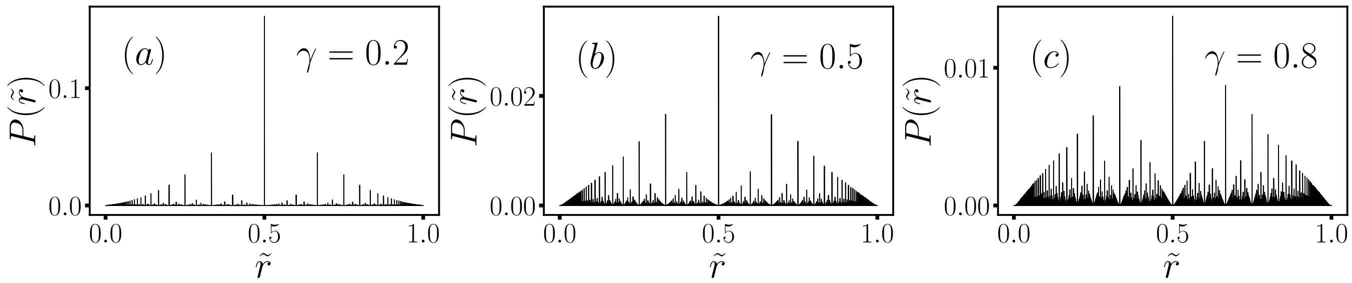

FIG. 5. The box counting dimension which is marker of fractal behavior increases as correlation of the time series decreases.

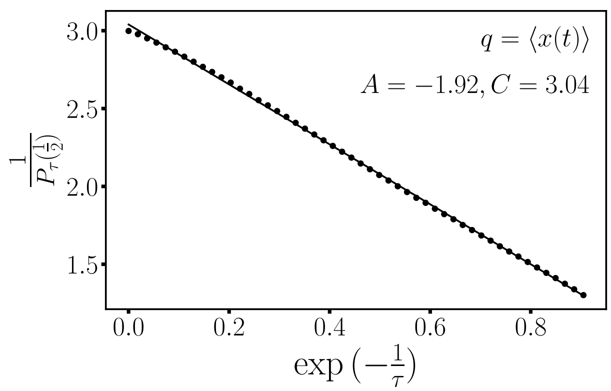

FIG. 6. For time series with a fixed correlation length $\tau = 3$, $P(R = \frac{1}{2})$, decrease as threshold is increased.

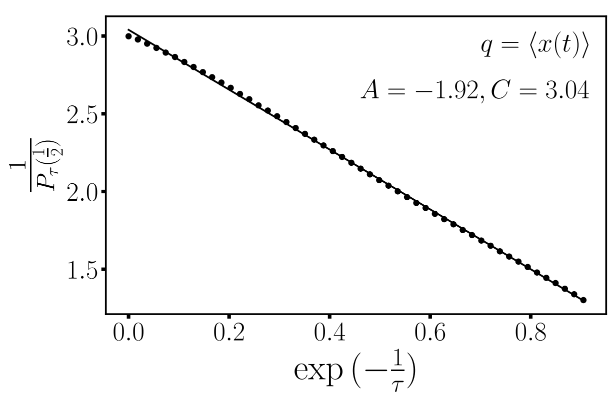

FIG. 7. For time series with a fixed correlation length $\tau = 3$, the box counting dimension increases as the threshold is increased.