

Large deviation probabilities for correlated Gaussian processes and intermittent dynamical systems

Mozhdeh Massah,¹ Matthew Nicol,² and Holger Kantz¹

¹Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, D 01187 Dresden, Germany

²Department of Mathematics, University of Houston, TX Houston 77204-3008, Texas, USA

(Dated: March 13, 2018)

In its classical version, the theory of large deviations makes quantitative statements about the probability of outliers when estimating time averages, if time series data are identically independently distributed. We study large deviation probabilities (LDP) for time averages in short and long range correlated Gaussian processes and show that long range correlations lead to sub-exponential decay of LDP. A particular deterministic intermittent map can, depending on a control parameter, also generate long range correlated time series. We illustrate numerically, in agreement with the mathematical literature, that this type of intermittency leads to a power law decay of LDP. The power law decay holds irrespective of whether the correlation time is finite or infinite, and hence irrespective of whether the central limit theorem applies or not.

PACS numbers: 05.45.Ac, 05.10.Gg, 05.20.-y

Keywords: large deviations, time averages

I. INTRODUCTION

A commonly used tool in data analysis is to compute a sample mean. Assuming a unimodal distribution, its mean provides valuable information about which value is typically found in an observation. Also, it is one of the simplest and therefore very robust statistics to compute and suffers much less from sampling effects of tails of the distribution than estimates of higher moments.

In the context of a time series, the sample mean is a time average. Due to correlations among successive data points, the information stored in a time series might be much less than the information stored in a sample of independently drawn data points of equal size. Hence, the issue of how close the sample estimate of a time average is to the true mean value of the process depends on correlations in data. In this paper, we will study the probability that a single time average deviates by more than some threshold value from the true process mean. This will be called the Large Deviation Probability (LDP), and it will be a function of the time interval over which the average is taken: The longer the time interval, the smaller will this probability be. However, it is the precise functional form of this decay which will be in the focus of this article.

Before we calculate this time-dependence of the large deviation probability for short range and long range correlated processes in Secs. 2 and 3 and for a deterministic intermittent map in Sec. 4, let us recall some relevant results about time averages of random variables.

Given is a time series $\{x_1, \dots, x_i, \dots, x_N\}$ of real numbers, where N is called the sample size. In applications, these data are results of some measurements, and the index i refers to time. We assume that the time intervals between each two successive measurements are equal, which is the case for most processes which are sampled at a constant rate.

We interpret the time series to be a given realization

of a stationary stochastic process. This includes deterministic settings and simply means that we assume the existence of well defined joint probability densities $p(x_i, x_{i+1}, \dots, x_{i+m})$ for the joint occurrence of m specific values, for all orders m . Due to the stationarity assumption, such probability densities are independent of i . This implies that the underlying process is operating at constant-in-time control parameters and that there are no transients. Stationarity is often violated in applications but nonetheless assumed for many data analysis tasks. In such a situation the mean value of the process is the mean of the marginal probability density $p_1(x)$, which we also call the ensemble mean μ .

Time averages are then defined as

$$S_N = \frac{1}{N} \sum_{i=1}^N x_i . \quad (1)$$

This is called the Birkhoff sum in the mathematical literature. The Birkhoff ergodic theorem [1] states that if a stationary process is ergodic then the time average S_N converges to the ensemble average $\mu = \langle x \rangle$ for $N \rightarrow \infty$. The Birkhoff ergodic theorem is also known as the *Law of Large Numbers* for identically independently distributed random numbers, so-called iid processes.

Since the number of observations N in all practical applications is finite, the crucial issue is how far a single time average for finite N can deviate from the limit μ . There are two statements which give partial answers in some situations: The *central limit theorem* (CLT) (see for example [2, Chapter 2]) tells us that the distribution of $\sqrt{N}S_N$ converges to a Gaussian with fixed variance for large N under certain assumptions, for example it suffices that the observations x_i are stationary, if their probability density $p_1(x)$ has a finite second moment, and if any dependencies between x_i and x_j decay exponentially fast in $|i - j|$. Therefore, all iid-processes with finite variance fulfil the central limit theorem, but processes with

long range correlations might violate it. Indeed, in Sec.4 we will discuss a deterministic map, which under iteration generates time series with long range temporal correlations and where under certain conditions the CLT is violated.

