Numerical Estimation of Limiting Large-Deviation Rate Functions

Peter Werner* and Alexander K. Hartmann†
Institut für Physik, Universität Oldenburg, 26111 Oldenburg, Germany
(Dated: December 15, 2025)

For statistics of rare events in systems obeying a large-deviation principle, the rate function is a key quantity. When numerically estimating the rate function one is always restricted to finite system sizes. Thus, if the interest is in the limiting rate function for infinite system sizes, first, several system sizes have to be studied numerically. Here rare-event algorithms using biased ensembles give access to the low-probability region. Second, some kind of system-size extrapolation has to be performed.

Here we demonstrate how rare-event importance sampling schemes can be combined with multi-histogram reweighting. We study two ways of performing the system-size extrapolation, either directly acting on the empirical rate functions, or on the scaled cumulant generating functions, to obtain the infinite-size limit. The presented method is demonstrated for a binomial distributed variable, a Markov process of random bits and the largest connected component of Erdős-Rényi random graphs. Analytical solutions are available in all cases for direct comparison. It is observed in particular that phase transitions appearing in the biased ensembles can lead to systematic deviations from the true result.

I. INTRODUCTION

Let S be an extensive random variable of a system at scale $\mathcal{N}$ that is under consideration. The scale $\mathcal{N}$ could be the duration of a process or the number of individual system components. For a very simple example, let S be the number of rainy days in a region within a total time span of $\mathcal{N}$ days. A corresponding intensive quantity is then $s = S/\mathcal{N}$, i.e. the fraction of rainy days. Since the number of days varies with time, a stochastic modeling is natural, described by a probability distribution $p(s;\mathcal{N})$. Often such distributions follow the *large-deviation principle* [1–3], which means that there exists a so-called *rate function* $I(s)$ such that

$$p(s;\mathcal{N}) \asymp e^{-\mathcal{N} I(s)} \,. \tag{1}$$

This is equivalent to the existence of the limit

$$\lim_{\mathcal{S}\to\infty} -\frac{1}{\mathcal{N}} \ln p(s;\mathcal{N}) = I(s). \tag{2}$$

There exists an analytical solution for a simple model of the rainy day example, as will be discussed below, but for more complex systems numerical methods are a viable alternative. Here many different systems have been investigated in the physics community like random graphs [4–6], exclusion processes [7, 8], the contact process [9], spin glasses [10], Brownian [11, 12] or fractional Brownian motion [13–15], traffic models [16], the Kardar-Parisi-Zhang equation [17–19], chaotic maps [20], non-equilibrium work processes [21–23], and many more. Here one has to take into account that the probability density $p(s;\mathcal{N})$ following Eq. (1) becomes exponentially small in $\mathcal{N}$ almost everywhere. Therefore, in ordder to obtain a good estimate of the rate function $I(s)$, rare-event samples have to be obtained from the distribution tails, which is a numerical challenge. For this purpose, various large-deviation algorithms and methods [24, 25] have been devised. One class of approaches is based on population dynamics, referred to as *cloning* or *splitting* [7–9, 26], with Adaptive Multilevel Splitting (AMS) [11, 27, 28] as a notable variant. In these, a set of systems is evolved simultaneously and at certain points systems are discarded and/or duplicated based on an importance function that biases the dynamics towards rare-events. Another approach, the adaptive power method [29], is centered around estimating the eigenvalue of the systems tilted transition matrix, which allows to get the rate function by means of the Gärtner-Ellis theorem [2]. In some methods, the maximum likelihood pathway of a rare event, the so called *instanton*, is numerically determined by mapping to a classical system, to gain insight into a systems large-deviation properties [30–32]. The final class of approaches are importance-sampling methods [5, 6, 21–23, 33]. Here, data points are sampled according to some Monte Carlo scheme but with respect to a tilted or *biased* distribution to obtain improved estimates in the original distributions tails. This later class of algorithms are of relevance here.

The purpose of this paper is to demonstrate how multi-histogram reweighting [34–38] can be effectively combined with importance-sampling large-deviation techniques and finite-size fitting extrapolation [9, 39, 40] to obtain a rate function estimate.

The paper is structured as follows: We start by explaining methods for estimating rate functions. Subsequently, three case studies are presented while applying the above mentioned approaches. First, a simple binomial distributed case is considered. Second, a Markov process of random bits. And third, we investigate the more complex case of the largest connected component of Erdős Rényi (ER) random graphs. For these investigations, the general behavior of the approaches in relation

* peter.werner@uni-oldenburg.de
† a.hartmann@uni-oldenburg.de

to their parameters are demonstrated. We conclude with an overarching discussion.

II. RATE FUNCTION ESTIMATION METHODS

A. Empirical Rate Function at a Large Scale

A naive way to estimate a rate function is to perform large-deviation simulations at a sufficiently large system scale $\mathcal{N}$, ideally chosen such that the empirical determined finite-size corrections are insignificant compared to the desired magnitude of statistical errors.

An estimate of the underlying probability distribution $p(s,\mathcal{N})$ for the quantity of interest in form of a histogram $h(s;\mathcal{N})$ is straightforward to calculate, see [21] for an example. From this, the empirical rate function is given by:

$$I(s;\mathcal{N}) = -\frac{\ln(h(s;\mathcal{N}))}{\mathcal{N}}. \tag{3}$$

This approach assumes that the chosen system size is "large enough" which can roughly be tested by performing at least some simulations for a larger size like $2\mathcal{N}$ and verifying that the results do not change as compared to the statistical error.

A minor but very common drawback of this method is that the histogram requires to select an appropriate binning, which can limit the spacial resolution in regions with less data samples. Statistical error propagation, on the other hand, is relatively easy accomplished with this method via simple Gaussian error propagation. The approach presented below is histogram-free and involves error propagation, too.

B. Direct Extrapolation of the Empirical Rate Function

An extension of the previous method is to obtain data for various system scales $\mathcal{N}$ and subsequently calculate the empirical rate function in Eq. (3) at each of them. Often one assumes that the finite-size effects behave like a power-law as a function of $\mathcal{N}$, which is very common for the finite-size behavior of physical systems, i.e.,

$$\begin{aligned} I(s;\mathcal{N}) =& \tilde{I}(s) + a(s)\mathcal{N}^{-b(s)} \\ & \text{with } b(s) \geq 0\,. \end{aligned} \tag{4}$$

This allows for a fit for selected values of s over the system scales $\mathcal{N}$, where $\tilde{I}$, a and b are s-dependent fit parameters. Thus, for every considered value of s an independent fit is performed. The extrapolated rate function is $\tilde{I}(s)$. There is no a-priory argument, why a power law is an appropriate choice in all cases. However, such shape of finite-size dependencies has been observed already in large-deviation contexts [9, 39, 41]. The discussion following Eq. (13) below will elaborate this briefly. When the measured quantity S only takes values from a finite set of elements that scales with $\mathcal{N}$, this approach comes with a further drawback. Since the empirical rate function is then evaluated at discrete points $s = S/\mathcal{N}$, too, which must be common at all system scales $\mathcal{N}$, the spatial resolution of the rate function estimate will be limited by the resolution at smallest system scale.

Finally, a benefit of extrapolation can sometimes be that it helps to reduce the overall computational effort, because smaller system sizes are generally easier to simulate than rather large ones. However, if finite size effects prevent a good extrapolation, i.e, the fitted function is not representative of the data, there is little to be gained (see for example the $q \approx 0.316$ curve in the lower panel of Figure 7).

C. SCGF Transformation with Multi-Histogram Reweighting

Instead of using Eq. (4) for a direct estimate of the rate function, the approach presented here is centered around the approximation of the *scaled cumulant generating function* (SCGF) that is given by

$$\Psi(q) := \lim_{\mathcal{N}\to\infty} \frac{1}{\mathcal{N}} \ln\left\langle e^{qS}\right\rangle, \tag{5}$$

where $S = \mathcal{N}s$ and $\langle \dots \rangle$ denotes the expectation value with respect to $p(S;\mathcal{N})$. From this estimate, the rate function is obtained via the Gärtner-Ellis theorem [2, 3], which states that the rate function is the Legendre-Fenchel (LF) transformation of the SCGF:

$$I(s) = \sup_q \left[qs - \Psi(q)\right]. \tag{6}$$

Note that the application of this transformation assumes that the underlying distribution follows the large-deviation principle Eq. (1). Furthermore, it will only yield the convex envelope of the rate function if the SCGF is non-convex [2]. There are benefits from taking this detour over the SCGF, compared to the direct methods. First, no histogram binning is required to calculate the expectation value in Eq. (5), which can just be approximated by an average of the data. Second, should the distributions undergo substantial changes when increasing the system scale $\mathcal{N}$, i.e., strong finite-size effects, Eq. (5) can be sometimes easier for extrapolateion by means of a fit than the empirical rate function in Eq. (3), see below.

