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

When or how often an earthquake of magnitude 8 Richter hits Japan? How often a place is hit with extreme cold waves or heat waves? When will the stock market crash next? These are a few of the many questions of interest in the context of extreme events because they are often devastating. It is in our best interest to predict those events, which can help us prepare for them. Or come up with a preventive measure. Though extreme events often have a negative connotation, there are scenarios where extreme events benefit the system (e.g.). Still, it is essential to gain insights into the distribution of extremes. Which led to a branch of statistics called extreme value statistics. And it has been shown that the maxima of iid random variables only falls into either of three distribution, Weibull, Fréchet, and Gumbel, and the distribution depends on the limiting behavior of the tail of the original distribution from where the iid's were drawn.

Though working with iid's provides essential insights and a null model, real-time series are highly complicated. Often they have a non-vanishing autocorrelation function at non-zero lag. which makes them extremely difficult to tackle analytically. There have been studies to discover a different facet of extreme events where long-range memory persists, But those analyses for weakly correlated time series are still limited. A weakly correlated time series has an autocorrelation function that decays exponentially and can model a time series for temperature data, whose autocorrelation functions are found to follow an exponential for the first few lags. As it is convenient to generate weekly correlated time series, it provides an accessible playground to study different facets of extreme events.

A few technical questions that help us to understand the question stated earlier are: What are the time gaps between two extreme events (recurrence time.) And what does this distribution look like? These questions have been asked before for time series with a long memory. We can ask another question: what are the ratios of these recurrence time and their distributions? We take inspiration from random matrix theory, which deals with spacing between eigenvalues and spacing ratios. spacing ratios are shown to provide a marker to quantum chaos, being more accessible to handle than spacing, which requires a nontrivial task of unfolding. Taking inspiration from that, we formulate a different definition of ratios in the context of extreme events. And we find it extremely important to mention that the ratios in this context are rational as they are ratios of whole numbers. Hence we study the discrete probability mass function of the ratios. Which provides us with a unique framework for studying extreme events.

Though previous attempts to find the spacing distribution manages to provide us with the behavior of the distributions at large space limit.which corresponds to sparsely distributed extreme events. Our framework helps us to study extreme events when their gaps are very close. which we find to be a characteristic of correlation. We extensively study the distribution at ratio = 1 /2, which corresponds to three consecutive extreme events with similar gaps; often, the gap is just 1. We also study the fractality of the ratio distribution, which is novel in this context to infer correlation from the distribution.

II. MODEL AND METHODS

Consider a process $x(t)$, which can denote the position of a particle under Brownian motion or the 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 autocorrelation 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 nonvanishing autocorrelation functions. In this paper, we primarily focus on time series with exponential autocorrelation function (weakly correlated) of the form $e^{-\frac{t_2-t_1}{\tau}}$. We adopt a well-known method to generate exponentially correlated time series []. 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 away the extremes are from the typical value. We first define recurrence time to formally study how extreme events are distributed over 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 the 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 on 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

First, we study the distribution of R for uncorrelated time series. As ratios are rations, which means they are countable, make $P(R)$ 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 the integer if all the possible combinations of ratios are reduced to their smallest form, we find $\frac{1}{2}$ to be formed the maximum time, and the similar argument goes for other small denominator fractions too. Fig 1 shows the distribution

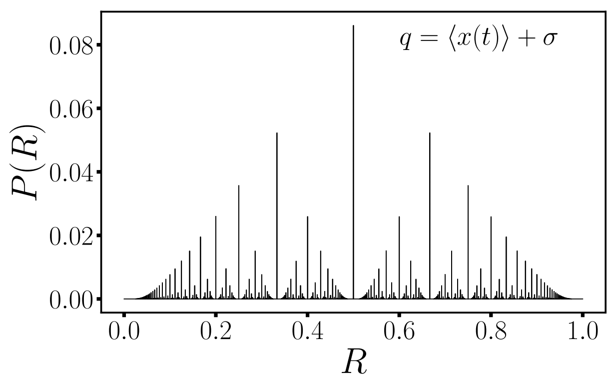

FIG. 1. The distribution of R, for uncorrelated time series, where extreme events are defined to exceed the threshold $q = \langle x(t) \rangle + \sigma$. The distribution is symmetric about $R = \frac{1}{2}$ and fractal in nature.

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

Next we systematically study the distribution of R as we vary the correlation length (τ) of the time series. We find that though the distribution remains symmetric, correlation changes the distribution in such a way that, at a rational point $R = \frac{p}{q}$ $P(R)$ is not just a function of q, as it was the case for uncorrelated time series. Still, the distribution depends on both p and q. We also find the peak at $\frac{1}{2}$ to be large compared to the other peaks when the correlation length (τ), which can be explained as: when correlation length is large the time series is highly correlated, hence the time series tend for follow a trend. That means if an event is extreme it is very likely the the next few events will also be extreme. Now three consecutive extreme events, leads to $R = \frac{1}{1+1} = \frac{1}{2}$. This is why we see a extremely high probability of R being $\frac{1}{2}$ in times series with large correlation length. Also as the correlation length decreases the time series is not dominated by consecutive extreme events, which leads ratios with larger denominator which in turn make the distribution to have much more non zero values at large denominator fractions, hence the fractal dimension of the distribution increases as the time series becomes less correlated.