A second relevant statement is made by the theory of large deviations[2, 3]. It makes a quantitative statement about the probability that a single S_N obtained from a time series deviates from the ensemble average μ by more than some tolerance ϵ , in the asymptotic regime of large N . More specifically, for an *iid* process such that $E[e^{\theta x_i}] < \infty$ for some $\theta > 0$ and satisfying other mild assumptions, this large deviation probability (LDP) reads:

$$P(|S_N - \mu| > \epsilon) \propto \exp(-I(\epsilon)N), \quad (2)$$

where the *rate function* $I(\epsilon)$ satisfies $I(\epsilon) > 0$ for $\epsilon \neq 0$. For classical distributions the rate function can be determined from the probability density $p_1(x)$ as the Legendre transform of the moment generating function of $p_1(x)$. E.g., for $p_1(x)$ being Gaussian with unit variance, $I(\epsilon) = \frac{\epsilon^2}{2}$. More relevant than the functional form of $I(\epsilon)$ is the universal exponential decay of this probability with N . It guarantees that large deviations of S_N from μ are exponentially suppressed in growing sample size N . Notice that the usual assumption is that the sample is an independent sample of identically distributed random variables, so that it is not applicable to time series with correlated elements.

In this article, we study this N dependence, and hence the speed of convergence of S_N to μ , for two classes of correlated Gaussian processes and for a family of one-dimensional deterministic maps. Whereas short range correlations will lead to a simple correction of sample size N , long range correlations with diverging auto-correlation time lead to a sub-exponential decay of LDPs. We then go beyond Gaussian processes and study a deterministic chaotic map, the Pomeau-Manneville map[4], where in a certain parameter regime the Central Limit Theorem breaks down and then also the LDP behavior differs from that of a long range correlated Gaussian stochastic process.

II. LDP FOR AR(1) PROCESSES

The *auto-regressive process of first order*, AR(1), is defined as a stationary Gaussian stochastic process with the following iteration rule:

$$x_{n+1} = ax_n + \xi_n \quad |a| < 1 \quad (3)$$

with Gaussian white noises ξ_n , $\langle \xi_i \xi_j \rangle = \delta_{i,j}$. By either taking the square of this equation or by multiplying by x_n and performing ensemble averages under the assumption of stationarity, one obtains its variance $\langle x^2 \rangle = \sigma_{AR(1)}^2 = \frac{1}{1-a^2}$ and its auto-correlation function is

$\langle x_n x_{n+k} \rangle = \frac{1}{1-a^2} a^{|k|}$. Hence, AR(1) processes are stable if $|a| < 1$, and their auto-correlation time is $-1/\ln|a|$.

We obtain the LDP from a calculation of the probability density function for S_N . Since S_N , Eq.(1), in this case is a linear superposition of Gaussian random variables x_i , its distribution is a Gaussian itself, regardless of correlations, i.e., without employing the central limit theorem and without employing the limit $N \rightarrow \infty$. The mean of this Gaussian distribution is the ensemble mean of S_N which is evidently 0, and the only parameter to calculate is its variance $\sigma_S^2(N)$ where we are particularly interested in the N -dependence.

A straightforward calculation of $\sigma_S^2(N) := \langle (\frac{1}{N} \sum_{k=1}^N x_k)^2 \rangle$ making use of the recursion relation Eq.(3) yields

$$\sigma_S^2(N) = \frac{1}{N^2} \left\langle \left(\sum_{k=0}^{N-1} a^k x_1 + \sum_{k=1}^{N-1} \xi_k \sum_{j=0}^{N-k-1} a^j \right)^2 \right\rangle \quad (4)$$

$$= \frac{1}{N^2} \left\langle \left(x_1 \frac{1-a^N}{1-a} + \sum_{k=1}^{N-1} \xi_k \frac{1-a^{N-k}}{1-a} \right)^2 \right\rangle \quad (5)$$

$$= \frac{1}{N^2} \left(\langle x_1^2 \rangle \left(\frac{1-a^N}{1-a} \right)^2 + \sigma_{\text{noise}}^2 \sum_{k=1}^{N-1} \left(\frac{1-a^k}{1-a} \right)^2 \right) \quad (6)$$

due to the pairwise independence of ξ_i, ξ_j for $i \neq j$ and the independence of ξ_i and x_1 . The mean of x_1^2 is the process variance σ_{AR}^2 . The variance of the noises σ_{noise}^2 is unity, so that the final expression reads:

$$\sigma_S^2(N) = \frac{1}{N^2} \frac{1}{(1-a)^2} \left(\frac{(1-a^N)^2}{1-a^2} + \sum_{k=1}^{N-1} (1-a^k)^2 \right) \quad (7)$$

For $a = 0$ this expression reduces to $\sigma_S^2(N, a = 0) = 1/N$, which is to be expected for the sum of white noises.