The starting point is to perform numerical simulations of the system of interest, for several scales $\mathcal{N}$ and various suitably chosen values of the *bias parameter* q, which appears already in the definition (5) of the SCGF. The simulations generate *exponentially-biased* data series $(S_1^q, \dots, S_{N_q}^q)$ of N_q values of the system quantity S of interest. For convenience, the data points are assumed to be uncorrelated. Note that for simplicity of notation we

do not denote the system-scale dependence of the data series. Exponentially biased means here that for the given value q of the bias parameter the data points are obtained according the biased probability

$$\tilde{p}(S;\mathcal{N},q) = \frac{p(S;\mathcal{N})\exp(qS)}{\mathcal{Z}(q,\mathcal{N})}\,, \tag{7}$$

instead of the original distribution $p(S;\mathcal{N})$. Note that $\tilde{p}(S;\mathcal{N},q=0) = p(S;\mathcal{N})$ for the unbiased case. While the angular brackets in Eq. (5) could be evaluated by directly sampling from the original non-biased distribution $p(S;\mathcal{N})$ and then applying a weight of e^{qS}, the important point here is the observation that this method becomes increasingly inefficient as q deviates more and more from zero. A simple argument for this is that since $p(S;\mathcal{N})$ and $p(S;\mathcal{N})e^{qS}$ usually do not peak around the same value of S for q values other than zero, most of the directly generated samples do not contribute significantly to the expectation value $\left\langle e^{qS}\right\rangle$ anymore, i.e, they are suppressed by the exponential weight. To overcome this limitation, it is then more efficient to equivalently sample according to the biased distribution $\tilde{p}(S;\mathcal{N},q)$ with a weight of one for the individual samples. Such biased data can be generated using corresponding large-deviation algorithms [21, 24, 42, 43]. The exponential bias $\exp(qS)$ is chosen to explicitly match the functional form of the expectation value argument in Eq. (5) and is frequently used in other previous studies that employ large-deviation algorithms [21, 22]. In these studies a temperature like parameter θ is usually denoted in the exponential bias, which relates to the parameter q through $q = -1/\theta$. In an importance sampling context, this setup is also referred to as an exponential change of measure [3]. Since the present work is concerned with the data analysis and extrapolation rather than the actual simulations, we refer to the literature on large-deviation simulation algorithms. However, in order to give a complete picture of the method the large-deviation algorithm from [21] will be explained in section III B for the Markov process. The same algorithm is used for the random graph case study, too.

The next step is to estimate from the numerical results the SCGF at finite size $\mathcal{N}$, namely $\Psi(q;\mathcal{N})$, and the *tilted* expectation value $\mu(q;\mathcal{N}) := \Psi'(q;\mathcal{N})$, i.e., the derivative of Ψ with respect to q. The latter one is needed to conveniently perform the LF transform, see below. We start with $\mu(q;\mathcal{N})$ and obtain

$$\begin{aligned}\mu(q;\mathcal{N}) &:= \frac{\partial}{\partial q}\Psi(q;\mathcal{N}) &(8)\\ &= \frac{\left\langle Se^{qS}\right\rangle_{p(S;\mathcal{N})}}{\mathcal{N}\left\langle e^{qS}\right\rangle_{p(S;\mathcal{N})}}\\ &= \frac{\langle S\rangle_{\tilde{p}(S;\mathcal{N},q)}}{\mathcal{N}} &(9)\\ &= \langle s\rangle_{\tilde{p}(S;\mathcal{N},q)}\,,\end{aligned}$$

at finite system scales $\mathcal{N}$. Details of the calculations for Eq. (9) are shown in appendix A. The expectation values $\langle\dots\rangle_{\tilde{p}(S;\mathcal{N},q)}$ are with respect to the biased distribution in Eq. (7) with the bias parameter q.

Thus, $\mu(q;\mathcal{N})$ can be estimated for a given size $\mathcal{N}$ directly for any value of q where simulations have been performed. But large-deviation simulations can be computationally costly, therefore it is impractical to perform these for many values of q at various scales $\mathcal{N}$. If the number of values q is large, this would be tedious. To address this issue, we apply histogram reweighting, also often referred to as Ferrenberg-Swendsen reweighting [34–37], to obtain estimates at arbitrary values of q. Actually, in the form that is presented below, the approach is histogram-free, which is very flexible. To ensure that the reweighting procedure works correctly, the requirement arises that the spacing between bias values q during the simulations is sufficiently small and therefore the relevant distribution support is decently covered by the sampled data. However, this depends severely on the investigated model and has to be determined empirically in each case.

Let q_i with $i = 1,\dots,K$ be the bias values used within the simulations, i.e., N_{q_i} denotes the number of sampled data points for the i'th simulation. An expectation value of a quantity $\mathcal{O} = \mathcal{O}(S)$, i.e., an arbitrary function depending on S, obtained at arbitrary bias q is then given by [36]

$$\langle\mathcal{O}\rangle_{\tilde{p}(S;\mathcal{N},q)} = \sum_{i=1}^{K}\sum_{a=1}^{N_{q_i}} \frac{\mathcal{O}_a^{q_i} e^{qS_a^{q_i}+f_q}}{\sum_{k=1}^{K} N_{q_k} e^{q_k S_a^{q_i}+f_{q_k}}}\,, \tag{10}$$

where $\mathcal{O}_j^{q_i} = \mathcal{O}(S_j^{q_i})$. The normalization constant f_q is determined by considering $\mathcal{O} = 1$ such that $\langle\mathcal{O}\rangle_{\tilde{p}(S;\mathcal{N},q)} = 1$ and dividing by e^{f_q} results in

$$e^{-f_q} = \sum_{i=1}^{K}\sum_{a=1}^{N_{q_i}} \frac{e^{qS_a^{q_i}}}{\sum_{k=1}^{K} N_{q_k} e^{q_k S_a^{q_i}+f_{q_k}}}\,. \tag{11}$$

While the corresponding constants f_{q_k} in the denominator are determined self consistently by setting $f_q = f_{q_j}$ with $j = 1,\dots,K$ on the left side of Eq. (11). The values of f_{q_j} can be obtained by iteration, e.g., starting with $f_{q_j} = 0$ for all values of j or by more sophisticated approaches [38]. Note that the values of f_{q_j} can be determined only up to an additive but irrelevant constant.

Reweighting by means of Eq. (10) is performed over the data points directly, i.e., there is the additional benefit that it is a histogram-less method. Thus, no selection of a specific binning for the measured values of S is necessary. The calculation of the corresponding statistical error-propagation is given in appendix B, which, according to the knowledge of the authors, has not been published in the literature so far. For only unbiased data samples, a discussion of statistical errors can be found in [44] and for importance-sampling estimators in [33].

