Convergence of large-deviation estimators

Christian M. Rohwer,[1,2,3] Florian Angeletti,[4] and Hugo Touchette[4,3,*]

[1]*Max-Planck-Institut für Intelligente Systeme, Heisenbergstraße 3, D-70569 Stuttgart, Germany*
[2]*Institut für Theoretische Physik IV, Universität Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart, Germany*
[3]*Department of Physics and Institute of Theoretical Physics, Stellenbosch University, Stellenbosch 7600, South Africa*
[4]*National Institute for Theoretical Physics (NITheP), Stellenbosch 7600, South Africa*

We study the convergence of statistical estimators used in the estimation of large-deviation functions describing the fluctuations of equilibrium, nonequilibrium, and manmade stochastic systems. We give conditions for the convergence of these estimators with sample size, based on the boundedness or unboundedness of the quantity sampled, and discuss how statistical errors should be defined in different parts of the convergence region. Our results shed light on previous reports of "phase transitions" in the statistics of free energy estimators and establish a general framework for reliably estimating large-deviation functions from simulation and experimental data and identifying parameter regions where this estimation converges.

I. INTRODUCTION

The generating function of a fluctuating quantity or random variable X, defined as

$$G(k) = \langle e^{kX} \rangle = \int_{-\infty}^{\infty} p(x)e^{kx}dx, \tag{1}$$

where $p(x)$ is the distribution of X, is related to many important physical quantities. Examples include velocity fields of turbulent fluids, whose generating function, estimated in terms of moments, is related to energy dissipation and multifractal exponents [1–3], the generating function of the energy of systems at thermal equilibrium, which is essentially the partition function [4–7], and the generating function of the work performed on nonequilibrium systems, which is linked via Jarzynski's equality to equilibrium free energy differences [8]. In large-deviation theory, generating functions are also related to so-called rate functions, which give the likelihood of rare fluctuations in stochastic systems. In recent years, this theory has been applied successfully to describe the fluctuations of equilibrium systems [4–7] and nonequilibrium systems driven in steady states [9–12], in addition to manmade systems such as queues and networks [13–17].

The problem that we consider in this paper is the statistical estimation of generating functions from empirical data which arises experimentally or numerically in all the applications mentioned above. To be more precise, we consider a finite sample $\{X^{(j)}\}_{j=1}^{M}$ of M realizations (copies, repetitions or empirical values) of the random variable X and study the convergence of the following statistical estimator of $G(k)$:

$$\hat{G}_M(k) = \frac{1}{M} \sum_{j=1}^{M} e^{kX^{(j)}}. \tag{2}$$

This function converges pointwise to $G(k)$ as the sample size M increases, but a major problem is that it does not converge *uniformly* in k, which means that the sample size M needed to achieve a given accuracy for $\hat{G}_M(k)$ depends on k. In fact, it is

known that, depending on the random variable considered, $\hat{G}_M(k)$ converges slowly for some k and, in some cases, does not converge at all. This problem, often referred to as the "linearization problem," has been studied in the context of multifractal analysis [18–20] and glassy phase transitions [21–23]. Convergence problems have also been studied for the so-called Jarzynski estimator, which is an estimator similar to (2) used to obtain free energy differences from nonequilibrium experiments [24–30]. The focus of these studies, however, is mostly on the statistical bias of $\hat{G}_M(k)$ [31–36], which disappears in the limit $M \to \infty$, rather than the convergence of $\hat{G}_M(k)$ as a function of k and M.

In this paper, we study this convergence for estimating large-deviation functions. Our starting point is a series of studies on data networks [37–40] showing that large-deviation functions can be estimated efficiently from generating functions for random variables having a finite number of values, such as random bits, and for bounded random variables, such as uniform variates. Here we extend these studies by considering *unbounded* random variables, which naturally arise when considering observables of equilibrium and nonequilibrium systems. For these, we show that the estimation of large-deviation functions based on $\hat{G}_M(k)$ converges up to some critical value k_c, which depends on M and the tail of the observable distribution. Moreover, we show that standard statistical errors for this estimator can be defined only up to $k_c/2$. The knowledge of k_c is thus important for properly evaluating, for a given sample size, the parameter range for which large-deviation functions are estimated reliably.

These functions play an important role, as mentioned, for characterizing the typical states and fluctuations of equilibrium, nonequilibrium, and manmade stochastic systems. The numerical computation of these functions for observables of nonequilibrium systems (e.g., particle and energy currents, work, heat, activity, entropy production) has been the subject of active studies in the last years (see, e.g., [41–49]), contributing to our understanding of nonequilibrium phase transitions and fluctuation symmetries. More recent works are now attempting to estimate rate functions in real experiments, for example, in active-matter systems [50,51]. On the experimental side, large-deviation estimations have also been carried out, as

*htouchette@sun.ac.za

mentioned, for data networks and provide in this context a real-time estimate of overflow probabilities in data servers [37–40].

Our results provide for these applications a general method for estimating large-deviation functions, their errors, and their convergence region from finite data samples. They can also be applied for computing multifractal spectra of random fields or time series, dispersion exponents in sheared flows [52], in addition to free energy differences from nonequilibrium work experiments [24–30]. Conceptually, these problems all fall in the topic of large-deviation estimation.

The outline of the paper is as follows. The large-deviation estimators that we consider are defined in Sec. II. Test cases involving bounded and unbounded random variables are considered in Sec. III to show how the estimators' convergence region depends on sample size and how this dependence can be determined, *a priori*, from the knowledge of the distribution considered or, *a posteriori*, from a sample of that distribution. Most of our results are illustrated for simplicity for sums of independent and identically distributed random variables. In Sec. IV we discuss applications for correlated Markov processes and systems composed of many interacting particles, for which the distributions of observables typically scale with time and the number of particles, respectively. Final conclusions are given in Sec. V.

II. METHOD AND RESULTS

A. Estimators

The theory of large deviations is concerned with random variables A_n, depending on some parameter n, whose probability distribution $p_n(a) = P(A_n = a)$ decays approximately exponentially as

$$p_n(a) \approx e^{-nI(a)} \tag{3}$$

when $n \to \infty$, with subexponential corrections in n [7,9,14]. This approximation appears naturally in many equilibrium and nonequilibrium systems, where A_n represent, for example, the energy or magnetization of N particles occupying a volume V, in which case $n = N$ or $n = V$ [5–7], or the current or heat integrated over a time T, so that $n = T$ [10–12]. In manmade systems, A_n can also be the number of "customers" served in a queue [13] or the fraction of time spent by a random walker on specific sites of a network after n time steps [15–17]. In all cases, the distribution of A_n is completely characterized to leading order in n by the function $I(a)$, which gives the likelihood of small and large fluctuations of A_n around its equilibrium or stationary value. This function is called the *rate function* in large-deviation theory [5] and has the interpretation in physics of an entropy function or a generalized potential, depending on the application considered [7].

Many techniques can be used to find $I(a)$. The most common proceeds by calculating the so-called *scaled cumulant generating function* (SCGF), defined as

$$\lambda(k) = \lim_{n \to \infty} \frac{1}{n} \ln \langle e^{nkA_n} \rangle. \tag{4}$$

Provided that this limit function exists for $k \in \mathbb{R}$ and is differentiable, it is known that $p_n(a)$ has the large-deviation form of (3) and that its rate function is given by the Legendre

transform of $\lambda(k)$,

$$I(a) = k_a a - \lambda(k_a), \tag{5}$$

k_a being the root of $\lambda'(k) = a$ [7,9,14].[1] Consider, for example, the case where A_n is a sample mean of n independent and identically distributed (IID) random variables:

$$A_n = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{6}$$

Then the SCGF takes the simple form

$$\lambda(k) = \ln \langle e^{kX} \rangle, \tag{7}$$

so that the large-deviation rate function is obtained as the Legendre transform of the cumulant function of a single random variable, denoted above by X without the subscript because of the IID property. For other observables A_n involving *correlated* random variables, the calculation of $I(a)$ is more involved, but still proceeds from $\lambda(k)$ as defined in (4).

In many applications, the SCGF must be estimated empirically from data samples. For the IID sample mean (6), to take the simplest example, this estimation proceeds by accumulating a sample $\{X^{(j)}\}_{j=1}^{M}$ of M IID copies of the random variable X, distributed according to the (*a priori* unknown) distribution $p(x)$ and by approximating the generating function of X by the estimator $\hat{G}_M(k)$ as defined in (2). The estimator of $\lambda(k)$ is then defined as [54]

$$\hat{\lambda}_M(k) = \ln \hat{G}_M(k). \tag{8}$$

Our goal in this paper is to understand the convergence of this estimator as a function of M and k. From now on, we consider the IID case to simplify the discussion; the case of correlated random variables and observables other than sums is discussed in Sec. IV.