For $a \neq 0$, exact numerical evaluations of Eq.(7) as functions of N for various a are shown in Fig.1, together with the $1/N$ behaviour of the white noise case. For large N , the variance decays like $1/N$ for all a , whereas for small N and $a \neq 0$ there is some a -dependent cross-over behaviour. For N large, the first summand in Eq.(7) stemming from the initial condition x_1 becomes negligible due to its rescaling with $1/N^2$. Since the individual terms $1-a^k$ in the second summand tend to unity for large k , this second sum grows like N , which, together with the $1/N^2$ prefactor, leads to the overall $1/N$ behaviour for large N .

Since the variance of the AR(1) random variables x_i depends on a , it is reasonable to rescale the variance $\sigma_S^2(N, a)$ accordingly. After normalizing Eq.(7) by σ_{AR}^2 the dominant second term has a prefactor $(1+a)/(1-a)/N^2$. Recalling that $1/N \sum_{k=1}^N (1-a^k)^2 \rightarrow 1$ for large N , we arrive at the universal result

$$\frac{\sigma_S^2(N, a)}{\sigma_{AR(1)}^2(a)} \rightarrow \frac{1+a}{1-a} \frac{1}{N} = \frac{1}{N_{\text{eff}}} \quad (8)$$

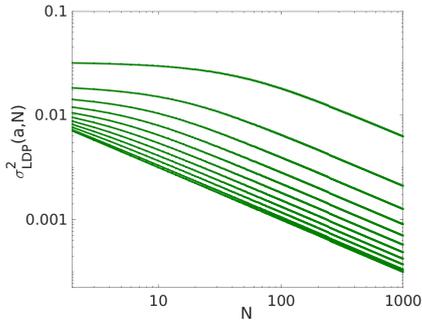


FIG. 1: Numerical evaluations of Eq.(7), the variance of the Large Deviation Probability distribution for $a = 0.993, 0.893, \dots, 0.193$ and for $a = 0$ from right top to bottom. The $1/N$ -behaviour for large N is evident, but for a close to 1, there is a complicated cross-over behaviour.

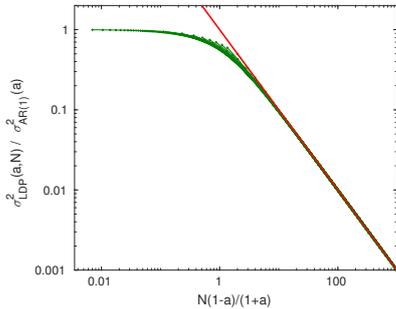


FIG. 2: The curves of Fig. 1 after rescaling the variance with the process variance and after rescaling of the sample size N by $(1-a)/(1+a)$ as mentioned in the text. The red line represents $1/N_{\text{eff}} = (1+a)/(1-a)/N$.

where we interpret $N(1-a)/(1+a) < N$ as an effective data set size for these short-range correlated AR(1) data: Due to lack of independence for $a \neq 0$, the Birkhoff sum S_N behaves as if the true time interval were much smaller. It is, however, remarkable that this rescaling factor for $|a| \approx 1$ is equivalent to (twice of) the correlation time $\tau = -1/\ln a$ as can be seen by an easy calculation. We call $1-a = \delta$ and make an expansion around $\delta = 0$ for the correlation time and for the scaling factor and find an identity in lowest order:

$$1/\tau = -\ln(1-\delta) \approx \delta + \delta^2/2$$

$$\frac{\delta}{2-\delta} \approx \frac{1}{2}\delta(1+\delta/2)$$

So in summary, we know that the probability density $\rho_{N,a}(s)$ for the Birkhoff sum S_N is a Gaussian distribution with zero mean and with a variance given by Eq.(7) and Eq.(8) for large N_{eff} . Fig. 2 shows that after these two rescalings, there is perfect data collapse.