With the approach explained above, $\mu(q,\mathcal{N})$ can be obtained for basically any desired value of q. Thus, we turn now to the determination of $\Psi(q;\mathcal{N})$, which can be written as $\frac{1}{\mathcal{N}}\ln\left\langle e^{qS}\right\rangle_{p(S;\mathcal{N})}$. By comparing with Eq. (10)

we observe that the calculation of the average can be achieved by first setting $q = 0$, which also means that $f_q = f_0$ in the equation, and then using $\mathcal{O} = e^{qS}$, where now q has become a free parameter. The latter yields $\mathcal{O}_a^{q_i} = e^{qS_a^{q_i}}$, when applied to the data points. This results in

$$\left\langle e^{qS} \right\rangle_{p(S;\mathcal{N})} = \sum_{i=1}^{K} \sum_{a=1}^{N_{q_i}} \frac{e^{qS_a^{q_i}+f_0}}{\sum_{k=1}^{K} N_{q_k} e^{q_k S_a^{q_i}+f_{q_k}}} \,. \tag{12}$$

While this looks very similar to Eq. (10) for $\mathcal{O} \equiv 1$, it differs in the fact that f_0 appears in the argument of the exponential, instead of f_q. Since the values of f_{q_k} and f_0 are already known, by taking the logarithm and dividing by $\mathcal{N}$, $\Psi(s;\mathcal{N})$ is readily obtained. It should be mentioned that $\Psi(q;\mathcal{N})$ could also be determined from Eq. (8) using thermodynamic integration [3], which is not done here to keep the error-propagation calculations simple for the SCGF estimator.

Since there is data for various scales $\mathcal{N}$, next we want to extrapolate $\mu(q;\mathcal{N})$ and $\Psi(q;\mathcal{N})$ toward infinite scales, i.e., taking the limit $\mathcal{N} \to \infty$. For this purpose, for any choice of q, a function $g(\mathcal{N};q)$ is fitted. Here, again power laws

$$\begin{aligned} g(\mathcal{N};q) =& C(q) + A(q)\mathcal{N}^{-B(q)} \\ & \text{with } B(q) \geq 0 \end{aligned} \tag{13}$$

are used, where A, B, and C are q dependent fit-parameters. The extrapolated value is given by $C(q)$ and is different when extrapolating μ or Ψ. The same symbol is only used for simplicity here. While the true functional form usually depends on the finite-size behavior at bias parameter q of the considered system and is unknown in general, the power law in Eq. (13) is a robust choice. For example, there are cases where this form of scaling is known to apply in a large-deviation context: In a cloning approach, the corresponding SCGF estimator is shown to follow a power-law scaling [9, 39]. Note that for problems which are admissible to the cloning approach, the SCGF can be estimated directly from "growth factors" describing the evolution of the population, while our approach is based on Eq. (12), which is very general. Finite-size corrections to eigenvalue statistics of random matrices were investigated in [41], where leading order correction terms are found to follow a power-law, too.

The final step is to perform the LF transform in Eq. (6). Instead of fixing s and then searching for the value of q that maximizes the supremum argument, it is easier to do a parametric LF transformation [2, 29]: For maximizing with respect to q one takes the derivative of the supremum argument and equals it to zero, which gives $\Psi'(q) = s = \mu(q)$. Substituting back into Eq. (6) yields

$$I(\mu(q)) = q\mu(q) - \Psi(q). \tag{14}$$

This form of the rate function can be directly evaluated for any desired value of q at $\mu(q)$ and $\Psi(q)$, which are the extrapolated values for $\mathcal{N} \to \infty$ by means of fitting Eq. (13). Doing so for sufficiently many values of q, each time yielding $s = \mu(q)$, will give arbitrary dense resolution on the rate function $I(s)$. This is computationally cheap, since arbitrary values for the bias parameter q can be utilized. As long as the data samples cover the region of s that is relevant for the chosen bias q, no new simulations are necessary. For error-propagation, the correlation between $\mu(q)$ and $\Psi(q)$ in Eq. (14) is ignored.

III. CASE STUDIES

Three systems are analyzed. First, a slight variation of binomial distributed data. This is mostly intended as a proof of concept for the presented method. Second, a Markov process of random bits, also referred to as *Balanced Markov Bits* [2]. The model has a less trivial yet convex rate function than the binomial distributed data. Here, it is used to compare the performance of the different rate function estimation methods. The third system is the distribution of the largest connected component in ER random graphs, which has a rather complicated and non-convex rate function in the highly connected phase, see Eq. (23)). It serves to test the method's performance in a complex system scenario, for example in the presence of phase transitions, as will be shown. ER graphs have been subject to large-deviation studies in various context before [4–6, 29, 45].

A. Binomial Distributed Data

Holding on to the terminology of the introductory example[46] in order to be less abstract, the probability distribution of a number S of rainy days in a period of $\mathcal{N}$ days is assumed to be binomial distributed, i.e.,

$$p(S;\mathcal{N}) = \binom{\mathcal{N}}{S} r(\mathcal{N})^S (1 - r(\mathcal{N}))^{\mathcal{N}-S} \,. \tag{15}$$

This is a very strong simplification of real precipitation events that are usually correlated in time and can be affected by seasonal variations, too. Equivalently, also the fraction of rainy days $s = S/\mathcal{N}$ can be considered, which is done in the following. To introduce a more challenging scaling behavior, which leads to a size-dependence of the rate function, the rain probability $r(\mathcal{N})$ is set to explicitly have a system scale $\mathcal{N}$ dependence of the form

$$r(\mathcal{N}) := r_\infty + c\mathcal{N}^{-\gamma}, \tag{16}$$

where $0 \leq r_\infty \leq 1$, $\gamma > 0$ and c is a constant. The parameters have to be chosen such that $0 \leq r(\mathcal{N}) \leq 1$ for all values of $\mathcal{N}$. The specific choices $r_\infty = 0.25$, $c = 7.4$ and $\gamma = 1$ are used here, while the considered system scales are $\mathcal{N}_i = \lfloor 10 \cdot (1.25)^i \rfloor$, with $i \in [2, 3, \ldots, 17]$.

A plot of the probability distributions in Eq. (15) resulting from the scale dependency of $r(\mathcal{N})$ in Eq. (16) is

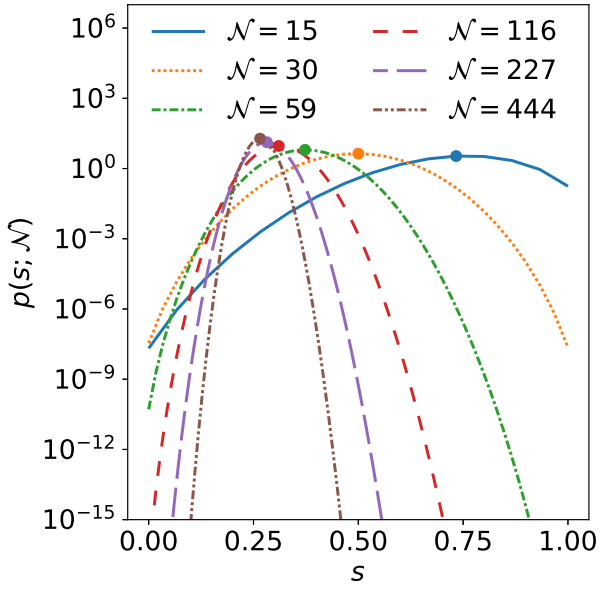

FIG. 1. Normalized binomial distributions with scale dependent parameter $r(\mathcal{N})$ as given in Eq. (16). In contrast to a binomial distribution with constant value of $r(\mathcal{N})$, the mean and variance change with the system scale $\mathcal{N}$. The circles mark the corresponding distribution maximum. Only a selection of system scales $\mathcal{N}$ is shown for visual purposes.

shown in figure 1 for various scales $\mathcal{N}$. The curves undergo significant change with increasing scale parameter, especially the peak of the distribution shifts from around $s \approx 0.76$ towards $s \approx r_\infty = 0.25$

The reason for making the parameter $r(\mathcal{N})$ explicitly dependent on the system scale $\mathcal{N}$ is that this causes the expectation value

$$\langle s \rangle_\mathcal{N} = \frac{\langle S \rangle}{\mathcal{N}} = r_\infty + c\mathcal{N}^{-\gamma} \tag{17}$$

and the SCGF

$$\Psi(q;\mathcal{N}) = \ln(1 - r(\mathcal{N})(1 - e^q)), \tag{18}$$

here given before taking the limit $\mathcal{N} \to \infty$, to be a function of the systems scale $\mathcal{N}$ as well. Finite-size behavior similar to this, which emerges for most complex systems naturally without including it explicitly, can make the estimation of rate functions difficult. The reason is that simply rescaling the data, i.e., dividing it with the scale $\mathcal{N}$, does not necessarily mitigate finite-size effects. For this particular case, it means that rescaling the data will result in a different estimate of the mean $\langle s \rangle_\mathcal{N}$ for every scale $\mathcal{N}$, in contrast to when $r(\mathcal{N})$ is constant.