The estimator of rate functions that we consider is defined from the Legendre transform (5) by noting that the estimator (8) of the SCGF is necessarily analytic, since it is a finite sum of exponentials and is thus differentiable for all $M < \infty$. As a result, we consider

$$\hat{I}_M(a) = k_a a - \hat{\lambda}_M(k_a) \tag{9}$$

as an estimator of $I(a)$, where k_a is the computed root of $\hat{\lambda}'_M(k) = a$ [54]. Alternatively, we can proceed parametrically by fixing k and obtain I at the estimated value

$$\hat{a}_M(k) = \lambda'_M(k) = \frac{\sum_{j=1}^{M} X^{(j)} e^{kX^{(j)}}}{\sum_{j=1}^{M} e^{kX^{(j)}}}, \tag{10}$$

using

$$\hat{I}_M(\hat{a}_M) = k\hat{a}_M - \hat{\lambda}_M(k). \tag{11}$$

Strictly speaking, the estimators (9) and (11) are different. We have found in all of our numerical tests, however, that they are nearly identical and differ only because of the discretization used for k. This is a minor, nonstatistical source of errors, which is not discussed further.

---

[1]This holds for convex rate functions. For results on nonconvex rate functions, see Sec. 4.4 of [7], and [53].

As statistical estimators, $\hat{\lambda}_M(k)$ and $\hat{I}_M(a)$ converge pointwise to $\lambda(k)$ and $I(a)$, respectively, in the limit of infinite sample size $M \to \infty$. Their speed of convergence was studied in [40], following previous results on overflow probabilities and bandwidth estimates of data networks [37–40]. These studies, however, consider only bounded random variables for which $\hat{\lambda}_M(k)$ and $\hat{I}_M(a)$ are known to converge quickly and uniformly. In this case, the probability distribution of both estimators has the large-deviation form of (3), which implies that these estimators converge exponentially fast for all k or a with M [40].

We extend these results in what follows by considering *unbounded* random variables. In this case, the convergence of $\hat{G}_M(k)$ is limited by two problems, namely, the linearization effect, which leads to noisy tails of $\hat{G}_M(k)$, and the nonuniform convergence of $\hat{G}_M(k)$ in k, which means that its statistical error is not uniform in k. These problems are explained next and lead us to define, as mentioned, a threshold value k_c depending on M for the convergence of $\hat{G}_M(k)$, $\hat{\lambda}_M(k)$, and $\hat{I}_M(a)$. Applications of these results are presented in the next section.

B. Linearization effect

The linearization effect refers to the fact that sums of exponentials, such as in (2), are dominated as $k \to \infty$ by the largest sample element,

$$X_{\max} = \max_{1 \leqslant j \leqslant M} X^{(j)}, \tag{12}$$

so that

$$\sum_{j=1}^{M} e^{kX^{(j)}} \approx e^{kX_{\max}}, \quad k \to \infty. \tag{13}$$

In this regime, the SCGF estimator thus becomes linear in k:

$$\hat{\lambda}_M(k) \approx k X_{\max}, \quad k \to \infty. \tag{14}$$

Similarly, for $k \to -\infty$, the sum is dominated by the smallest element

$$X_{\min} = \min_{1 \leqslant j \leqslant M} X^{(j)}, \tag{15}$$

so that

$$\hat{\lambda}_M(k) \approx k X_{\min}, \quad k \to -\infty. \tag{16}$$

If the sample $\{X^{(j)}\}$ is obtained from a discrete or continuous distribution with *bounded* support, then the values of $X_{\max}$ and $X_{\min}$ are also bounded and the linearization effect is real: That is, the asymptotic linear branches of $\hat{\lambda}_M(k)$ seen for $|k| \to \infty$ correspond in this case to actual linear branches of $\lambda(k)$ and are unlikely to change much as the sample size M is increased, since the sample will most likely "fill" the range of the bounded distribution for M large enough. However, if the sample is obtained from an *unbounded* distribution, then the linear tails of $\hat{\lambda}_M(k)$ coming from $X_{\max}$ and $X_{\min}$ are an artifact of the finite-size sample: $X_{\max}$ and $X_{\min}$ fluctuate from sample to sample, which implies that $\hat{\lambda}_M(k)$ has fluctuating linear tails for large $|k|$ which are not related to the actual tails of $\lambda(k)$.

This problem affects not only large-deviation computations, as mentioned in the Introduction: Any estimation of

exponential sums, such as those entering in free energy computations and multifractal analysis [18–20], is limited by the linearization effect whenever unbounded random variables are considered. The main problem in these cases is to identify the *onset of linearization*, that is, the threshold value k_c such that, for $|k| < k_c$, $\hat{\lambda}_M(k)$ is not affected artificially by linearization and is therefore a good representation of $\lambda(k)$.

In general, k_c depends on M as well as the particular distribution considered. Moreover, for asymmetric distributions, two threshold values must be considered: k_c^- for the left tail of $p(x)$ determining the distribution of $X_{\min}$, and k_c^+ for the right tail of $p(x)$ determining $X_{\max}$. The convergence and errors of estimators thus depend on whether $k \in [k_c^-, k_c^+]$.

In general, it is difficult to determine k_c^- and k_c^+ exactly; for practical purposes, however, it is sufficient to approximate their growth as $M \to \infty$. This can be done by approximating $G(k)$ in the limit $k \to \infty$ using the saddle-point or Laplace approximation [55] as

$$G(k) \approx e^{kx^*(k) + \ln p[x^*(k)]}, \tag{17}$$

where $x^*(k)$ satisfies

$$kp(x^*) + p'(x^*) = 0. \tag{18}$$

This shows that $G(k)$ is determined for large k by a narrow region of the distribution $p(x)$ located around the *saddle* or *concentration point* $x^*(k)$ [56–58].[2] As a result, $\hat{G}_M(k)$ must be a good estimator of $G(k)$ when the empirical distribution or density histogram $\hat{p}_M(x)$ of the sample $\{X^{(j)}\}_{j=1}^M$ is close to $p(x)$ around $x^*(k)$.

To express this more quantitatively, we define a typicality region for the random variable X by considering the probability,

$$P(X^{(1)}, \dots, X^{(M)} < \bar{x}) = P(X_{\max} < \bar{x}), \tag{19}$$

that all the sample elements $X^{(j)}$ are smaller than a given bound $\bar{x}$.[3] This probability is given in terms of the cumulative distribution

$$F(x) = P(X < x) = \int_{-\infty}^{x} p(z)\,dz \tag{20}$$

by

$$P(X^{(1)}, \dots, X^{(M)} < \bar{x}) = F(\bar{x})^M \tag{21}$$

and can be approximated for $\bar{x}$ and M large by

$$P(X^{(1)}, \dots, X^{(M)} < \bar{x}) \approx 1 - M\overline{F}(\bar{x}), \tag{22}$$

where $\overline{F}(x) = 1 - F(x)$ is the complementary cumulative distribution of X. From this, we see that, if $\bar{x}$ is a constant independent of M, then the probability (21) vanishes as $M \to \infty$, as all the samples eventually reach $\bar{x}$. However, if we scale $\bar{x}$ as a function of M, then the same probability will, in general, not go to zero. In particular, if we set

$$\bar{x} = \bar{x}(M, \tau) = \overline{F}^{-1}\left(\frac{\tau}{M}\right), \tag{23}$$

---

[2]Including the Gaussian correction to the saddle point only leads to subdominant corrections to the scaling of k_c with M.

[3]We could also consider only a fraction of the $X^{(j)}$'s to be below $\bar{x}$; however, this does not significantly alter the scaling of k_c.

where τ is an arbitrary small constant and $\overline{F}^{-1}$ is the inverse of $\overline{F}$, then the probability of having all the samples smaller than $\bar{x}$ is equal to $e^{-\tau}$ for all M. The region $(-\infty,\bar{x}]$ therefore defines a typical region for the sample $\{X^{(j)}\}$: As M grows, all samples will fall in that region with constant probability.

With this result we now define the truncated generating function

$$G_{M,\tau}(k) = \int_{-\infty}^{\bar{x}(M,\tau)} e^{kx} p(x) \, dx. \qquad (24)$$

Depending on k and M, two different situations then arise for $x^*(k)$. On the one hand, if $x^*(k) < \bar{x}(M,\tau)$, then

$$\hat{G}_M(k) \approx G_{M,\tau}(k) \approx G(k), \qquad (25)$$

and we are away from the linearization regime. On the other hand, if $x^*(k) > \bar{x}(M,\tau)$, then $G(k)$ is not well approximated by $G_{M,\tau}(k)$ or $\hat{G}_M(k)$ since $x^*(k)$, the concentration point of $G(k)$, lies outside the typical values covered by the sample. In this case, one must either increase M for a given k so that $\bar{x}(M,\tau)$ reaches $x^*(k)$ or decrease k for a given M so that $x^*(k)$ reaches $\bar{x}(M,\tau)$. The threshold value of k for which $x^* = \bar{x}$ defines k_c; thus

$$x^*(k_c) = \bar{x}(M,\tau). \qquad (26)$$

This result yields the upper bound k_c^+; a similar calculation yields the lower bound k_c^- mentioned before by considering $P(X_{\min} > \bar{x})$ instead of $P(X_{\max} < \bar{x})$.