We can now express the LDP through the complemen-

tary error function $\text{erfc}(\epsilon) := \frac{2}{\sqrt{\pi}} \int_{\epsilon}^{\infty} e^{-t^2} dt$ as

$$\text{Prob}(S_N > \epsilon) = \int_{\epsilon}^{\infty} \rho_{N,a}(s) ds = \frac{1}{2} \text{erfc}(\epsilon/\sqrt{2}\sigma_S(N)) \quad (9)$$

The asymptotic expansion of erfc for large arguments yields in lowest order

$$\text{Prob}(S_N > \epsilon) \approx \sqrt{\frac{2}{\pi}} \frac{\sigma_S(N)}{\epsilon} e^{-\epsilon^2/2\sigma_S^2(N)}. \quad (10)$$

Using Eq.(8) and inserting $\sigma_{\text{AR}(1)}^2$, we find asymptotically for large N and ϵ

$$\text{Prob}(S_N > \epsilon) \approx \sqrt{\frac{2}{\pi}} \frac{1}{N(1-a)^2 \epsilon} e^{-N \frac{\epsilon^2}{(1-a)^2}}. \quad (11)$$

Again, if we wish to study the a -dependence of the LDP, it makes sense to rescale the threshold value ϵ by the a -dependent standard deviation of the corresponding AR(1) process. The rescaled threshold then determines a given percentile of the distribution, independent of a . We call the new threshold $\tilde{\epsilon} := \epsilon/\sigma_{\text{AR}} = \epsilon\sqrt{(1-a)^2}$. This introduces in a natural way the ratio $\sigma_S^2(N)/\sigma_{\text{AR}(1)}^2$ and we find

$$\begin{aligned} \text{Prob}(S_N > \tilde{\epsilon}) &= \frac{1}{2} \text{erfc}\left(\frac{\tilde{\epsilon}\sigma_{\text{AR}}}{\sqrt{2}\sigma_S(N)}\right) \\ &= \frac{1}{2} \text{erfc}\left(\frac{\tilde{\epsilon}}{2} \sqrt{\frac{N(1-a)}{1+a}}\right) \\ &\approx \frac{1}{\sqrt{2\pi}} \frac{1}{2\tilde{\epsilon}\sqrt{N_{\text{eff}}}} \exp(-N_{\text{eff}} \frac{\tilde{\epsilon}^2}{2}). \end{aligned} \quad (12)$$

for large N_{eff} .

The analytical calculations for the AR(1) process therefore demonstrate that for large N the exponential decay of the LDP in sample size sets in, but that the effective time window is given by the true time window divided by twice the correlation time of the process. For $|a|$ close to zero, the rescaling factor tends to unity, reflecting the time-discrete nature of the process, i.e., the correct behaviour of white noise is recovered, whereas a rescaling with the true correlation time 0 would not make sense.

When rescaling the threshold value ϵ by the standard deviation of the AR(1) process which is reasonable when comparing processes with different a , the asymptotic exponential decay in N is dressed by a rate function $I(\tilde{\epsilon}) = \tilde{\epsilon}^2/2$ which is identical to the rate function of Gaussian iid random variables with unit variance.

In Fig.3 we show a comparison of numerically obtained LPDs for a range of a values versus the analytical prediction of the asymptotic behaviour for fixed $\tilde{\epsilon} = 0.1$ and as a function of the effective sample size $N_{\text{eff}} = N(1-a)/(1+a)$. We observe indeed an excellent agreement between numerical results and the asymptotic behaviour.

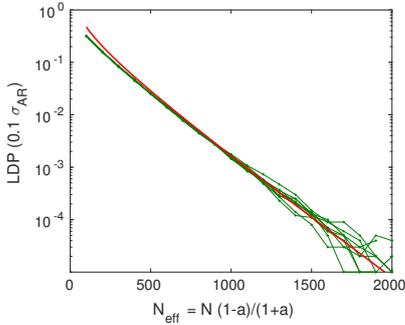


FIG. 3: The Large deviation probability LDP for AR(1) processes for 10 different a -values ranging from $a = 0.05$ to $a = 0.95$. The thresholds were $\epsilon = 0.1 \frac{1}{\sqrt{1-a^2}}$, i.e., they are a fixed fraction of the processes' standard deviation. Plotting such LDPs versus the effective sample size $N(1-a)(1+a)$ leads to a perfect data collapse (apart from statistical fluctuations for small LDP values, ensemble mean over 10^5 realizations each). The red curve represents the asymptotic theory $\text{LDP}(\epsilon) = \frac{1}{\sqrt{2\pi}} \frac{1}{\epsilon \sqrt{N_{\text{eff}}}} \exp(-\epsilon^2 N_{\text{eff}}/2)$.