The scaling behavior in this simple binomial example is known and generally follows a power law due to the specific choice for the parameter $r(\mathcal{N})$ in Eq. (16). For the expectation value $\langle s \rangle$, this can be seen directly in Eq. (17). The SCGF Eq. (18) by Talyor expansion around $q = 0$, approximately follows this behavior too, i.e.,

$$\Psi(q;\mathcal{N}) \approx qr(\mathcal{N}) = qr_\infty + qc\mathcal{N}^{-\gamma}\,. \tag{19}$$

Therefore, the extrapolation by means of Eq. (13) is well justified.

We tested the numerical simulation of biased data, subsequent determination of $\mu(q,\mathcal{N})$ and $\Psi(q;\mathcal{N})$, extrapolation to infinite $\mathcal{N} \to \infty$, and finally applying the LF to obtain the limiting rate function $I(S)$. For this purpose, 36 bias values q from the regime $q \in [-5,\ldots,5]$ were selected in a way that the value ranges close to the peak of the biased distributions cover the entire set of s values.

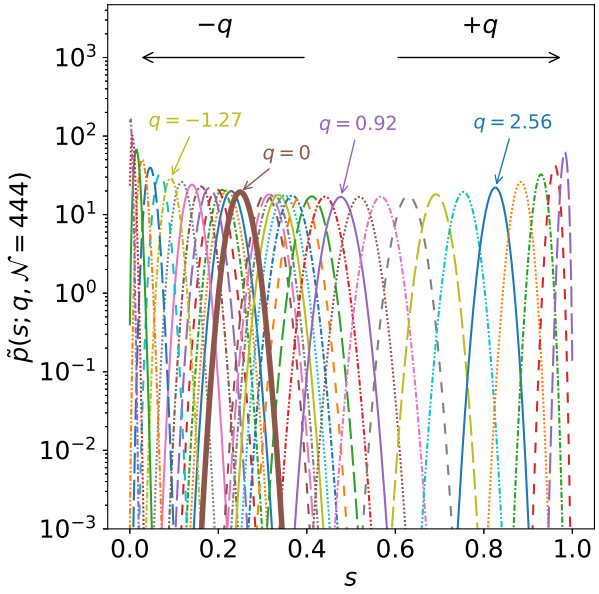

FIG. 2. Biased binomial distribution $\tilde{p}(s;\mathcal{N},q) \sim p(s;\mathcal{N})\exp(qs)$ at system scale $\mathcal{N} = 444$ for the various bias values q. The thicker curve shows the unbiased case, i.e., $q = 0$. Depending on the specific bias value q, the distributions are shifted towards higher or smaller values of the fraction of rainy days s. Each distributions peak region stretches over a different subset of s, such that in total the entire range of s is covered.

An example of the resulting biased probability distributions $\tilde{p}(s;\mathcal{N},q) \sim p(s;\mathcal{N})\exp(q\mathcal{N}s)$ is depicted in figure 2 for the largest system scale $\mathcal{N} = 444$, where the unbiased case, i.e. $q = 0$, is highlighted with a thicker line. Figure 2 illustrates that the biased distributions cover a regime of s values with high probability that otherwise only have very small values in the original unbiased distribution. Employing a large-deviation algorithm to sample realizations according to the biased distributions $\tilde{p}(s;\mathcal{N},q)$ is generally, and in particular at larger system scales, more efficient when numerically estimating the necessary expectation values with $q \neq 0$, e.g., in Eq. (5), compared to direct sampling at $q = 0$.

However, the system scales used for the binomial case are in fact so small that direct sampling is possible, i.e., calculating all probabilities $\tilde{p}(s;\mathcal{N},q)$ explicitly and then draw for each sample the number S of rainy days accordingly. For every bias value q, $N = 3 \cdot \mathcal{N}$ data points are generated, thus a very small number as compared to typical number of samples needed for statistical analyzes.

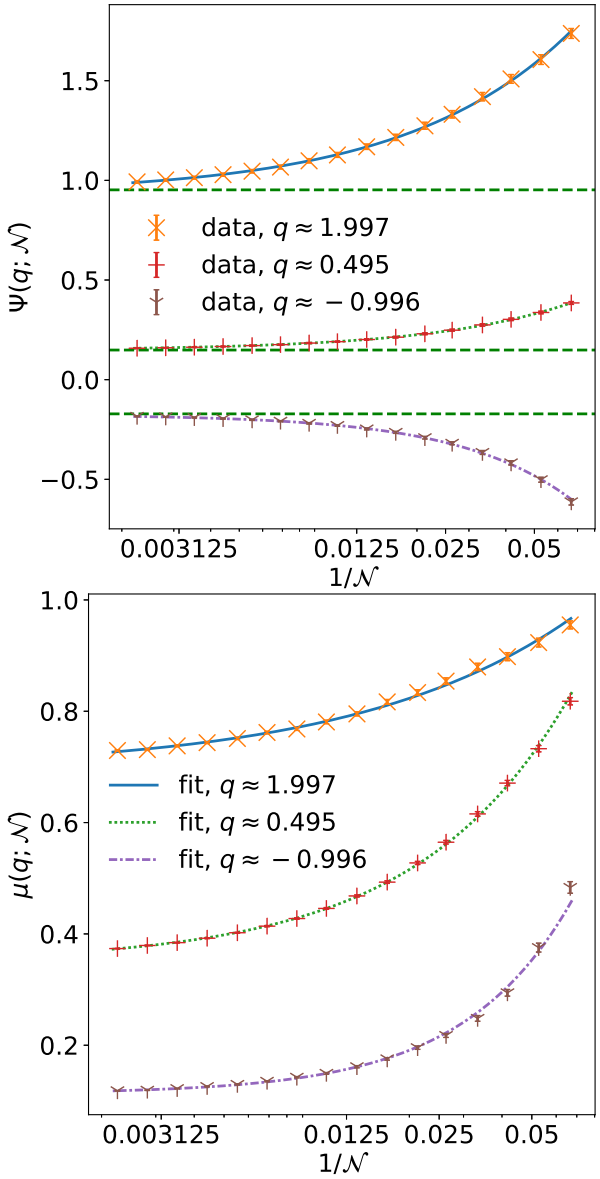

FIG. 3. Fit of the power-law in Eq. (13) for the binomial distributed data to extrapolate towards infinite system sizes $\mathcal{N}$. Curves for a selection of bias values q are shown together with the corresponding fits. Top: SCGF $\Psi(q;\mathcal{N})$. Also shown are the correct limiting values according to Eq. (18) as dashed horizontal green lines. Bottom: Biased expectation value $\mu(q;\mathcal{N})$. Fit parameters are summarized for both panels in Table I under Appendix C.

Exemplary extrapolation fits for the SCGF and the expectation value at various biases parameters q are shown in figure 3. The power law behavior is found to match the data decently.

In figure 4 the parametric LF transformation (see Eq. (14)) of the extrapolated SCGF is depicted. This is compared to the analytical rate function

$$I_{\mathrm{a}}(s) = s \ln \frac{s}{r_\infty} + (1-s) \ln \frac{1-s}{1-r_\infty}\,, \tag{20}$$

which is identical to the form when $r(\mathcal{N})$ is scale independent [46]. Note that due the parametric LF transformation also the horizontal axis for s exhibits error bars. Mostly, good agreement is found within errors for the rate function $I(s)$, while for s values towards one, a slight systematic deviation is visible There, the estimated result is higher than the analytical curve. This is somewhat expected, due to the fact that in this regime the power law behavior for the SCGF extrapolation (see Eq. (19)) becomes less accurate.

