# 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.*

Extreme events such as earthquakes, heatwaves, and stock market crashes can have devastating consequences, making it crucial to predict and prepare for them. By examining the distribution of ratios of recurrence times for extreme events, we develop a framework for studying extreme events in weakly correlated time series. We find the distribution of the ratio closely follows the Thomae function, in uncorrelated time series, But the distribution modifies significantly when correlation is introduced the in the series. Our findings provide insight into the behavior of extreme events in close proximity and the relationship between correlation and the distribution of extreme events.

## I. INTRODUCTION

Questions regarding the frequency of earthquakes, extreme weather events, and stock market crashes are of interest in the study of extreme events. Predicting such events can help us prepare for or prevent them. Extreme value statistics is a branch of statistics that focuses on the distribution of extremes. It has been shown that the maxima of independent and identically distributed random variables can only fall into one of three distributions: Weibull, Fréchet, and Gumbel. The specific distribution depends on the behavior of the tail of the original distribution from which the iid random variables were drawn.

Although using independent and identically distributed random variables provides valuable insights and a useful null model, real-time series are often more complex. In particular, they often have non-vanishing autocorrelation functions at non-zero lags, which makes them difficult to analyze. Previous studies have attempted to explore the different facets of extreme events in the presence of long-range memory. But these analyses for weakly correlated time series are limited. A weakly correlated time series is one whose autocorrelation function decays exponentially, and such a model can be used to represent temperature data, for which the autocorrelation function is found to follow an exponential decay for the first few lags. It is convenient to generate weekly correlated time series, which provides an accessible setting for studying the various aspects of extreme events.

To better understand and predict the occurrence of such events, it is important to study the patterns and regularities in their occurrence. One way to do this is to examine the time gaps between extreme events, which can provide insight into their recurrence patterns and help to identify potential triggers or precursors.

Previous research on this topic has focused on the distribution of time gaps in time series with long memory, but here we propose a new approach inspired by random matrix theory. Specifically, we take inspiration from the use of spacing ratios in quantum chaos, which have been shown to provide a useful marker for identifying chaotic behavior. In the context of extreme events, we define ratios in a similar way and study their discrete probability mass function. This allows us to develop a unique framework for studying extreme events and can help to shed light on the underlying mechanisms that govern their occurrence.

By examining the time gaps between extreme events and the distribution of these gaps, we can gain valuable insights into their occurrence patterns. This can help us to understand the factors that contribute to the stability of complex systems and identify potential triggers or precursors for extreme events. Overall, this work provides a new approach for studying extreme events and can help to advance our understanding of their underlying mechanisms.

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

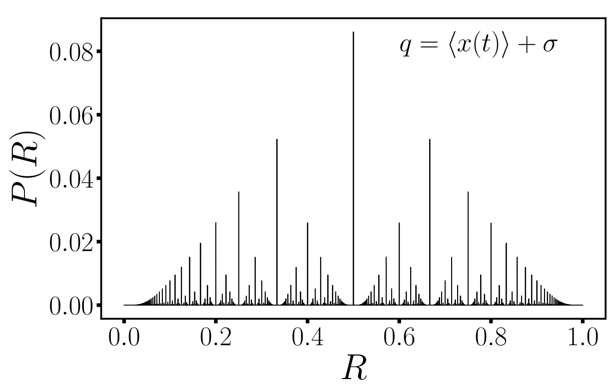


FIG. 1. The distribution of $R$, for uncorrelated time series, where extreme events are defined to exceed the threshold $q = \langle x(t) \rangle + \sigma$. The distribution is symmetric about $R = \frac{1}{2}$ and fractal in nature.

Next we systematically study the distribution of $R$ as we vary the correlation length ($\tau$) 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 ($\tau$), 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-counting 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:

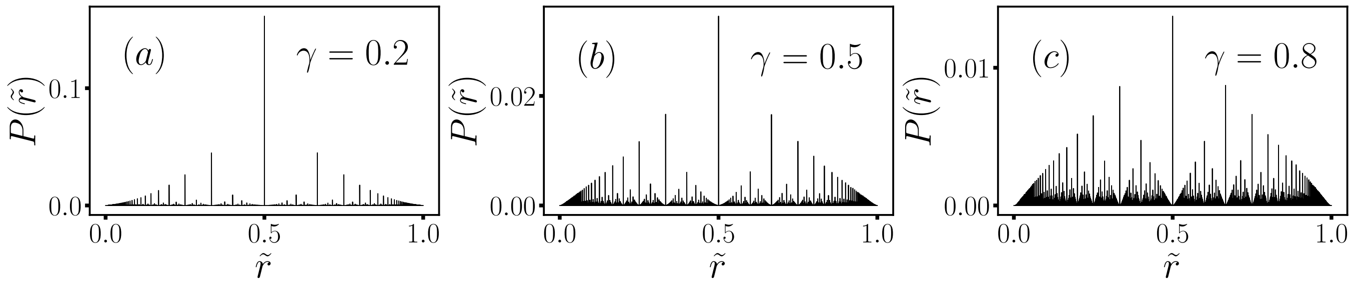


FIG. 2. Distribution of $R$ for time series with different correlation lengths. Panels (a), (b) and (c) correspond to the distribution of of $R$ of the time series with correlation length 1, 2 and 10 respectively for a fixed threshold $q = \langle x(t) \rangle + \sigma$. $P(R = \frac{1}{2})$, is much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

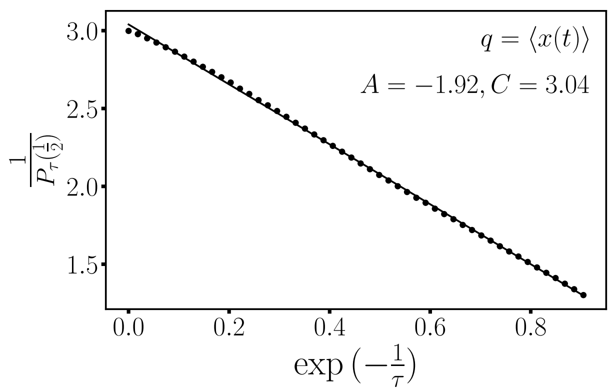


FIG. 3. For small threshold limit (i.e., Threshold for extreme events is $q = \langle x(t) \rangle$), $P(R = \frac{1}{2})$ follows $\frac{1}{P_\tau(\frac{1}{2})} = A \exp -\frac{1}{\tau} + C$, where $A$ and $C$ are constants.

## 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 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 the same denominator fractions, which are same for uncorrelated time series. and Thomae function. This also provides us with 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.

Our approach to studying ratios in the context of extremes opens up a new avenue to study ratios and rational numbers that appear in various contexts, it has been shown that these kinds of fractal distributions appear in biological systems as well in the election as vote share distributions.