This analysis gives estimates for k_c^- and k_c^+ that are good enough for practical purposes, as they capture the scaling of the linearization effect with M based on the tail behavior of $p(x)$ in (17).[4] For example, if X is distributed according to the Gaussian density with $p(x) \sim e^{-x^2/2}$ as $|x| \to \infty$, then we obtain from (23)

$$\bar{x}(M,\tau) = 2\,\mathrm{erfc}^{-1}\left(\frac{2}{M}\right). \qquad (27)$$

Moreover, the concentration point for this distribution is located at $x^*(k) = k$. Combining this with (27) in (26) and using known asymptotics for the complementary error function then yields

$$k_c^{\pm} \approx \pm\sqrt{\ln M}. \qquad (28)$$

More generally, if

$$p(x) \approx e^{-|x|^{\rho}}, \quad \rho > 1, \qquad (29)$$

as $|x| \to \infty$, then (26) yields

$$k_c^{\pm} \approx \pm(\ln M)^{1-1/\rho}. \qquad (30)$$

The full derivation of this result can be found in [60]. For both cases, the estimate of k_c does not depend on τ, as shown in [60]. If, however, $p(x)$ is bounded from above at $x_{\max}$ and behaves like

$$p(x) \approx (x_{\max} - x)^{\beta}, \quad \beta > 0, \qquad (31)$$

for $x < x_{\max}$ as $x \to x_{\max}$ from below, then

$$k_c^+ \approx \left(\frac{M}{\tau}\right)^{\frac{1}{\beta+1}}. \qquad (32)$$

In this case, there is an explicit dependence on τ, which for applications can be set to some fixed but otherwise arbitrary value. A similar scaling is obtained for bounded (e.g., uniform) random variables and finite, discrete random variables.

C. Statistical errors

Estimators must be supplemented by statistical errors to be meaningful. Commonly, this is done by assuming that the distribution of the sum defining an estimator converges to a Gaussian distribution around its mean, which leads to defining the "dispersion" or error of the estimator as its standard error. In our case, we have to be careful with this error definition: Since the variance of the random variable e^{kX} is

$$\mathrm{var}(e^{kX}) = G(2k) - G(k)^2, \qquad (33)$$

the variance of $\hat{G}_M(k)$ is defined only on half the range on which $\hat{G}_M(k)$ converges. Moreover, although that estimator is known to converge for $k < k_c$, that convergence may not be to a Gaussian random variable, which prevents us from using the standard deviation as an error measure already from $k_c/2$.

This basic observation is supported by rigorous mathematical results obtained recently by Ben Arous and collaborators [61] which show for a general class of random variables[5] that $\hat{G}_M(k)$ converges, when properly rescaled, to a Gaussian random variable for all $k = k_c(M)/\alpha$ when $\alpha > 2$. It then converges to a Lévy α-stable random variable for $k = k_c(M)/\alpha$ when $1 < \alpha \leqslant 2$, whereas there is no convergence when $\alpha < 1$. This means overall that we have to consider three regions for defining error bars.

(1) $k \leqslant k_c(M)/2$. $\hat{G}_M(k)$ is *self-averaging*, meaning that it converges in probability to $G(k)$ as $M \to \infty$. Moreover, this estimator is Gaussian-distributed around $G(k)$, so that its standard variance can be used as an error estimate.

(2) $k_c(M)/2 < k \leqslant k_c(M)$. $\hat{G}_M(k)$ is self-averaging, but is not Gaussian-distributed around $G(k)$, so that standard (Gaussian) error bars are inadequate.

(3) $k > k_c(M)$. $\hat{G}_M(k)$ is not self-averaging, so there is no convergence to $G(k)$ as $M \to \infty$.

We detail each region next and explain its consequences for defining errors for $\hat{G}_M(k)$, $\hat{\lambda}_M(k)$, and $\hat{I}_M(a)$. For simplicity, we only discuss the upper bound k_c^+; errors concerning k_c^- are defined similarly.

1. Error estimates below $k_c/2$

In this region, the error bar for $\hat{G}_M(k)$ can be defined as its standard deviation, which is estimated from the empirical variance:

$$\mathrm{err}[\hat{G}_M(k)] = \sqrt{\mathrm{var}[\hat{G}_M(k)]} = \frac{1}{\sqrt{M}}\sqrt{\hat{G}_M(2k) - \hat{G}_M(k)^2}. \qquad (34)$$

---

[4]A similar analysis of sample extremes was developed for a more specific model by Hurtado and Garrido [59] to study statistical errors in the cloning algorithm [41–43].

[5]This class includes the Gaussian distribution, the Gamma distribution, and the stretched exponential distribution (29), among many others.

Computing from this error the error of $\hat{\lambda}_M(k)$ is not trivial because the latter is a nonlinear function of $\hat{G}_M(k)$. However, for small errors we can perform a Taylor expansion of (8) to first order, as commonly done in physics [62], to obtain

$$\text{err}[\hat{\lambda}_M(k)] = \frac{\text{err}[\hat{G}_M(k)]}{\hat{G}_M(k)}. \tag{35}$$

With this error, we can define the error of the rate function estimator $\hat{I}_M(a)$ parametrically by varying k, as explained before. For a given $k < k_c(M)/2$, we first compute $\hat{a}_M(k)$ as in (10). Denoting the numerator of the right-hand side of (10) by $\hat{H}_M(k)$, we next estimate the error of $\hat{a}_M(k)$ as

$$\text{err}[\hat{a}_M(k)] = \sqrt{\left\{\frac{\text{err}[\hat{H}_M(k)]}{\hat{G}_M(k)}\right\}^2 + \left\{\frac{\text{err}[\hat{G}_M(k)]\hat{H}_M(k)}{\hat{G}_M(k)^2}\right\}^2}, \tag{36}$$

which follows by applying a Taylor expansion to the definition of $\hat{a}_M(k)$ and by neglecting the correlation between the numerator and denominator.[6] Given the Legendre transform (5) or (9), it is then natural to define the error for $\hat{I}_M(a)$ at $a = \hat{a}_M(k)$ as the sum of the errors:

$$\text{err}\{\hat{I}_M[\hat{a}_M(k)]\} = \sqrt{k^2 \text{err}[\hat{a}_M(k)]^2 + \text{err}[\hat{\lambda}_M(k)]^2}. \tag{37}$$

This quantity probably overestimates the error; however, it is the simplest error that one can define, based on the linear form of the Legendre transform, which does not assume any dependency between $\hat{a}_M$ and $\hat{\lambda}_M$.

2. Error estimates between $k_c/2$ and k_c

In this region, linearization sets in from $k_c(M)$, leading $\hat{G}_M(k)$ to converge to $G(k)$ but in a non-Gaussian way, which implies that the standard deviation calculated from M realizations cannot be used as an error estimate. In this case, it is common to define the error of estimators not from one sample $\{X^{(j)}\}_{j=1}^M$, but from R such samples of size M, called *repetitions*. For the SCGF this means generating R independent samples of size M, yielding R estimators $\hat{\lambda}_M^{(\ell)}(k), \ell = 1, \ldots, R$, which are averaged to yield the following estimate of $\lambda(k)$:

$$\hat{\lambda}_{R \times M}(k) = \frac{1}{R} \sum_{\ell=1}^R \lambda_M^{(\ell)}(k). \tag{38}$$

The error for this estimation is then obtained by computing the standard error over the R repetitions:

$$\text{err}[\hat{\lambda}_{R \times M}(k)] = \frac{\text{err}[\hat{\lambda}_M(k)]}{\sqrt{R}}. \tag{39}$$

The error estimate for $\hat{I}_M(a)$ can be defined similarly using repetitions and the error method presented for one sample. In this case, the repetition error $\text{err}[\hat{a}_{R \times M}(k)]$ must be computed as in (39) and added as in (37) to the repetition error of $\hat{\lambda}_{R \times M}(k)$.

---

[6]The numerator and denominator of (10) are not independent, but this is a necessary approximation to be able to obtain an error estimate.

Though more computationally intensive, the use of repetitions provides better error estimates for $\hat{\lambda}_M$ and $\hat{I}_M$, as the logarithm in $\hat{\lambda}_M(k)$ has the effect of regularizing the extreme values (and thus the linearization) of $\hat{G}_M(k)$. In practice, a sufficiently large sample can be partitioned into R smaller samples to apply this method. Alternatively, bootstrap methods can be used to generate new samples from the empirical distribution of the sample already obtained [63–65].