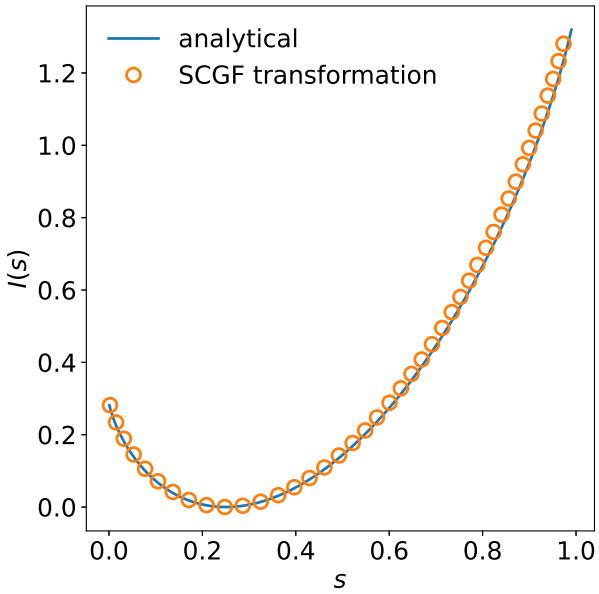

FIG. 4. Comparison of the exact analytical rate function in Eq. (20) and the rate function obtained from the parametric Legendre-Fenchel transformation (see Eq. (14)) of the numerically estimated SCGF for the number of rainy days. The procedure to obtain the estimated curve in described under subsection II C. Errorbars are smaller than symbol size.

B. Balanced Markov Bits

The second case study is about a two state Markov process with sequence $S = (S_1, S_2, \ldots, S_{\mathcal{N}})$ where $S_i \in \{0,1\}$ and $\mathcal{N}$ constitutes the length of the chain. If the system is currently in state S_i, then the probability of transitioning to the other state is given by $\alpha \in (0,1)$, which for the next state would mean $S_{i+1} = S_i + 1 \bmod 2$. Consequently, the probability of remaining in the current state, i.e., $S_{i+1} = S_i$, is $1-\alpha$. The inset in the top panel of figure 5 depicts the corresponding state graph of the Markov process. The initial state S_0 of the sequence is chosen randomly with uniform probability. The quantity of interest is the number of ones in the chain: $S = \sum_{i=1}^{\mathcal{N}} S_i$. Note that in contrast to the Bernoulli case, the fraction of ones always converges towards $s = \frac{S}{\mathcal{N}} \to 0.5$ as $\mathcal{N} \to \infty$.

Following the reasoning in section 4.3 of [2], the SCGF

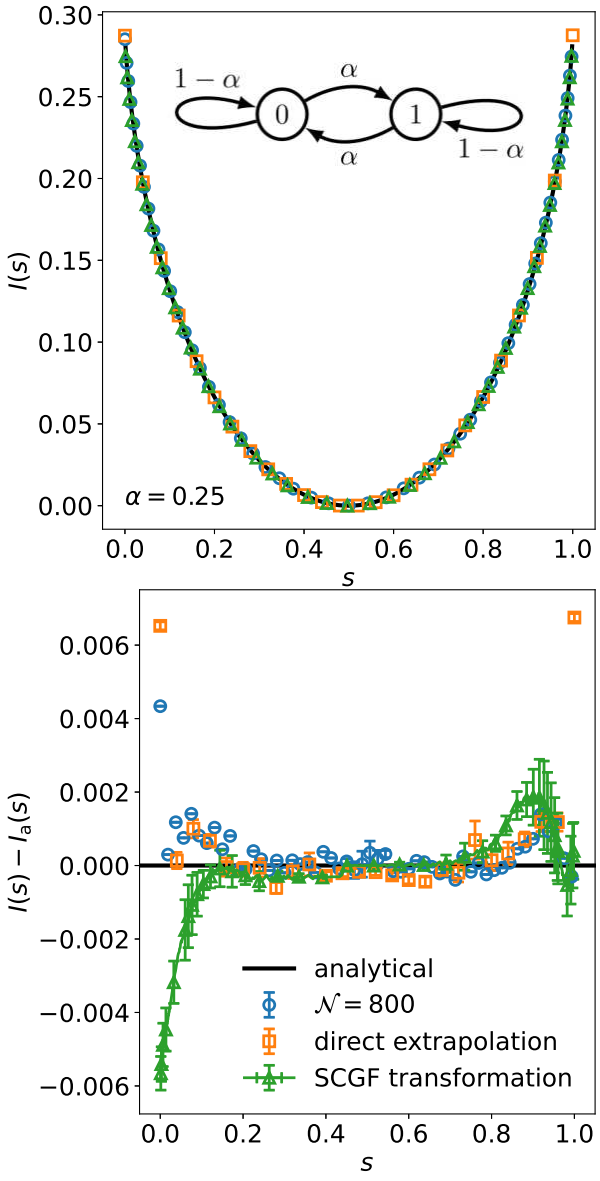

FIG. 5. Top: Comparison of the analytical rate function in Eq. (23), the Legendre-Fenchel transformed extrapolated SCGF (see subsection II C), the empirical rate function at system size $\mathcal{N} = 800$ (see subsection II A) and the directly extrapolated empirical rate function (see subsection II B) for the two state Markov process case study with transition probability $\alpha = 0.25$. The inset shows the state transition diagram of the Markov chain. Bottom: Corresponding residuals, highlighting the difference of the presented estimation methods to the exact analytical solution.

for the Balance Markov Bits is given by

$$\Psi(q) = \ln \zeta(\Pi_q), \tag{21}$$

where $\zeta(\Pi_q)$ is the dominant eigenvalue of the tilted transition matrix Π_q

$$\Pi_q = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha e^q & (1-\alpha)e^q \end{pmatrix}. \tag{22}$$

The corresponding rate function $I(s)$ can then be obtained via a parametric LF transformation of the SCGF $\Psi(q)$. For $\alpha = 0.5$, the limiting rate function is identical to that of the previous Bernoulli case study with $r_\infty = 0.5$.

For the numerical simulations, the transition probability $\alpha = 0.25$ and chain lengths $\mathcal{N} \in \{25, 50, 100, 200, 400, 800\}$ were considered. There are 25 bias values from the regime $q \in [-4.21, 4.21]$, explicitly including $q = 0$, for which up to 3810 independent samples were generated each.

The large-deviation sampling procedure then goes as follows: First, note that in standard simulations all random numbers are generated on demand by calling a pseudo random-number generator. Usually it is sufficient to generate $U(0,1)$ random numbers, i.e., uniformly distributed in $[0,1]$, since more complex distributions can be obtained by transformations [47]. To implement a large-deviation approach, instead of drawing the necessary random numbers on demand, they are computed in advance [48] and stored in a vector $\xi = (\xi_1, \xi_2, \ldots, \xi_\mathcal{N})$ with $\xi \in [0,1]$. The algorithm relies on the fact that the number S of ones within one realization of the Markov chain is a deterministic function of the random numbers ξ utilized during the simulation, i.e., $S = S(\xi)$. During step $i = 1, \ldots, \mathcal{N}$ of the simulation for example, the current state remains, if $\xi_i < \alpha$ and vice versa. Sampling according to the biased probability distribution $\tilde{p}(S; \mathcal{N}, q) \propto p(S(\xi); \mathcal{N}) \exp(qS(\xi))$ in Eq. (7) is then realized by treating the random number vector ξ as the state variable in another overarching Markov chain Monte Carlo (MC) simulation. In this overarching MC simulation, each MC step then consist of first redrawing with $U(0,1)$ uniform density a number of entries in ξ to obtain a trial state ξ'. For this trial state, the underlying Markov chain is simulated and the number of ones determined, i.e., $S' = S(\xi')$. Finally, the trial state is accepted using the corresponding Metropolis probability $p_{\text{Metr.}}(\xi \to \xi') = \min\{1, \exp(q(S' - S))\}$ for Eq. (7). For the number of entries that are redrawn in ξ during every MC step of the overarching MC simulation, one aims for an empirical acceptance probability close to 0.5. To ensure the equilibration of the Markov chain, multiple simulations can be started at different extremal initial values for the state ξ, i.e., all entries zero or all entries one. When the resulting values for S that are generated have become sufficiently similar between the multiple simulations, i.e., they have converged to each other, the Markov chains are likely sufficiently equilibrated, i.e., have "forgotten" their initial state.