III. LDP FOR ARFIMA MODEL

ARFIMA models[5] generate time series with long range correlations. Their formal definition involves a backshift operator which can be interpreted as the time-discrete version of a fractional time differentiation. Thereby, the process properties are comparable to fractional Gaussian noise[6]. We use the simplest version, ARFIMA(0, d ,0), which can be mapped to an MA(∞)-model

$$x_n = \sum_{k=1}^{\infty} a_k \xi_{n-k} \quad (13)$$

where ξ_k are iid Gaussian random variables with unit variance and zero mean (i.e., white noise), and the coefficients are given by $a_k = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} = \prod_{0 < j < k+1} \frac{j-1+d}{j}$ [5]. The auto-correlation function $c(\tau)$ of this process exhibits asymptotically a power law decay $\tau^{-\gamma}$ for large time lags τ with $\gamma = 2d - 1$. Hence for $0 < d < 1/2$ this process has positive long range temporal correlations (LRC), since the integral over all time lags of the auto-correlation function, which is interpreted as the correlation time, diverges[5]. Apart from this, it is a Gaussian process, i.e., all joint probability densities are multivariate Gaussians.

The Birkhoff sum S_N of a time series generated by Eq.(13) is therefore a linear combination of Gaussian random variables and hence its probability density is Gaussian itself. Its mean is zero by construction, hence we have to determine its variance as in the last section. We determine its scaling behaviour in N using slightly different arguments than for the AR(1) process. A Brownian path $W(N)$ is obtained by adding

up white noises, $W(N) = \sum_{i=1}^N \xi_i$, $\langle \xi_i \xi_j \rangle = \delta_{ij}$. Its essential feature is that the Mean Squared Displacement $\text{MSD}(N) := \langle (W(N) - W(0))^2 \rangle$, scales linearly in N . If instead we replace the noises ξ_i by the output x_i of the ARFIMA model, the corresponding MSD scales like N^{1+2d} , for $d \in [-1/2, 1/2]$. ARFIMA models create (a time discrete version of) fractional Gaussian noise whose integration leads to fractional Brownian motion. The relationship between $\text{MSD}(N)$ and S_N is evident: $\text{MSD}(N) = N^2 \langle S_N^2 \rangle$. With $\langle S_N \rangle = 0$, the MSD is N^2 times the variance of the probability distribution of S_N . Hence we know that the variance of the distribution of S_N scales like N^{2d-1} with an unknown prefactor which we call α .

Like in Sec.II, the LDP is the integral over the tail of the Gaussian:

$$\text{LDP}(\epsilon, N) \propto \int_{\epsilon}^{\infty} \exp(-\alpha \epsilon'^2 / 2 N^{1-2d}) d\epsilon' \quad (14)$$

and can be expressed through the complementary error function. We approximate this integral by its first order asymptotic expansion and find:

$$\text{LDP}(\epsilon, d, N) \propto \frac{\sqrt{2}}{\sqrt{\pi \alpha \epsilon}} N^{d-1/2} \exp(-\alpha N^{1-2d} \epsilon^2 / 2). \quad (15)$$

For $d = 0$ the ARFIMA-process is identical to white noise, and indeed with Eq.(15) we recover the asymptotically exponential decay in N , in this case we know that $\alpha = 1$. For other values of d , the LDP behaves like a stretched exponential with a power law prefactor.

In Fig.4 we show the comparison between numerically obtained LDPs, of Eq.(14) and of Eq.(15), for $\alpha = 1$.

If we ignore the power law-prefactor, the decay is sub-exponential if the MSD grows faster than linear (superdiffusion) and is super-exponential if the MSD grows slower than linear in time (sub-diffusion), and the relationship is that the stretching exponent $\beta = 1 - 2d$ in the decay of the LDP is related to the exponent $2H$ of the growth of the MSD, N^{2H} , by:

$$\beta = 2 - 2H. \quad (16)$$

The exponent H of the growth of the MSD is often also called the Hurst exponent, and the relationship between H and the power of the decay of the auto-correlation function is known to be $H = 1 + \gamma/2$ where $-1 < \gamma < 0$ (i.e., $d > 0$).

In summary, within this model class of correlated Gaussian processes, a diverging correlation time due to a power law decay of the auto-correlation function induces a sub-exponential decay of the LDP and hence a sub-exponential convergence of time averages towards the ensemble mean μ .