3. Error estimates above k_c

In this region, estimators do not converge, leaving the computation of $\hat{G}_M(k), \hat{\lambda}_M(k)$, and $\hat{I}_M(a)$ meaningless. To have an idea of the variability of these estimators, one can estimate them over R repetitions involving M samples, as before, and extract the first decile and last decile of these realizations. This can be taken as a measure of the error. Our results indicate, however, that such an error is typically very large and only confirms the fact that no useful information can be inferred about $\lambda(k)$ beyond $k_c(M)$.

III. TEST CASES

We illustrate in this section the previous results about estimator convergence for four types of distributions: Gaussian, exponential, Bernoulli, and power law. Gaussian distributions have been extensively studied in the context of the Jarzynski estimator [31–36] and are revisited here to illustrate the case of unbounded random variables. The exponential distribution is considered as a limiting case of the saddle-point analysis, whereas Bernoulli random variables illustrate our results for the bounded case and are relevant for data network applications [37–40]. Finally, power-law distributions are considered to discuss the case where $\lambda(k)$ diverges and large-deviation functions do not exist. Other distributions fall, as will be explained, in each of these cases with only minor differences in the behavior of $\hat{\lambda}_M(k)$ and k_c. Physical applications and non-IID random variables are discussed in the next section.

A. Unbounded distributions

We consider as in (6) a sample mean A_n of n IID random variables and assume that these are distributed according to a Gaussian distribution with mean μ and variance σ^2. The exact SCGF in this case is

$$\lambda(k) = \mu k + \frac{\sigma^2}{2} k^2, \quad k \in \mathbb{R}. \tag{40}$$

Figure 1(a) shows the estimation of this SCGF based on the estimator $\hat{\lambda}_M(k)$ using a sample $\{X^{(j)}\}_{j=1}^M$ of $M = 1000$ Gaussian random variables with $\mu = 0$ and $\sigma^2 = 1$. The relatively low sample size is used to obtain visible error bars. Moreover, rather than plotting $\hat{\lambda}_M$ as a function of k and identifying k_c for varying M, we plot the estimator as a function of k/k_c using the *a priori* estimate given in (28), so that the convergence region is fixed at $|k/k_c| = 1$ for all M. In this way, plots of $\hat{\lambda}_M$ obtained for different M look similar, provided that M is large enough; hence, we show only one value of M in Fig. 1(a). Note that, because of the choice $\mu = 0$, we have $-k_c^- = k_c^+ = k_c$; for asymmetric Gaussian distributions, $|k_c^-|$ is slightly different from k_c^+, but this does not affect the scaling of k_c.
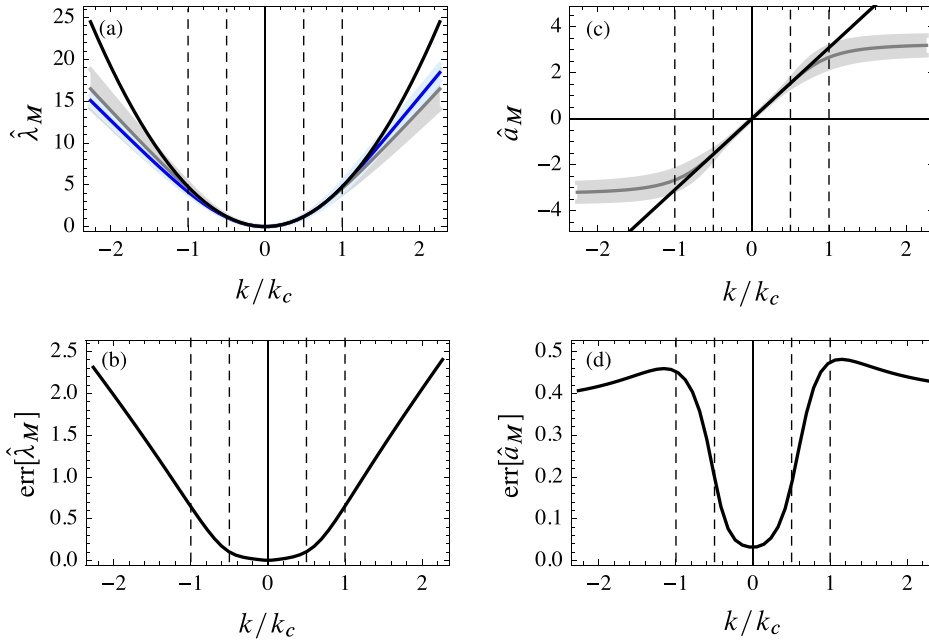
FIG. 1. (Color online) SCGF estimator for Gaussian random variables. Parameters: $\mu = 0$, $\sigma = 1$, $M = 1000$, and $R = 200$ repetitions (except otherwise noted). (a) Blue curve, $\hat{\lambda}_M(k)$ with statistical error (blue shaded area) for a single repetition ($R = 1$); gray curve, $\hat{\lambda}_M(k)$ with statistical error (gray shaded area); black curve, exact $\lambda(k)$. (b) Statistical error of $\hat{\lambda}_M(k)$ showing the linearization effect. (c) Estimator $\hat{a}_M(k)$ of the derivative of $\lambda(k)$ with statistical error (gray shaded area). (d) Statistical error for $\hat{a}_M(k)$.

The convergence and linearization regions of $\hat{\lambda}_M(k)$ are clearly visible in Fig. 1(a). For $|k/k_c| < 1$, $\hat{\lambda}_M(k)$ agrees with the exact $\lambda(k)$ of (40) within the statistical errors determined from either $R = 1$ or $R > 1$ repetitions. For $|k/k_c| < 1/2$, we have checked that both errors have similar magnitude (not shown), whereas for $1/2 < |k/k_c| < 1$ the two errors differ slightly (also not shown). More importantly, for $|k/k_c| > 1$, $\hat{\lambda}_M(k)$ starts to differ significantly from the exact $\lambda(k)$ because of the linearization coming from the maximum and minimum sample values. The gray curve in Fig. 1(a) shows this linearization for the R repetition estimate as compared to the single repetition (blue curve). The former is, in general, more stable than the latter because of the averaging coming from the R repetitions; however, both estimators give results that are off the exact SCGF because $\hat{\lambda}_M(k)$ and its error do not converge for $|k/k_c| > 1$.

The linearization effect is also seen in the repetition error of $\hat{\lambda}_M$ [Fig. 1(b)] and the estimator $\hat{a}_M(k)$ [Fig. 1(c)] of the derivative of $\lambda(k)$. Linearization appears for $\hat{a}_M(k)$ as plateaus with heights given in the R repetition case by the mean of the different maxima and minima contained in the repeated samples. Since the variance of these minima and maxima is independent of k, the statistical error of $\hat{a}_M$ is constant, as seen in Figs. 1(c) and 1(d). Inside the convergence region, $|k/k_c| < 1$, err[$\hat{a}_M$] decreases sharply from $|k/k_c| = 1$ to $|k/k_c| = 1/2$, below which it converges to 0 as $M \to \infty$ for any $R \geqslant 1$. This error behavior is interesting for two reasons. First, it can be interpreted as a "phase transition" or crossover as k is varied, reflecting the transition from Gaussian to non-Gaussian errors at $|k/k_c| = 1/2$. A similar crossover was reported in the behavior of the bias of the Jarzynski estimator [36] and the partition function of glassy systems [56]. Second, it provides a simple way of estimating k_c numerically without knowing the

distribution of the random variables considered: We simply have to find the function $k(M)$ that aligns the maximum of err[$\hat{a}_M(k)$] for different M.

From the estimators $\hat{\lambda}_M$ and $\hat{a}_M$ we may now estimate the rate function $I(a)$ using the parametric relations (10) and (11). The result is shown in Fig. 2 together with the exact result,

$$I(a) = \frac{(a - \mu)^2}{2\sigma^2}, \quad a \in \mathbb{R}. \quad (41)$$

We show in this plot the vertical error bars for the ordinate $\hat{I}_M[\hat{a}_M(k)]$ obtained from (37), as well as horizontal error bars for the abscissa $\hat{a}_M(k)$ corresponding to the repetition error err[$\hat{a}_M(k)$] given in (36). Also indicated is the value $\hat{a}_M(k_c)$, corresponding for $\mu = 0$ and $\sigma = 1$ to $a = 1$, which
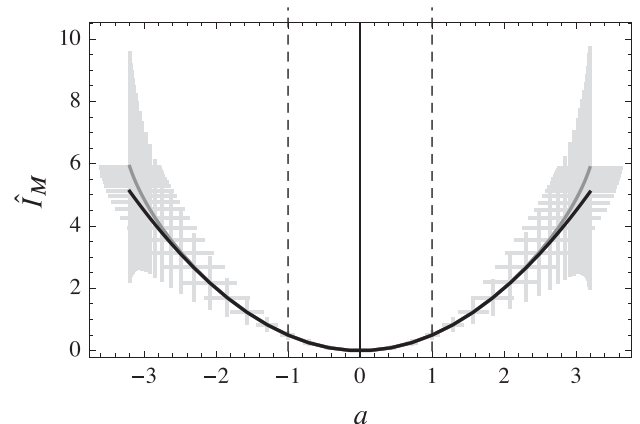