We also find the effect of increasing threshold q on the $P(R)$. As increasing threshold increases the average spacing more rationals with larger denominators are now possible, which reduces the probability of the ratios with small denominators and also increases the box-counting dimension of the distributions.

We also find $P(R = \frac{1}{2})$ to decrease as the correlation length decreases, and for small fixed threshold $q = \langle x(t) \rangle$, it follows $\frac{1}{P_\tau(\frac{1}{2})} = A \exp -\frac{1}{\tau} + C$, where A and C are constants. For fixed correlation length $\tau = 3$, the peak at $\frac{1}{2}$ also follow this () relation.

Our framework also allows us to connect the correlation in a time series to the box-counting dimension of the distribution, which we find to increase as correlation length increases and as the threshold increases. We find the box-countin dimension to follow these () relations.

Uncorrelated to correlated: Generalized Thomae function to a modified Thomae function. In generalized Thomae function the value of the $P(R)$ at any rational point $\frac{p}{q}$ only depends q, For example, The values of $P(R)$ at $\frac{1}{5}, \frac{2}{5}, \frac{3}{5}$ and $\frac{4}{5}$ are the same of uncorrelated time series. But as correlation is introduced in the time series they no longer remain the same. We show the relative height of $P(R)$ for R's with a fixed denominator has a connection to the correlation length of the system. which we find to to be this()/

We demonstrate the results of our findings in a real-world temperature data set of India. We find the temperature time series to have an exponential auto-correlation function with a correlation length of $5-7$. Fig x shows an excellent match to the temperature data's $P(R)$ with the simulated data of correlation length 6. The match holds for larger thresholds, too, which can be seen in panel (b).

A generalized thomae function:

IV. DISCUSSIONS

To summarize we have used a novel framework in the context of extreme events which is inspired by ratios of spacing in the random matrix theory. Though in the context of RMT the spacing ratios can take up any real values, in the case of time series, the ratios can only take rational values and by construction ratios only take values from 0 to 1. This framework allows us to study consecutive extreme events in great detail, which led to the finding that extreme events apart from small time gaps are markers of high correlation. From the study of the box-counting dimension of the ratio distribution, we find a different approach to related correlation with the fractal dimension of the distribution. Earlier the approach

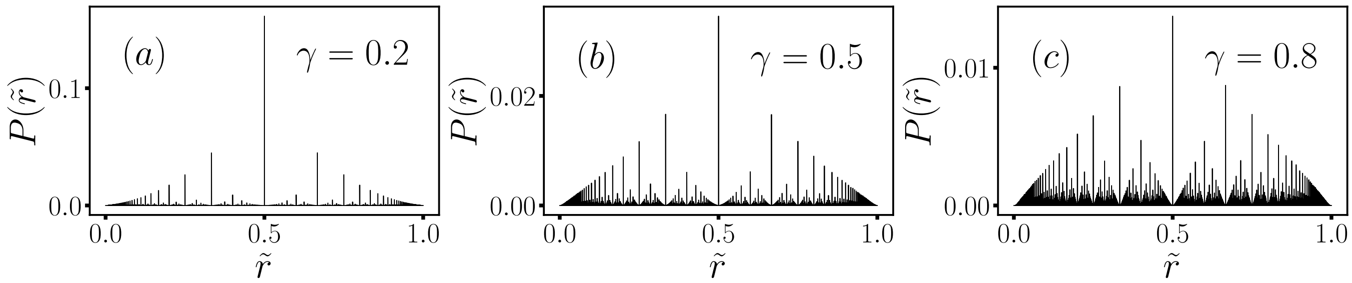

FIG. 2. Distribution of R for time series with different correlation lengths. Panels (a), (b) and (c) correspond to the distribution of of R of the time series with correlation length 1, 2 and 10 respectively for a fixed threshold $q = \langle x(t) \rangle + \sigma$. $P(R = \frac{1}{2})$, is much larger 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.

relate to fractality and correlation was to study the fractal dimension of the time series itself. We also introduce the thomae function in the context of ratios studied the effect of correlation on it. and find that a large correlation changes the height of the peaks at same denominator fractions, which are same for uncorrelated time series. and Thomae function. This also provides us yet another approach to studying extreme events in correlated time series. Together with all this we demonstrate the applicability of our approach in a real-world data set.