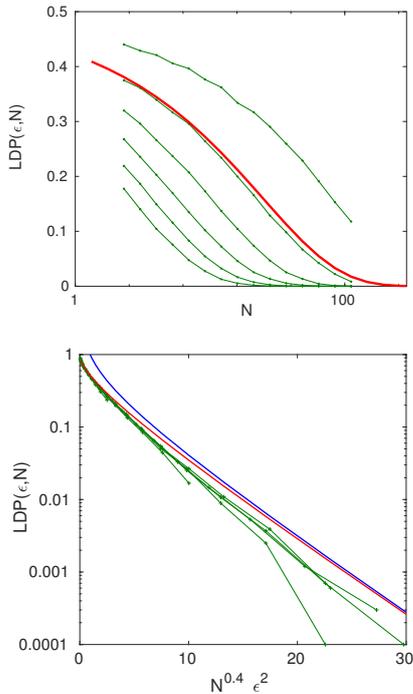


FIG. 4: (color online) Upper panel: The Large Deviation Probability $LDP(N, \epsilon)$ of an ARFIMA(0,d,0) model with $d = 0.3$. The green solid lines show the numerically obtained LDPs for an ensemble of 10000 trajectories and various threshold values ϵ . The red line is the numerical evaluation of the error function Eq.(14) ($\alpha = 1$), which is the (numerically) exact representation of our theoretical prediction. Lower panel: The same LDPs as in the upper panel, but plotted versus $\epsilon^2 N^{1-2d}$, which, according to Eq.(15) should lead to a data collapse in the asymptotic regime. The blue curve is the approximation by the error function Eq.(15), and in this plot we used $\alpha = 1/12$ to ensure the perfect match of numerics and theoretical curves.

IV. LARGE DEVIATIONS IN THE POMEAU-MANNEVILLE MAP

The last model system for our study of LRC on LDP is the deterministic Pomeau-Manneville-map[4]. We study a variant of this map due to Liverani, Saussol and Vaienti [7]. Under iteration, this interval map

$$x_{n+1} = f(x_n) = \begin{cases} x(1 + (2x)^{z-1}) & 0 \leq x < 1/2 \\ 2x - 1 & 1/2 \leq x \leq 1 \end{cases} \quad (17)$$

creates intermittent dynamics, i.e., an alternation of laminar phases where $x_i \approx 0$ for many iterations, and chaotic bursts where $x_i > 1/2$. This is caused by the fact that the fixed point $x = 0$ is marginally stable, i.e. $f'(x=0) = 1$, and trajectories starting from $x = \delta$ with $0 < \delta \ll 1$ take many iterates till they deviate significantly from zero. This is illustrated in Fig.5. It has been shown[8] that this type of intermittency causes a power law decay of the auto-correlation function with the exponent γ .

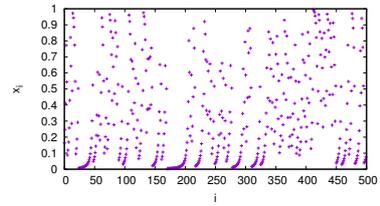


FIG. 5: A typical trajectory of the Pomeau-Manneville map Eq.(17) for $z = 1.6$. The origin for the power law decay of the auto-correlation function lies in the power law distribution of the durations of the laminar phases where $x_i \ll 1$.

$$C(\tau) \propto \tau^{-\gamma}; \gamma = \frac{2-z}{z-1}. \quad (18)$$

Hence, for $1 < z < 3/2$, $\gamma < 1$, the process is not “long range correlated” since the correlation time does not diverge, which is, however, the case for $z > 3/2$. We were able to reproduce the result $\gamma = \frac{2-z}{z-1}$ numerically for the range $z \in [1.12, 1.96]$ with some effort: The power law is asymptotic, hence we need to study time lags where the auto-correlation function is already very close to 0. To suppress fluctuations sufficiently we used quadruple precision in order to avoid trajectory trapping in round-off induced periodic orbits [9].

For $1 < z < 3/2$, the power law decay of correlations is sufficiently fast so that the central limit theorem still holds[10], and therefore, asymptotically for large N , $\sqrt{N}S_N$ is a Gaussian random variable as in the case of the ARFIMA model. Due to the lack of LRC, one might expect a classical exponential decay of the LDP. However, as shown in Fig.6, the convergence of the distribution of $\sqrt{N}S_N$ to a Gaussian is slow. For every finite N , the distribution has an almost Gaussian tail towards large arguments, whereas there is a pronounced non-Gaussian tail towards $S_N = 0$. Large deviations of S_N from μ are dominated by this tail towards values smaller than μ . The anomalous decay of these tails has the consequence that the decay of LDP in N is neither exponential nor stretched exponential, but, as mathematically proven in [11], follows a power law asymptotically:

$$LDP(\epsilon, N, z) \propto N^{\frac{z-2}{z-1}} \text{ if } 1 < z < 2; . \quad (19)$$

Numerically, it is not easy to reach the asymptotic regime, but for z close to $3/2$ we succeed, as shown in Fig.7, after some exponential cross-over behaviour for moderate N .