FIG. 2. Rate function estimator for Gaussian random variables. Parameters: $\mu = 0$, $\sigma = 1$, $M = 1000$, and $R = 200$. Gray curve, $\hat{I}_M(a)$ with statistical error bars; black curve, exact rate function.
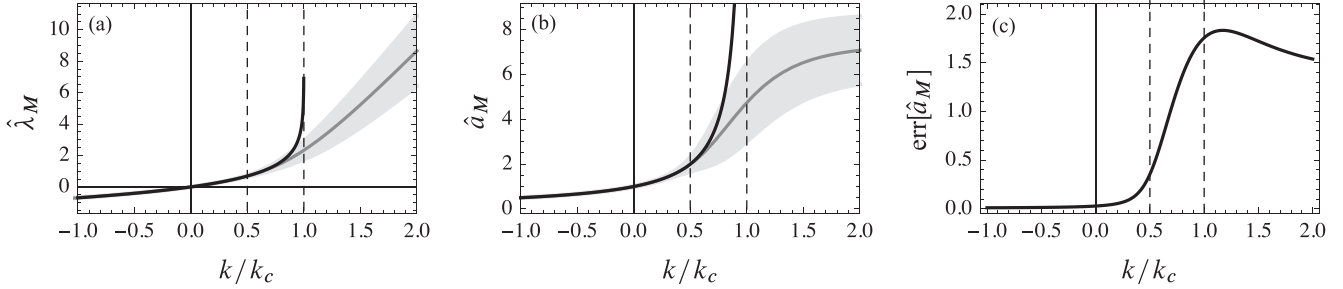
FIG. 3. SCGF estimators for exponential random variables. Parameters: $\mu = 1$, $M = 1000$, $R = 200$, $k_c = k_c^+ = 1$. (a) Gray curve, $\hat{\lambda}_M(k)$ with statistical error (gray shaded area); black curve, exact $\lambda(k)$. (b) Gray curve, $\hat{a}_M(k)$ with statistical error (gray shaded area); black curve, exact derivative of $\lambda(k)$. (c) Statistical error of $a_M(k)$.

bounds the convergence region of $I(a)$ where the error bars decrease as R and M are increased. In the results shown in Fig. 2, the errors for $|a/\hat{a}_M(k_c)| < 1$ are actually smaller than the width of the curves, whereas they increase substantially for $|a/\hat{a}_M(k_c)| > 1$. This comes again from the linearization problem affecting all estimators above k_c, but also from the k factor in the Legendre transform (11), which magnifies the error on $\hat{a}_M$ following (37).

Similar results will be obtained for other distributions which, as for the Gaussian, are unbounded for $x > 0$ and $x < 0$. In this case, $|k_c^{\pm}|$ will grow with M, as in the Gaussian case, with a speed given by the tail behavior of the distribution considered, following our results of Sec. II.

B. One-sided exponential distributions

We consider next the exponential distribution,

$$p(x) = \frac{1}{\mu} e^{-x/\mu}, \quad x \geqslant 0, \quad (42)$$

with mean μ, as representative of random variables that are bounded from below. This distribution corresponds to the limit case $\rho = 1$ in (29) and therefore falls, in principle, outside the results of [61]. Given, however, that its SCGF

$$\lambda(k) = -\ln(1 - \mu k) \quad (43)$$

is defined for $k < 1/\mu$, we can set $k_c^+ = 1/\mu$, since $\hat{\lambda}_M(k)$ is defined for all $k \in \mathbb{R}$, whereas $\lambda(k)$ exists only for $k < 1/\mu$, so that the part of $\hat{\lambda}_M(k)$ beyond k_c^+ is a finite sample artifact. This constant k_c^+ is also consistent with our estimate (26) of $k_c(M)$ and arises for any distributions with exact or asymptotic exponential tails. On the other hand, we find $k_c^- = -\infty$, since $p(x)$ is bounded below at $x_{\min} = 0$, so that the minimum of the sample $\{X^{(j)}\}_{j=1}^M$ converges rapidly to 0.

Figure 3(a) shows the result of $\hat{\lambda}_M(k)$ for a sample size $M = 1000$ and statistical errors calculated with $R = 200$ repetitions, plotted as a function of k/k_c. The linearization effect is clearly seen for $\hat{\lambda}_M(k)$, as well as for $\hat{a}_M(k)$, which correctly saturates to the lower bound $x = 0$ for $k \to -\infty$, but incorrectly saturates for $k > k_c^+$; see Fig. 3(b). The main difference with the Gaussian case is that, since k_c^+ is now constant, the convergence of $\hat{\lambda}_M(k)$ to $\lambda(k)$ is not accompanied by an increased region of k, where this convergence takes place; all that changes as $M \to \infty$ is the slope of $\hat{\lambda}_M$ or, equivalently, the value $\hat{a}_M(k)$, which diverges to reach the

asymptote of $\lambda(k)$. Because $k_c^- = -\infty$, we also see that the repetition error of both $\hat{\lambda}_M(k)$ and $\hat{a}_M(k)$ converges uniformly to 0 for all $k < k_c^+/2$. In Figs. 3(a) and 3(b), the error bars in that region are actually smaller than the width of the black lines representing $\lambda(k)$ and $a_M(k)$, respectively. For $k > k_c^+/2$, the error is similar to the Gaussian case: It sharply increases between $k_c^+/2$ and k_c^+ and saturates for $k > k_c^+$, providing again a way to estimate k_c.

The asymmetric convergence of $\hat{\lambda}_M$ and $\hat{a}_M$ leads naturally to an asymmetry in the estimation of the rate function, shown in Fig. 4. For $a < a(k_c^+/2)$ the rate function is correctly estimated and matches the exact rate function,

$$I(a) = \frac{a}{\mu} - 1 - \ln \frac{a}{\mu}, \quad a > 0, \quad (44)$$

with errors bars smaller than the width of the curve representing this function, whereas for $a > a(k_c^+/2)$, the linearized $\hat{\lambda}_M$ and $\hat{a}_M$ lead to an estimation of $I(a)$ with large error bars.

Other distributions with asymptotic exponential tails lead to similar results. In particular, for distributions with left and right exponential tails, k_c^- and k_c^+ are both constant with M.

C. Bounded distributions

The last test case of interest is the class of bounded random variables for which $\lambda(k)$ is exactly or asymptotically linear
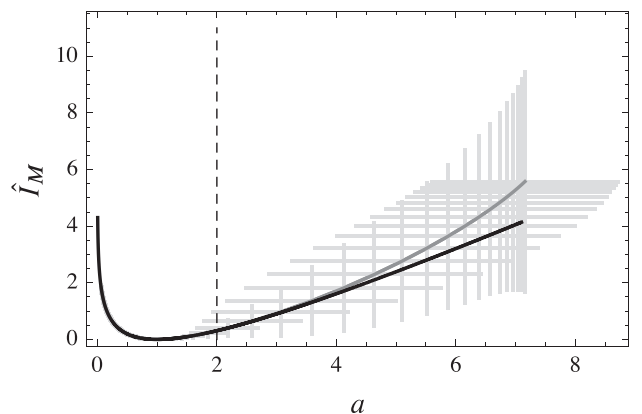

FIG. 4. Rate function estimator for exponential random variables. Gray curve, $\hat{I}_M(a)$ with statistical error bars; black curve, exact result; dashed line, $a(k)$ for $k_c^+/2 = 1/(2\mu)$. Parameters: $\mu = 1$, $M = 1000$, $R = 200$.
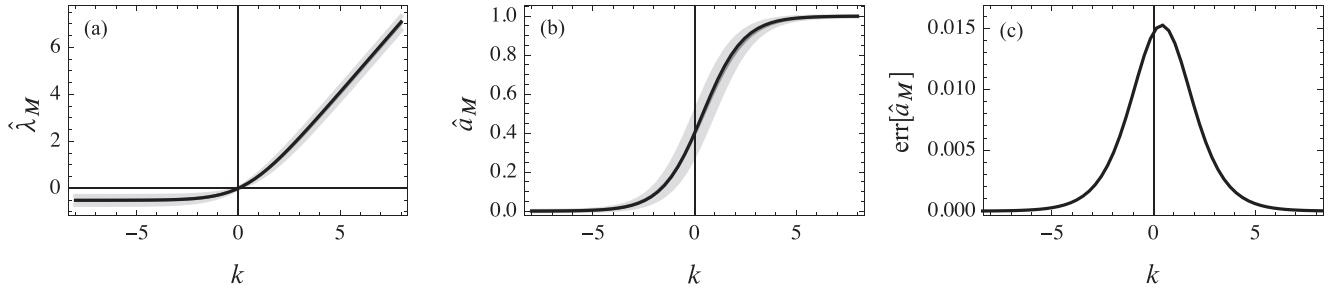
FIG. 5. SCGF estimators for Bernoulli random variables. Parameters: $\alpha = 0.4$, $M = 20$, $R = 1$. (a) Gray curve, $\hat{\lambda}_M(k)$ with statistical error (gray shaded area); black curve, exact $\lambda(k)$. (b) Gray curve, $\hat{a}_M(k)$ with its statistical error (gray shaded area); black curve, exact derivative of $\lambda(k)$. (c) Statistical error of $a_M(k)$.