The large-deviation rate function resulting from the various estimation methods are depicted in the top panel of figure 5 together with the analytical curve. While on first glance there seems to be good agreement with the analytical prediction, the residuals in the bottom panel, which show the difference to the analytical curve, reveal that there are small systematic deviations across all methods, especially towards the boundaries of s. While

the result for the rate function obtained via SCGF transformation (see method in subsection II C) seems to be rather prone to these systematic deviation near the definition boundaries, it attributes higher statistical uncertainty in these regions in contrast to the other methods. Furthermore, it has the advantage that it has in principle arbitrary spatial resolution along the s domain, since arbitrary bias values q can be considered. Please note that for visual purposes in figure 5, only a representative selection of points is depicted. The other methods are in this regard limited by either the smallest simulated system size for the direct extrapolation (see subsection II B), or the largest system size that can reasonably be simulated to obtain the empirical rate function (see subsection II A).

C. Largest Connected Component of Erdős-Rényi Random Graphs

As third case study, the largest connected component size of ER random graphs [4] is considered. Each graph consists of $\mathcal{N}$ nodes. For each pair i, j of nodes an edge is randomly and independently inserted with probability $c/(\mathcal{N}-1)$, where the connectivity c denotes the mean number of edges that are incident to a node. This graph ensemble exhibits a percolation transition at $c = 1$. For $c < 1$ the graph is composed of many small tree-like components, where for $c > 1$ there exists in the limit $\mathcal{N} \to \infty$ one large component, which is of the order of the system size.

The analytical rate function for the relative size of the largest connected component $s = S/\mathcal{N}$ is known [49]. It is given piecewise on intervals $[s_k, s_{k-1}] \subset [0, 1]$ which are arranged from right to left:

$$\begin{aligned} s_0 &= 1 \\ s_k &= \sup_s \left\{ \frac{s}{1-ks} = 1 - e^{-cs} \right\} \\ m(y) &= \log\left(1 - e^{-y}\right) \\ I_{\mathrm{a}}(s) &= -ksm(cs) + ks\log(s) + (1-ks)\log(1-ks) \\ &\quad + cks - k(k+1)cs^2/2 \\ &\quad \text{for} \quad s_k < s \leq s_{k-1}, \end{aligned} \tag{23}$$

Note that for $c \leq 1$, $s_1 = 0$, thus there is only one interval $[s_1, s_0]$, where for $c > 1$ in principle infinitely many intervals are needed to describe $I(s)$, but almost all of them accumulate near $s = 0$.

Here, the two representative connectivities $c \in \{0.5, 2.0\}$ for graphs of order $\mathcal{N} \in \{50, 75, 100, 150, 200, 300, 400, 800\}$ were investigated. We considered 34 bias values $q \in [-3.85, \ldots, 3.85]$, where one is always the unbiased case, i.e., $q = 0$. For sampling, the same methods as described in [5] are used, which involves the application of the large-deviation algorithm explained in the previous subsection. Only the actual model driven by the random vector ξ is different. Now the random numbers are used to generate a random graph for which the size S of the largest connected component is measured, i.e., $S = S(\xi)$ is again a deterministic function of the random numbers ξ. Up to 5912 independent data points were generated for each value of q.

For connectivity $c = 0.5$, the top panel of figure 6 shows the mean relative size of the largest connected component $\langle s \rangle$ as a function of the bias value q. The analytical curve is obtained numerically from the rate function in Eq. (23), which means first performing a LF transformation of $I(s)$ and then calculating the derivative, since $\Psi'(q) = \langle s \rangle$. Except for a region around $q \approx 0.3$, the analytical result is reproduced within error by the extrapolation via Eq. (13).

At this value of q, the biased ensemble for the smaller system sizes experiences a phase-transition like behavior as can be seen from the double peak structure in the corresponding probability distribution of the largest component size, which is shown in the bottom panel of figure 6. This causes a different scaling behavior, making the extrapolation by means of a power law inaccurate around this bias value regime (see figure 7). A similar break-down of scaling in presence of a phase transition was observed in [40], too.

In figure 8, the exact rate function, the LF transformed size-extrapolated SCGF (section II C), the empirical rate function Eq. (3) for the largest graph at $\mathcal{N} = 800$ (section II A) and the directly extrapolated rate function (section II B) are shown for $c = 0.5$ together with the corresponding residuals (bottom row). To asses finite-size effects, especially in the region around $s \approx 0.25$, the analysis is done twice: First, incorporating all simulated graph sizes (left column in figure 8) and second only system scales $\mathcal{N} \geq 200$ (right column). Around $s \approx 0.25$ the directly extrapolated curve shows rather large deviations from the analytical result, which could not be alleviated by only considering $\mathcal{N} \geq 200$ system sizes. Also the transformed SCGF curve fails to reliably produce estimates in this region, i.e., there is an unexpected gap in the parametric LF transformation instead of a continuous line. The empirical rate function at $\mathcal{N} = 800$ does not suffer from finite size effects in this region of s. Generally, the empirical rate function and the directly estimated rate function are systematically above or below the analytical result. In contrast, the transformed SCGF is overall closer to the exact result, however compared to the estimated errors, there are still systematic deviations resulting from the accuracy of the extrapolation fits. Except for these caveats, all methods are able to get decently close to the analytical result.

For connectivity $c = 2.0$, the rate function has concave regions and its minimum is no longer located at $s = 0$, as can be seen in figure 9. Both, the analytical solution and empirical rate functions (section II A) at various scales $\mathcal{N}$ are depicted. Convergence of the empirical rate function towards the true curve is observed for increasing graph order $\mathcal{N}$, while the convergence is slower for smaller val-

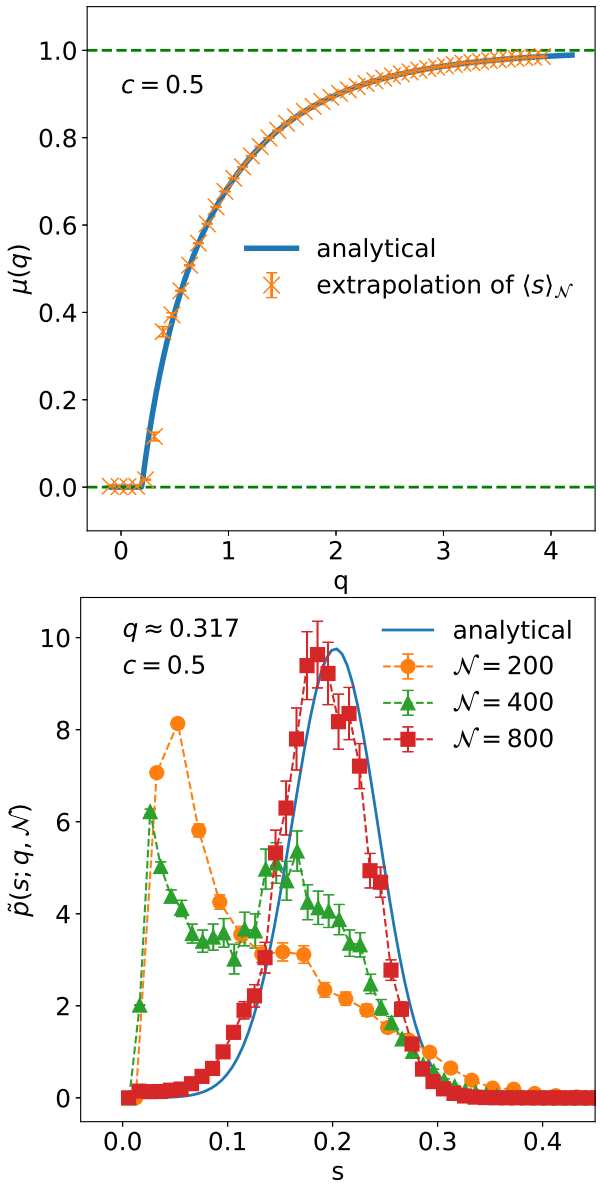

FIG. 6. Top: For connectivity $c = 0.5$, analytical and extrapolated relative size $\mu(q)$ of the larges connected component of the Erdős Rényi random graph ensemble. The extrapolation is according to Eq. (13). The dashed horizontal lines indicate the limiting values of the expectation value $\mu(q)$ for $q \to \pm\infty$. The gap at $q \approx 0.3$ for the extrapolated curve is caused by a an apparent first-order phase transition in the corresponding biased ensemble, resulting in deviations from the assumed power law scaling behavior. Bottom: Biased distribution of the largest connected component relative size at $q \approx 0.317$. The apparent phase transition is indicated by the double peak structure that vanishes for increasing graph order $\mathcal{N}$. For comparison, the analytical curve is shown, too and which is given by $\tilde{p}(s; q, \mathcal{N}) \sim \exp(-800\, I(s)) \exp(800\, q\, s)$ with rate function $I(s)$ from Eq. (23). Lines are a guide to the eye only.