As we discussed before the distribution for S_N in terms of the MSD, let us stress here that the MSD increases linearly with time if the autocorrelation function is a power law with power larger than one. By expanding the inner sum, the MSD, $\langle (\sum_{i=1}^N x_i - x_0)^2 \rangle$, can be readily expressed as a summation over the auto-correlation function $c(\tau)$ with some τ -dependent prefactor. This expression has two dominant terms: one stemming from the

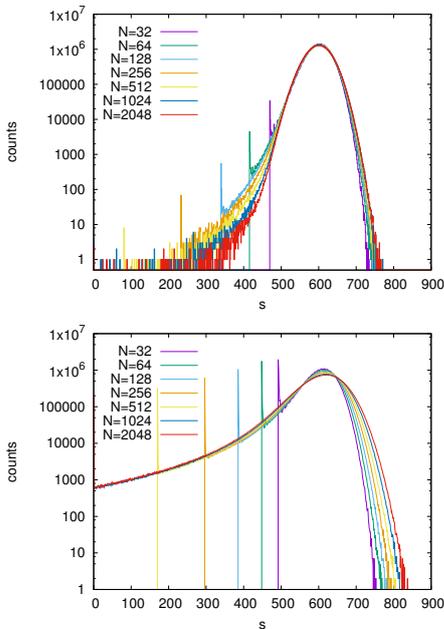


FIG. 6: Histograms of numerically obtained distributions for S_N , for various N , showing the convergence to a Gaussian for $z < 3/2$ (upper panel: $z = 1.24$), and the lack of convergence for $z > 3/2$ (lower panel: $z = 1.54$). Rescaling: $s = (S_N - \mu)\sqrt{N}$, i.e., if x_i were iid Gaussian, then all histograms should be identical parabolae.

range of small $|\tau|$, where the auto-correlation function is finite, the other range for large $|\tau|$ covering the asymptotic power law. In combination with the prefactors, the first one dominates when $\gamma > 1$, and yields a linear increase of MSD in N . For $0 < \gamma < 1$, the second term dominates and results in $\text{MSD} \propto N^{2-\gamma}$. Therefore, there is no contradiction between the validity of the CLT, the linear increase of the MSD, and a power law decay of the LDP: LDP is also sensitive to the tails of the distribution which have an anomalous scaling behaviour in N .

For $3/2 < z < 2$, the CLT is violated, as proven in [12] and as visible in the right panel of Fig.6, and the process is LRC with $0 < \gamma < 1$. When $\sqrt{N}S_N$ is not asymptotically Gaussian random variable, the auto-correlation function, which is a two-point-statistics, is insufficient to fully characterise joint probability distributions. It is therefore plausible that LDP indeed behaves differently from the simple ARFIMA model. Indeed, [13] has studied this case mathematically and shown that the decay of the LDP is given by the same power law as Eq.(19), i.e., in terms of LDP, there is no effect of whether or not the CLT is fulfilled. In our numerical simulations we are able to reach this asymptotic regime and thereby reproduce the theoretical results, see Fig.7.

For $z > 2$, the invariant density cannot be normalized and the system exhibits more complicated features such as weak ergodicity breaking and ageing[14]. In that case, even the ensemble mean μ is ill-defined and hence LDP cannot be studied.

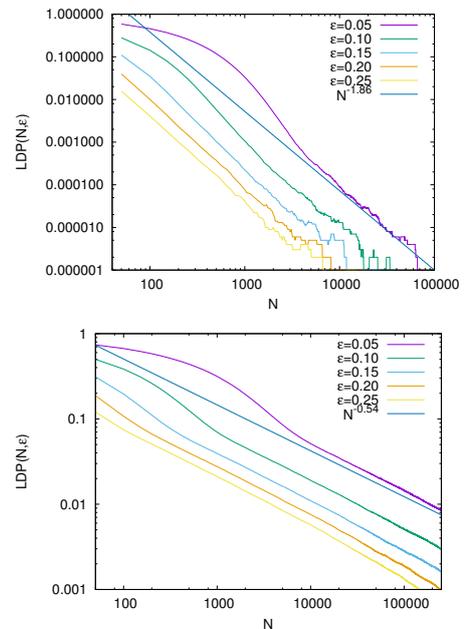


FIG. 7: The decay of the LDP on Pomeau-Manneville trajectories: In both parameter regimes, $z = 1.35 < 3/2$ (upper panel) and $z = 1.65 > 3/2$ (lower panel), the LDP asymptotically decays like a power law in N , after some exponential cross-over. The numerical results are consistent with the theoretical result Eq.(19) which is included in both panels as straight lines.