and so for which $k_c^- = -\infty$ and $k_c^+ = \infty$. To illustrate this case, we consider Bernoulli random variables taking values $\{0,1\}$ with probabilities $p(1) = \alpha$ and $p(0) = 1 - \alpha$, where $\alpha \in [0,1]$, so that

$$\lambda(k) = \ln(\alpha\, e^k + 1 - \alpha), \quad k \in \mathbb{R}. \tag{45}$$

Figure 5 shows the results of estimating this SCGF obtained for $M = 20$ and plotted now as a function of k and not k/k_c, since $k_c = \infty$. Already for such a small sample size, the estimators $\hat{\lambda}_M$ and $\hat{a}_M$ are very accurate, compared to $M = 1000$ used in the Gaussian and exponential cases. The single and repetition errors essentially agree for all k and decrease uniformly for all k as $M \to \infty$. Figure 5(c) shows more precisely that the statistical error of $\hat{a}_M(k)$ is largest at $k = 0$ and decreases to 0 as $k \to \pm\infty$. This is due to the fact that the "true" linear behavior of $\lambda(k)$ as $k \to \pm\infty$ is correctly estimated as soon as the sample $\{X^{(j)}\}$ contains one 0 and one 1, whereas the exact form of $\lambda(k)$ around $k = 0$ depends on α, which is correctly estimated as $M \to \infty$. However, both regions have errors of the same magnitude, which translate into uniform errors for the estimated rate function, shown in Fig. 6. Here we see that, despite the small sample size $M = 20$, the estimator is close to the exact rate function

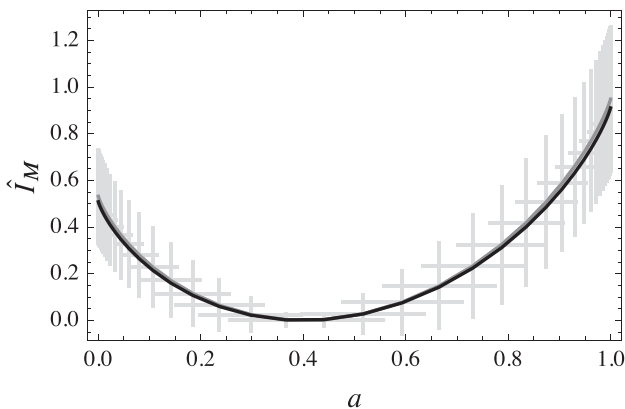$$I(a) = a \ln \frac{a}{\alpha} + (1 - a) \ln \frac{1 - a}{1 - \alpha}, \quad a \in [0,1], \tag{46}$$

FIG. 6. Rate function estimator for Bernoulli random variables. Gray curve, $\hat{I}_M(a)$ with statistical error bars; black curve, exact result. Parameters: $\alpha = 0.4$, $M = 20$, $R = 1$.

with error bars that are significantly reduced if we were to use $M = 1000$. This comes again from the fact that linearization is not an artifact in this case: The bounded support of $p(x)$ is covered by the sample for a finite M, which means essentially that $k_c = \infty$.

These results confirm previous results obtained for data networks [37–40], showing that the estimation of large-deviation functions from a data stream of bits converge fast and uniformly. For other distributions with bounded support, convergence is expected to be as fast as for the Bernoulli case, with the difference that k_c may not be equal to ∞ following our results (31) and (32). For a distribution $p(x)$ that vanishes linearly, for example, we obtain $k_c \sim M^{1/2}$ from (26), whereas if $p(x)$ decays like a parabola, we obtain $k_c \sim M^{1/3}$.

Distributions that have a fixed "window" or "vertical cutoff," such as the uniform distribution or the Bernoulli distribution, represent a limit case of bounded distributions for which $k_c = \infty$. These distributions behave similarly, whether they are discrete or continuous, because their SCGFs have asymptotic linear branches, which is the property responsible for $k_c = \infty$.

D. Divergent generating functions

To close this section, we briefly discuss the case where the theoretical SCGF diverges everywhere except at $k = 0$, which arises when X is distributed for example according to a Lévy stable distribution or, more generally, any power-law tail distributions [7]. Assuming that the distribution of interest is two-sided with both tails behaving as a power-law, then $\lambda(k) = \infty$ for $k \neq 0$, which is consistent with the fact that the distribution of A_n does not have a large-deviation form; see Example 4.2 of [7].

The estimator $\hat{\lambda}_M(k)$ in this case exists for all k when computed on finite samples, since it is a finite sum of exponentials. It is easy to check, however, that it diverges for all $k \neq 0$ as the sample size M is increased. Thus, there is no convergence region for the SCGF estimator, except at $k = 0$ where we trivially have $\hat{\lambda}_M(0) = \lambda(k) = 0$ for all M, so that k_c is effectively equal to 0. This is similar to the exponential case for which k_c is set to the limit of the convergence region of $\lambda(k)$. Here we have $k_c^- = k_c^+ = 0$ simply because the convergence region of a distribution with left and right power-law tails is $\{0\}$. If the distribution has only one power-law tail, say to the right, then $k_c^+ = 0$, whereas k_c^- will behave according to the shape of the other tail following the cases above.

This applies if there is no bound (experimental or numerical) on the values of X that can be measured. If we increase M without increasing measurement bounds, then $\hat{\lambda}_M(k)$ will, of course, behave as if the quantity sampled is bounded and will thus represent the distribution of that bounded quantity having a finite SCGF rather than the theoretical unbounded quantity having an infinite SCGF. One could then argue that all physical quantities are bounded because of the limited range of measuring devices. However, this is not a fundamental limit: Measurement bounds can always be pushed, in principle, with better devices. From this point of view, a quantity should be considered as unbounded when the theoretical range of values that can be measured can always be made large enough to include the maximum and minimum values actually measured in any given large but finite samples.

IV. CORRELATED OBSERVABLES

We assumed in the previous sections that the X_i's were independent to illustrate in the simplest way possible the linearization effect limiting the estimation of large-deviation functions. We now consider observables involving correlated random variables, representing, for example, the individual state of interacting particles or the state of a stochastic process sampled in time. In many cases of interest, these observables involve weakly interacting components (in space or time) which can be grouped into independent or asymptotically independent blocks. This is the basis of the *block averaging* method, proposed independently in the context of free energy calculations [34] and large-deviation theory [37–40].

We explain this method following [40]; see also [54]. We consider again the sample mean

$$A_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \tag{47}$$

but assume now that the sequence of random variables $X_1, X_2, \ldots, X_n$ forms a Markov chain. In this case, the SCGF of A_n defined in (4) does not simplify to the simple cumulant function (7). However, it is possible to "group" the X_i's into blocks as

$$\underbrace{X_1 + \cdots + X_b}_{Y_1} + \underbrace{X_{b+1} + \cdots + X_{2b}}_{Y_2} + \cdots$$

$$+ \underbrace{X_{n-b+1} + \cdots + X_n}_{Y_K}, \tag{48}$$

where $K = n/b$ is the number of blocks of size b, so as to rewrite the sample mean as

$$A_n = \frac{1}{bK} \sum_{i=1}^{K} Y_i. \tag{49}$$

For *mixing* Markov chains having a finite correlation length, it can be shown that the blocks Y_i become independent in the limit where $n \to \infty$ and $b \to \infty$ but with b growing slower than n so that $K \to \infty$ [40]. Moreover, if the chain is ergodic, then the Y_i's become identically distributed for i large enough, so that

$$\frac{1}{n} \ln \langle e^{nkA_n} \rangle \approx \frac{K}{n} \ln \langle e^{kY_i} \rangle = \frac{1}{b} \ln \langle e^{kY_i} \rangle. \tag{50}$$

We are thus back to the problem of estimating the SCGF for an IID sequence of random variables formed by the Y_i's instead of the X_i's, so that our estimator for $\lambda(k)$ is now

$$\hat{\lambda}_M(k) = \frac{1}{b} \ln \frac{1}{M} \sum_{j=1}^{M} e^{kY^{(j)}}, \tag{51}$$

where $Y^{(j)}$, $j = 1, \ldots, M$, are M IID copies of the blocked random variable Y. The estimation of the rate function follows as before from (9).