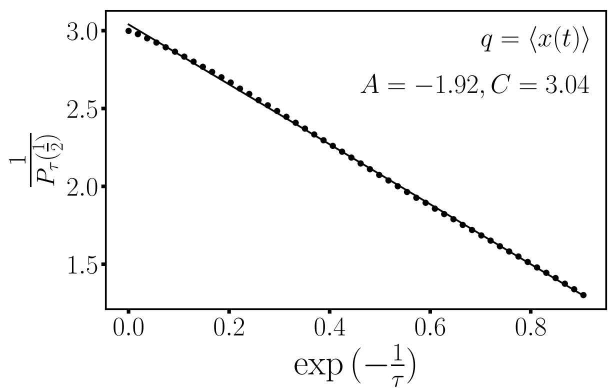

FIG. 3. For small threshold limit (i.e., Threshold for extreme events is $q = \langle x(t) \rangle \rangle$, $P(R = \frac{1}{2})$ follows $\frac{1}{P_\tau(\frac{1}{2})} = A \exp -\frac{1}{\tau} + C$, where A and C are constants.

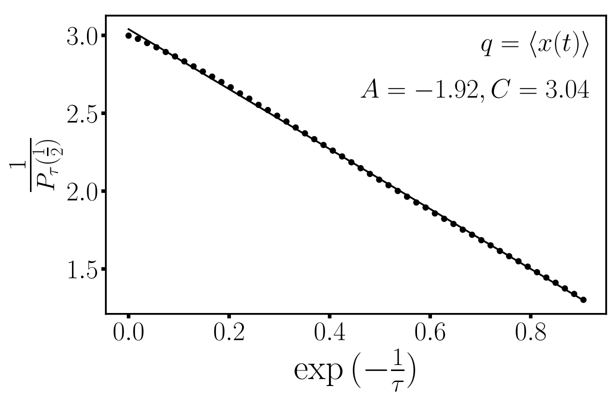

FIG. 4. For small threshold limit (i.e.Threshold for extreme events is $q = \langle x(t) \rangle \rangle$, $P(R = \frac{1}{2})$ follows $\frac{1}{P_\tau(\frac{1}{2})} = A \exp -\frac{1}{\tau} + C$, where A and C are constants.

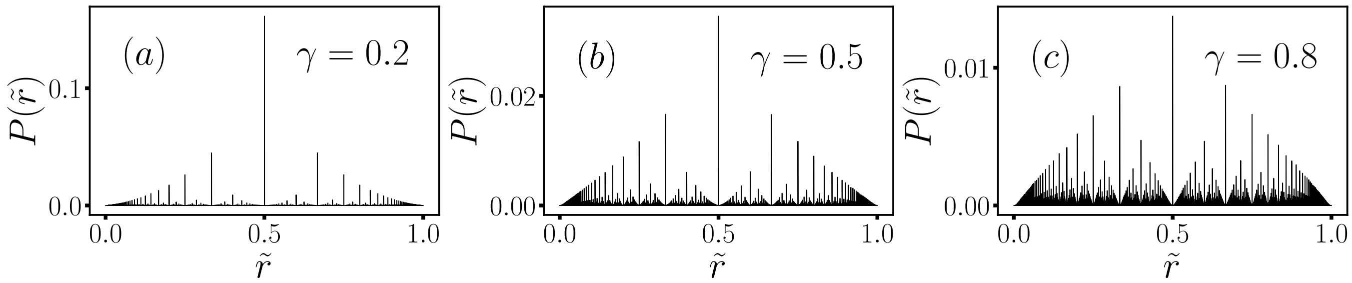

FIG. 5. The box-counting dimension which is marker of fractal behavior increases as correlation of the time series decreases.

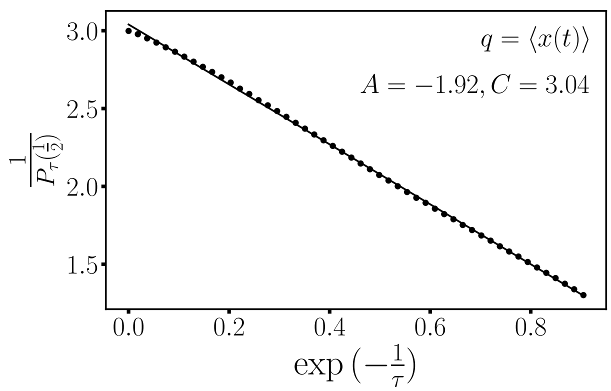

FIG. 6. For time series with a fixed correlation length $\tau = 3$, $P(R = \frac{1}{2})$, decrease as threshold is increased.

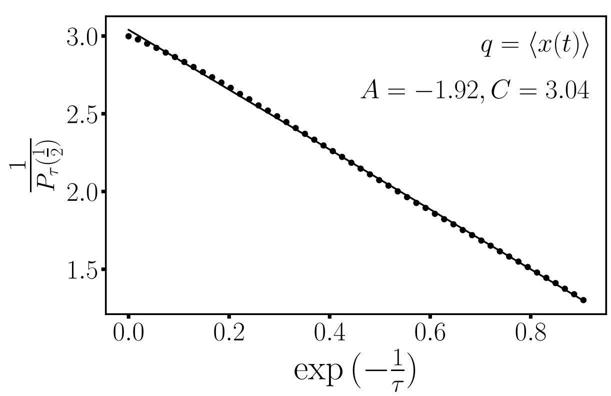

FIG. 7. For time series with a fixed correlation length $\tau = 3$, the box counting dimension increases as the threshold is increased.