V. SUMMARY AND CONCLUSION

We have studied the behaviour of the large deviation probability LDP of Birkhoff sums for correlated processes. For two Gaussian processes, AR(1) and ARFIMA(0,d,0), we were able to find exact analytical expressions. In these two model classes, the distribution of the individual random variables is Gaussian, and since these are linear models, also the distribution of the Birkhoff sums S_N is Gaussian. Hence, the variance of this distribution is enough to determine the LDP. For AR(1) we find that $\text{LDP}(N)$ always decays exponentially fast in N , where the time scale is given by an effective sample size related to the decay of correlations of the AR-process. For ARFIMA we obtain stretched exponentials, depending on the value of d , which is an interesting deviation from the standard behaviour.

Mathematicians had already before obtained analytical results for a class of intermittent maps of the Pomeau-Manneville family. In contrast to our Gaussian processes, the decay of LDP in N follows a power law. We were able to interpret these results in view of the distribution of S_N and the decay of correlations, and also to review the validity of the CLT for certain parameter ranges. Asymptotic validity of the CLT does not mean that the Birkhoff sum behaves as a Gaussian random variable for finite N . Nonetheless, it remains a surprise that in terms of LDP, there is no qualitative difference between the

regime $1 < z < 3/2$ and the regime $3/2 < z < 2$, although in the first one the correlation time is finite and the CLT is valid.

What our study also illustrates is the in principle well known fact, which, however, is rarely discussed in the literature about LRC in data: A power-law decay of the auto-correlation function can have different dynamical

origins, so the LRC does not define a single class of processes with unique features. Here we have presented two different such classes, Gaussian stochastic processes and deterministic intermittent maps, whose parameters can be tuned to possess identical power law decays of their auto-correlations, but which behave quite differently in terms of the N -dependence of their LDP.

-
- [1] G. D. Birkhoff. *Proof of the Ergodic Theorem*, PNAS MATHEMATICS, 17 (12) 656-660, (1931)
 - [2] R. Durrett. *Probability: Theory and Examples*, Second Edition, Duxbury Press, (2004)
 - [3] R. S. Ellis. *Entropy, large deviations and statistical mechanics*. Springer-Verlag, New York, (1985)
 - [4] Y. Pomeau and P. Manneville. Intermittent transition to turbulence in dissipative dynamical systems. *Comm. Math. Phys.* **74** 189–197 (1980)
 - [5] Time Series Analysis: Forecasting and Control, Box-Jenkins (4th edition 2008, page 430):Eq.(10.3.6)
 - [6] B. Mandelbrot. Fractional Gaussian Noise and Fractional Brownian Motion and applications, SIAM Rev., 10(4), 422-437
 - [7] C. Liverani, B. Saussol and S. Vaienti, *A probabilistic approach to intermittency*, Ergodic Theory Dynam. Systems **19**(3), 671–685 (1999)
 - [8] L.-S. Young. Recurrence times and rates of mixing. *Israel J. Math.* **110** 153–188 (1999)
 - [9] Hubbard, J. H., and West, B. H. (1991) *Differential Equations: A Dynamical Systems Approach, Part I* (Springer, New York)
 - [10] L.-S. Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math.* **147** 585–650 (1998)
 - [11] I. Melbourne and M. Nicol, *Large deviations for nonuniformly hyperbolic systems*, Transactions of the AMS, **360**, 6661–6676 (2008)
 - [12] S. Gouëzel. *Central limit theorem and stable laws for intermittent maps*, Probability Theory and Related Fields **128**:82-122, (2004)
 - [13] I. Melbourne. *Large and moderate deviations for slowly mixing dynamical systems*. Proc. Amer. Math. Soc. **137** 1735-1741 (2009)
 - [14] Philipp Meyer and Holger Kantz, Infinite invariant densities due to intermittency in a nonlinear oscillator Phys. Rev. E **96**, 022217 (2017)