In practice, this block averaging method works well by choosing a finite b greater than the correlation length of the Markov chain or, equivalently, by varying b until the estimated SCGF of A_n ceases to depend on b. This avoids taking the double limit $n \to \infty$, $b \to \infty$, with $K = n/b \to \infty$.

The method can also be used for integrated observables of continuous-time Markov processes having the form

$$A_T = \frac{1}{T} \int_0^T f(X_t) dt, \tag{52}$$

where f is an arbitrary function of the state X_t of the Markov process. In this case, the integral is "blocked" in segments of length b to obtain

$$A_T = \frac{1}{bK} \sum_{i=1}^{K} Y_i, \tag{53}$$

where Y_i is the integral of $f(X_t)$ over the time interval $[(i-1)b, ib]$ and $K = T/b$ is as before the number of blocks over the total interval $[0, T]$.

Other observables that can be expressed in the block form (53) include the total activity of interacting particle systems, defined as the total number of particle jumps accumulated over a time T, the total integrated current which depends on the jumps and their direction, and observables of equilibrium systems. For example, one can divide the energy E_N of an N-particle system into additive blocks Y_i involving $b < N$ particles which become asymptotically decoupled as the limits $N \to \infty$ and $b \to \infty$ are taken, with b growing slower than N. In this limit, E_N is thus effectively treated as a sample mean of $K = N/b$ IID random variables. This works as long as the interactions between particles are weak or short-range, which is the spatial analog of a mixing Markov process.

In all cases, the distribution of the IID or near-IID blocks Y_i determines how quickly the estimated SCGF converges to its theoretical value according to the test cases studied previously. If this distribution has an unbounded support and decays faster than an exponential, then the convergence threshold k_c of the SCGF is expected to grow slowly with the sample size M, as in the Gaussian case, whereas if its tails decay like an exponential, then k_c is essentially constant, as seen before. Bounded block distributions, on the other hand, are characterized by a k_c that grows rapidly with M, and represent the best possible case in terms of estimation.

The use of block averaging techniques is important as it yields an exponential gain in estimation compared to the direct sampling of large-deviation probabilities. To see this, suppose that we want to estimate the probability p of an event known to scale in a large-deviation way as $p \sim e^{-nI}$ with the parameter n, which can be N or T as above. It is known (see [66] or [67])

that the direct sampling of that probability requires roughly $M \sim e^n$ samples to obtain a relative error,

$$r_M = \frac{\hat{P}_M - p}{p}, \tag{54}$$

for the estimate $\hat{P}_M$ of p that is constant in n. By contrast, the estimation of p via $\hat{\lambda}_M$ and $\hat{I}_M$ leads from our results to an error on the actual rate of decay I that decreases with M as $1/\sqrt{M}$ in the convergence region. As this error is multiplied in the large-deviation form of p by n, we must therefore choose $M > n^2$ to obtain a constant relative error for $\hat{P}_M = e^{-n\hat{I}_M}$ as a function of n.

This exponential sampling gain (e^n vs n^2) can obviously be exploited if A_n can be divided into independent or asymptotically independent blocks for a large enough block size b. If this cannot be done or if A_n does not have an additive structure, then we can still obtain the rate function of A_n, in principle, by directly sampling its generating function $G_n(k) = \langle e^{nkA_n} \rangle$ and obtaining its SCGF using (4). However, in this case the estimation is inefficient: The saddle point a^* of $G_n(k)$, which does not scale with n, can be reached only with a sample size $M \sim e^n$ because of the exponential form of $P(A_n = a)$.

Considering our result (26), this means that M must grow exponentially with n for k_c to remain constant as n is increased. Since $M \sim e^n$ is also, as just mentioned, the sample size needed to obtain the rate function of A_n by direct sampling, we see that the generating function method offers no real gain over direct sampling when A_n has no obvious additive structure [54]. Similar results were obtained in the context of free energy estimation [58], where A_n is the work performed on an N-particle system over a time T so that $n = NT$, and for multifractals [56], where A_n is the local dimension measured over a spatial or temporal scale $\epsilon = 1/n$.

For experiments, there is no obvious way to overcome this problem of sampling observables that are not additive; however, for simulations, faster convergence can be achieved using modified sampling techniques, such as importance sampling [54], escort distributions [68–70], and transition path sampling [71], which modify the distribution of A_n to center it essentially at the saddle point $a^*(k)$. Cloning techniques [41–43], which are not based on sampling but rather attempt to obtain $\lambda(k)$ from the multiplicative property of generating functions, can also be used and prove efficient in simulations.

V. CONCLUSION

We have developed in this paper a general method for estimating large-deviation functions from simulation or experimental data and have provided convergence results for estimators of these functions and their errors. Our results establish a separation between bounded random variables, for which the estimation of large-deviation functions converges quickly as a function of sample size, and unbounded random variables, for which convergence is guaranteed only for a certain parameter region, which depends on the tail of the distribution considered. We have proposed a way to determine this convergence region without the *a priori* knowledge of that distribution, based on the fact that statistical errors behave differently inside and outside of that region, and have illustrated our approach for various distributions of interest.

These results can be applied to compute rate functions of any additive observables of equilibrium, nonequilibrium, and manmade systems, in addition to computing multifractal spectra, dispersion exponents, and equilibrium free energies using the Jarzynski estimator, as these are also based on estimating generating functions. Our focus on large deviations brings a new and general insight into these computations, which should play an important role in future experiments designed to probe the fluctuations of microscopic and mesoscopic systems.

[1] J. L. McCauley, Introduction to multifractals in dynamical systems theory and fully developed fluid turbulence, Phys. Rep. **189**, 225 (1990).

[2] B. B. Mandelbrot, *Multifractals and* $1/f$ *Noise* (Springer, New York, 1999).

[3] D. Harte, *Multifractals: Theory and Applications* (CRC Press, New York, 2001).

[4] O. E. Lanford III, Entropy and equilibrium states in classical statistical mechanics, in *Statistical Mechanics and Mathematical Problems*, edited by A. Lenard, Lecture Notes in Physics Vol. 20 (Springer, Berlin, 1973), pp. 1–113.

[5] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics* (Springer, New York, 1985).

[6] Y. Oono, Large deviation and statistical physics, Prog. Theor. Phys. Suppl. **99**, 165 (1989).

[7] H. Touchette, The large deviation approach to statistical mechanics, Phys. Rep. **478**, 1 (2009).

[8] C. Jarzynski, Nonequilibrium Equality for Free Energy Differences, Phys. Rev. Lett. **78**, 2690 (1997).

[9] F. den Hollander, *Large Deviations*, Fields Institute Monograph (American Mathematical Society, Providence, RI, 2000).

[10] B. Derrida, Non-equilibrium steady states: Fluctuations and large deviations of the density and of the current, J. Stat. Mech. (2007) P07023.

[11] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, Stochastic interacting particle systems out of equilibrium, J. Stat. Mech. (2007) P07014.

[12] R. J. Harris and H. Touchette, Large deviation approach to nonequilibrium systems, in *Nonequilibrium Statistical Physics*

of Small Systems: Fluctuation Relations and Beyond, edited by R. Klages, W. Just, and C. Jarzynski, Reviews of Nonlinear Dynamics and Complexity Vol. 6 (Wiley-VCH, Weinheim, 2013), pp. 335–360.

[13] A. Shwartz and A. Weiss, *Large Deviations for Performance Analysis*, Stochastic Modeling Series (Chapman and Hall, London, 1995).

[14] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. (Springer, New York, 1998).

[15] A. Montanari and R. Zecchina, Optimizing Searches via Rare Events, Phys. Rev. Lett. **88**, 178701 (2002).

[16] V. Kishore, M. S. Santhanam, and R. E. Amritkar, Extreme events and event size fluctuations in biased random walks on networks, Phys. Rev. E **85**, 056120 (2012).

[17] C. De Bacco, A. Guggiola, R. Kühn, and P. Paga, Rare events statistics of random walks on networks: Localization and other dynamical phase transitions, arXiv:1506.08436.

[18] P. Abry, V. Pipiras, and H. Wendt, Extreme values, heavy tails and linearization effect: A contribution to empirical multifractal analysis, in *21st GRETSI Symposium on Signal and Image Processing* (Troyes, France, 2007).

[19] J.-F. Muzy, E. Bacry, R. Baile, and P. Poggi, Uncovering latent singularities from multifractal scaling laws in mixed asymptotic regime. Application to turbulence, Europhys. Lett. **82**, 60007 (2008).

[20] E. Bacry, A. Gloter, M. Hoffmann, and J.-F. Muzy, Multifractal analysis in a mixed asymptotic framework, Ann. Appl. Prob. **20**, 1729 (2010).

[21] B. Derrida, Random-energy model: An exactly solvable model of disordered systems, Phys. Rev. B **24**, 2613 (1981).