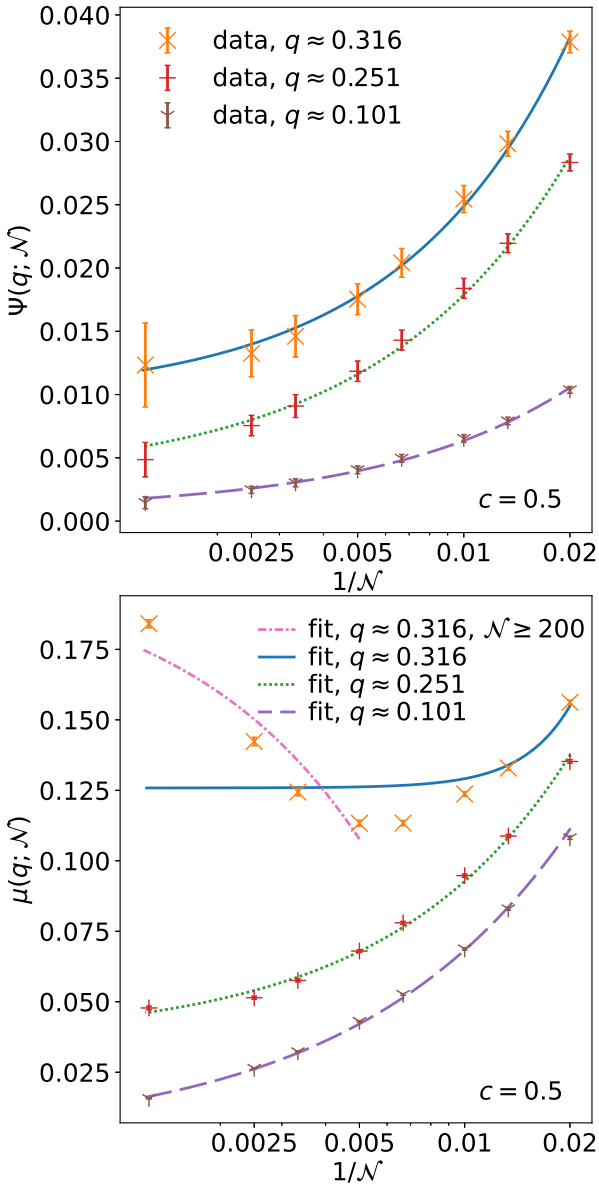

FIG. 7. Infinite system size $\mathcal{N}$ extrapolation fit by means of the power-law in Eq. (13) for the largest connected component in the Erdős-Rényi random graph ensemble with connectivity $c = 0.5$. Curves for a selection of bias values q are shown together with the corresponding fits. Top: SCGF $\Psi(q; \mathcal{N})$. Bottom: Biased expectation value $\mu(q; \mathcal{N})$. The scaling behavior for $q \approx 3.17$ shows substantial deviations from that of a power law. Also shown is a fit at $q \approx 3.17$ only including data values with $\mathcal{N} \geq 200$. Fit parameters are summarized for both panels in Table II under Appendix C.

ues of s.

A comparison of the rate function estimates for $c = 2.0$ from all methods is shown in figure 10. The residuals in the bottom panel show that all methods are able to decently reproduce the analytical curve for values $s \geq 0.75$. For the empirical rate function with $\mathcal{N} = 800$ (section II A) and the directly extrapolated curve (section II B)

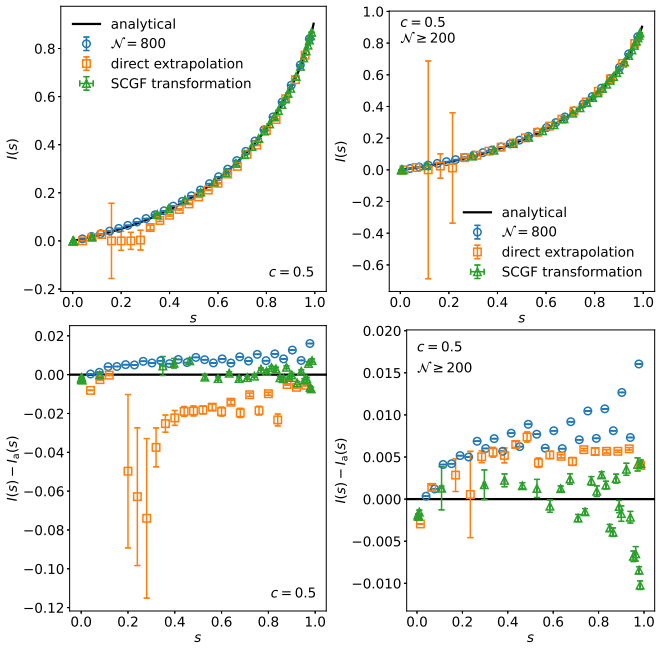

FIG. 8. The left column shows analysis results that include data points at all simulated system scales, while the right column only considers graph sizes where $\mathcal{N} \geq 200$. Top row: Comparison of the analytical rate function in Eq. (23), the Legendre-Fenchel transformed extrapolated SCGF (see subsection II C), the empirical rate function at system size $\mathcal{N} = 800$ (see subsection II A) and the directly extrapolated empirical rate function (see subsection II B) for the Erdős Rényi random graph ensemble. The connectivity is $c = 0.5$. The gap in the transformed SCGF curve at $s \approx 0.25$ is due to the finite-size apparent first-order phase transition of the biased graph ensemble at this point and the resulting breakdown of the assumed scaling power law in Eq. (13) used for extrapolation. Bottom row: Corresponding residuals, highlighting the difference of the presented estimation methods to the exact analytical solution. For visual purposes, the data point with the largest error value is omitted in the left plot and likewise the two data points with the largest errors in the right plot.

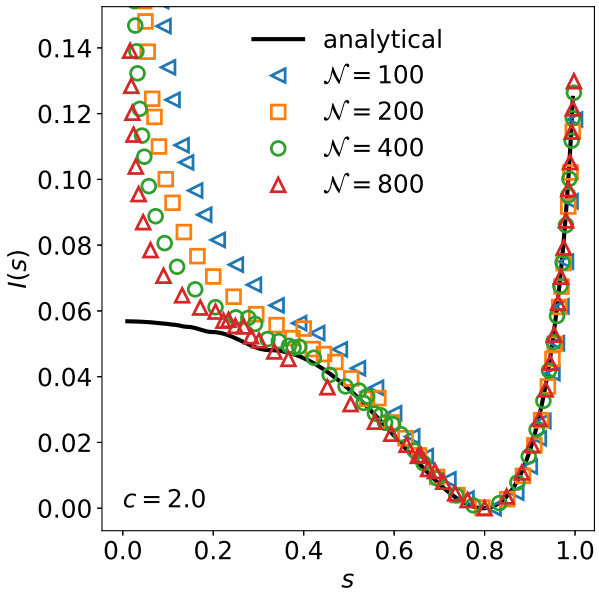

FIG. 9. For the Erdős Rényi random graph ensemble at connectivity $c = 2.0$, the analytical rate function in Eq. (23) and empirical rate functions at various system scales $\mathcal{N}$ (see method under subsection II A). While the analytical curve is already matched for values $s \geq 0.8$ by the curves at finite system sizes $\mathcal{N}$, the convergence at smaller values of s is slower.

this is event true for values $s \geq 0.6$. For small values of $s \approx 0.2$, the directly extrapolated results are closest to the analytical curve compared to the other methods. Due to the convexity of the rate function on certain intervals for $s < 0.75$, it is expected that the transformed SCGF curve (section II C) can not yield the analytical result but rather the convex envelope, which the method clearly failed to do.