[22] J.-P. Bouchaud and M. Mézard, Universality classes for extreme value statistics, J. Phys. A **30**, 7997 (1997).

[23] L. Berthier and G. Biroli, Theoretical perspective on the glass transition and amorphous materials, Rev. Mod. Phys. **83**, 587 (2011).

[24] G. Hummer and A. Szabo, Free energy reconstruction from nonequilibrium single-molecule pulling experiments, Proc. Nat. Acad. Sci. USA **98**, 3658 (2001).

[25] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality, Science **296**, 1832 (2002).

[26] N. C. Harris, Y. Song, and C.-H. Kiang, Experimental Free Energy Surface Reconstruction from Single-Molecule Force Spectroscopy Using Jarzynski's Equality, Phys. Rev. Lett. **99**, 068101 (2007).

[27] G. Hummer and A. Szabo, Free energy profiles from single-molecule pulling experiments, Proc. Natl. Acad. Sci. USA **107**, 21441 (2010).

[28] A. N. Gupta, A. Vincent, K. Neupane, H. Yu, F. Wang, and M. T. Woodside, Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements, Nat. Phys. **7**, 631 (2011).

[29] A. Alemany, A. Mossa, I. Junier, and F. Ritort, Experimental free-energy measurements of kinetic molecular states using fluctuation theorems, Nat. Phys. **8**, 688 (2012).

[30] S. Kim, Y. W. Kim, P. Talkner, and J. Yi, Comparison of free-energy estimators and their dependence on dissipated work, Phys. Rev. E **86**, 041130 (2012).

[31] R. H. Wood, W. C. F. Muhlbauer, and P. T. Thompson, Systematic errors in free energy perturbation calculations due to a finite sample of configuration space: Sample-size hysteresis, J. Phys. Chem. **95**, 6670 (1991).

[32] J. Gore, F. Ritort, and C. Bustamante, Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements, Proc. Natl. Acad. Sci. USA **100**, 12564 (2003).

[33] D. M. Zuckerman and T. B. Woolf, Theory of a Systematic Computational Error in Free Energy Differences, Phys. Rev. Lett. **89**, 180602 (2002).

[34] D. M. Zuckerman and T. B. Woolf, Overcoming finite-sampling errors in fast-switching free-energy estimates: Extrapolative analysis of a molecular system, Chem. Phys. Lett. **351**, 445 (2002).

[35] M. Palassini and F. Ritort, Improving Free-Energy Estimates from Unidirectional Work Measurements: Theory and Experiment, Phys. Rev. Lett. **107**, 060601 (2011).

[36] A. Suárez, R. Silbey, and I. Oppenheim, Phase transition in the Jarzynski estimator of free energy differences, Phys. Rev. E **85**, 051108 (2012).

[37] S. Crosby, I. Leslie, B. McGurk, J. T. Lewis, R. Russell, and F. Toomey, Statistical properties of a near-optimal measurement-based CAC algorithm, in *Proceedings of the IEEE ATM Workshop 1997* (IEEE, 1997), pp. 103–112.

[38] N. G. Duffield, J. T. Lewis, N. O'Connell, R. Russell, and F. Toomey, Entropy of ATM traffic streams: A tool for estimating QoS parameters, IEEE J. Sel. Areas Commun. **13**, 981 (1995).

[39] J. T. Lewis, R. Russell, F. Toomey, B. McGurk, S. Crosby, and I. Leslie, Practical connection admission control for ATM networks based on on-line measurements, Comp. Commun. **21**, 1585 (1998).

[40] K. Duffy and A. P. Metcalfe, The large deviations of estimating rate functions, J. Appl. Probab. **42**, 267 (2005).

[41] V. Lecomte and J. Tailleur, A numerical approach to large deviations in continuous time, J. Stat. Mech. (2007) P03004.

[42] J. Tailleur and J. Kurchan, Probing rare physical trajectories with lyapunov weighted dynamics, Nat. Phys. **3**, 203 (2007).

[43] C. Giardina, J. Kurchan, V. Lecomte, and J. Tailleur, Simulating rare events in dynamical processes, J. Stat. Phys. **145**, 787 (2011).

[44] M. Gorissen, J. Hooyberghs, and C. Vanderzande, Density-matrix renormalization-group study of current and activity fluctuations near nonequilibrium phase transitions, Phys. Rev. E **79**, 020101 (2009).

[45] M. Gorissen and C. Vanderzande, Finite size scaling of current fluctuations in the totally asymmetric exclusion process, J. Phys. A: Math. Theor. **44**, 115005 (2011).

[46] M. Merolle, J. P. Garrahan, and D. Chandler, Space-time thermodynamics of the glass transition, Proc. Natl. Acad. Sci. USA **102**, 10837 (2005).

[47] L. O. Hedges, R. L. Jack, J. P. Garrahan, and D. Chandler, Dynamic order-disorder in atomistic models of structural glass formers, Science **323**, 1309 (2009).

[48] D. Chandler and J. P. Garrahan, Dynamics on the way to forming glass: Bubbles in space-time, Ann. Rev. Chem. Phys. **61**, 191 (2010).

[49] P. I. Hurtado, C. P. Espigares, J. J. del Pozo, and P. L. Garrido, Thermodynamics of currents in nonequilibrium diffusive systems: Theory and simulation, J. Stat. Phys. **154**, 214 (2014).

[50] N. Kumar, S. Ramaswamy, and A. K. Sood, Symmetry Properties of the Large-Deviation Function of the Velocity of a Self-Propelled Polar Particle, Phys. Rev. Lett. **106**, 118001 (2011).

[51] N. Kumar, H. Soni, S. Ramaswamy, and A. K. Sood, Anisotropic isometric fluctuation relations in experiment and theory on a self-propelled rod, Phys. Rev. E **91**, 030102 (2015).

[52] P. H. Haynes and J. Vanneste, Dispersion in the large-deviation regime. Part 1: Shear flows and periodic flows, J. Fluid Mech. **745**, 321 (2014).

[53] H. Touchette, Methods for calculating nonconcave entropies, J. Stat. Mech. (2010) P05008.

[54] H. Touchette, A basic introduction to large deviations: Theory, applications, simulations, in *Modern Computational Science 11: Lecture Notes from the 3rd International Oldenburg Summer School*, edited by R. Leidl and A. K. Hartmann (BIS-Verlag der Carl von Ossietzky Universität Oldenburg, 2011).

[55] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978).

[56] F. Angeletti, M. Mézard, E. Bertin, and P. Abry, Linearization effect in multifractal analysis: Insights from the random energy model, Physica D (Amsterdam, Neth.) **240**, 1245 (2011).

[57] H. Touchette and C. Beck, Asymptotics of superstatistics, Phys. Rev. E **71**, 016131 (2005).

[58] C. Jarzynski, Rare events and the convergence of exponentially averaged work values, Phys. Rev. E **73**, 046105 (2006).

[59] P. I. Hurtado and P. L. Garrido, Current fluctuations and statistics during a large deviation event in an exactly solvable transport model, J. Stat. Mech. (2009) P02032.

[60] F. Angeletti, E. Bertin, and P. Abry, Critical moment definition and estimation, for finite size observation of log-exponential-power law random variables, Signal Proc. **92**, 2848 (2012).

[61] G. Ben Arous, L. V. Bogachev, and S. A. Molchanov, Limit theorems for sums of random exponentials, Probab. Theory Rel. Fields **132**, 579 (2005).

[62] A. Pohorille, C. Jarzynski, and C. Chipot, Good practices in free-energy calculations, J. Phys. Chem. B **114**, 10235 (2010).

[63] B. Efron, Bootstrap methods: Another look at the jackknife, Ann. Stat. **7**, 1 (1979).

[64] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Monographs on Statistics & Applied Probability (Chapman & Hall/CRC, New York, 1994).

[65] A. C. Davison and D. V. Hinkley, *Bootstrap Methods and Their Application*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, Cambridge, UK, 1997).

[66] J. A. Bucklew, *Introduction to Rare Event Simulation* (Springer, New York, 2004).

[67] S. Asmussen and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Stochastic Modelling and Applied Probability (Springer, New York, 2007).

[68] D. D. L. Minh, Free-energy reconstruction from experiments performed under different biasing programs, Phys. Rev. E **74**, 061120 (2006).

[69] D. D. L. Minh and A. B. Adib, Optimized Free Energies from Bidirectional Single-Molecule Force Spectroscopy, Phys. Rev. Lett. **100**, 180602 (2008).

[70] S. Vaikuntanathan and C. Jarzynski, Escorted Free Energy Simulations: Improving Convergence by Reducing Dissipation, Phys. Rev. Lett. **100**, 190601 (2008).

[71] C. Dellago and P. Bolhuis, Transition path sampling and other advanced simulation techniques for rare events, in *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, edited by C. Holm and K. Kremer, Advances in Polymer Science Vol. 221 (Springer, Berlin, 2009), pp. 167–233.