IV. DISCUSSION

We presented an approach to estimate large-deviation rate functions. The method is based on biased sampling of the quantity of interest, multi-data histogram-less reweighting, followed by a finite-size extrapolation fit of the SCGF and the tilted expectation value, with a final application of the Legendre-Fenchel transform. Also an analysis of the resulting statistical errors was performed, as shown in the appendix. The intended targets of the approach are models which exhibit strong finite-size behavior in the context of numerical simulations, specifically when applying importance-sampling approaches. A comparison to two other rate function estimation methods, namely the empirical rate function and its direct extrapolation, was performed, too.

Three example systems with known analytical solutions were considered: First, a modified binomial distributed variable, for which the rate function could be reliably reproduced. Second, a Markov process of random bits. Third, the distribution of the size of the largest connected component for ER graphs. For the low connectivity case, a finite-size first-order phase-transition like behavior in the biased system distribution changed the finite-size behavior such that the extrapolation fit had bad quality, which led to a gap in the estimated rate function. In the high connectivity phase, the rate function could be reproduced in the convex regions. Due to this non-convexity at very small values of the relative size of the largest connected component, which exhibited very slow convergence with increasing system scale, the method can not outperform other methods.

The comparison of the different rate function estimation procedures showed that they all usually show some systematic deviations of similar magnitude from the analytical result, except for cases where the extrapolation fit became inaccurate. There, the ordinary empirical rate function at a sufficiently large system size yields in com-

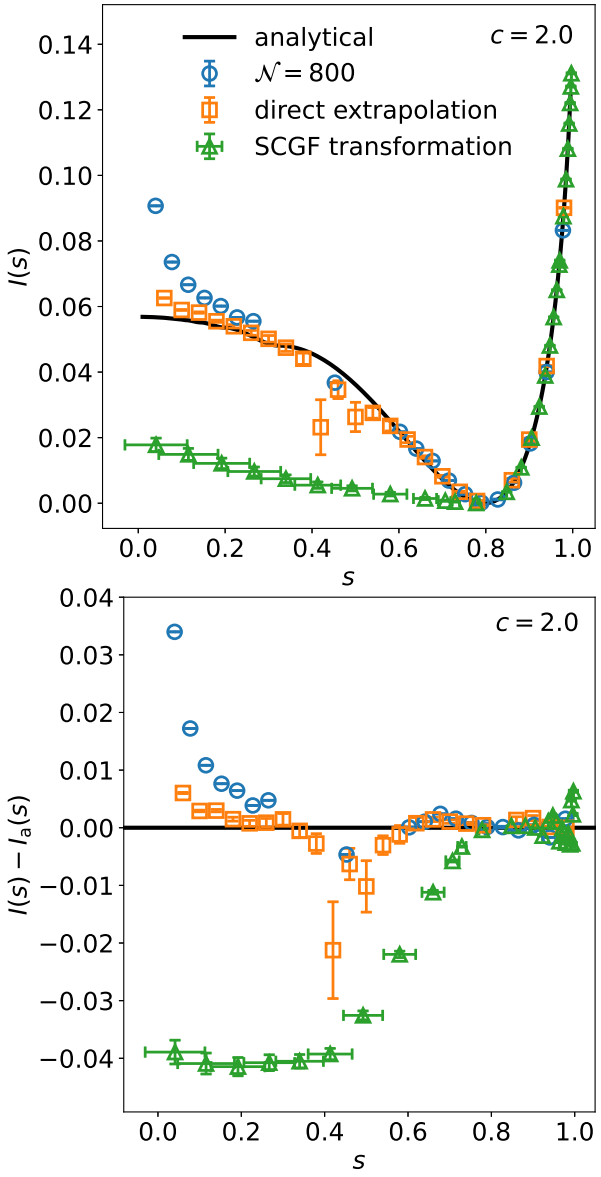

FIG. 10. top: Plot of the analytical rate function in Eq. (23), the Legendre-Fenchel transformed extrapolated SCGF (see subsection II C), the empirical rate function at system size $\mathcal{N} = 800$ (see subsection II A) and the directly extrapolated empirical rate function (see subsection II B) for the Erdős Rényi random graph ensemble. The connectivity is $c = 2.0$. The transformed SCGF only is located velow the convex envelope of the true rate function in the regime $s \approx 0.3 \ldots 0.75$. There the analytical rate function is partially concave. Bottom row: Corresponding residuals, highlighting the difference of the presented estimation methods to the exact analytical solution.

parison rather robust estimates.

It could be interesting to test different models with entirely convex rate functions, but strong finite-size effects, in the future. This could further reveal strengths and shortcomings of the method.

ACKNOWLEDGMENTS

The authors wish to thank Mylène Maïda for pointing out the useful reference [41] on finite-size corrections of the eigenvalue statistics of random matrices.

Appendix A: Tilted Expectation Value

Using the exponentially biased distribution

$$\tilde{p}(S;\mathcal{N},q) := \frac{e^{qS}}{\mathcal{Z}(q,\mathcal{N})} p(S;\mathcal{N}), \tag{A1}$$

where $\mathcal{Z}(q,\mathcal{N})$ is a normalization constant, the following relation

$$\begin{aligned} \left\langle S e^{qS} \right\rangle_{p(S;\mathcal{N})} &:= \int_{-\infty}^{\infty} S e^{qS} p(S;\mathcal{N}) dS \\ &= \mathcal{Z}(q,\mathcal{N}) \int_{-\infty}^{\infty} S \tilde{p}(S;\mathcal{N},q) dS \\ &= \mathcal{Z}(q,\mathcal{N}) \left\langle S \right\rangle_{\tilde{p}(S;\mathcal{N},q)} \end{aligned} \tag{A2}$$

is obtained. Inserting this result into the tilted expectation value yields

$$\begin{aligned} \mu(q;\mathcal{N}) &= \frac{\left\langle S e^{qS} \right\rangle_{p(S;\mathcal{N})}}{\mathcal{N} \left\langle e^{qS} \right\rangle_{p(S;\mathcal{N})}} \\ &= \frac{\mathcal{Z}(q,\mathcal{N}) \left\langle S \right\rangle_{\tilde{p}(S;\mathcal{N},q)}}{\mathcal{N} \mathcal{Z}(q,\mathcal{N}) \left\langle 1 \right\rangle_{\tilde{p}(S;\mathcal{N},q)}} \\ &= \frac{\left\langle S \right\rangle_{\tilde{p}(S;\mathcal{N},q)}}{\mathcal{N}}, \end{aligned} \tag{A3}$$

which is exactly Eq. (9).

Appendix B: Error-propagation in Multiple Histogram-Reweighting

The variance of the observable S simulated at bias value q_i is estimated by

$$(\sigma_S^{q_i})^2 = \frac{1}{N-1} \sum_{a=1}^{N_{q_i}} \left(S_a^{q_i} - \frac{1}{N_{q_i}} \sum_{b=1}^{N_{q_i}} S_b^{q_i} \right)^2. \tag{B1}$$

To obtain the corresponding variance for the expectation value μ and the SCGF Ψ, Gaussian error propagation is performed. This means, for a function $g(x_1,\ldots,x_n)$, where the random numbers $x_1,\ldots,x_n$ exhibit variances $\sigma_1^2,\ldots,\sigma_n^2$ and are uncorrelated, the total variance is estimated by $\sum_i \left(\frac{\partial g}{\partial x_i}\right)^2 \sigma_i^2$.

In both cases μ and Ψ, the derivative of the normalization constants f_q, defined in Eq. (11), with respect to

a particular data point $S_a^{q_l}$ $(a = 1, \ldots, N_{q_l})$ is required, which is given by

$$\frac{\partial f_q}{\partial S_a^{q_l}} = -\frac{q e^{q S_a^{q_l}+f_q}}{\sum_{i=1}^K N_i e^{q_i S_a^{q_l}+f_i}} + e^{q S_a^{q_l}+f_q} \frac{\sum_{j=1}^K N_{q_j} q_j e^{q_j S_a^{q_l}+f_{q_j}}}{\left(\sum_{i=1}^K N_{q_i} e^{q_i S_a^{q_l}+f_{q_i}}\right)^2} + \sum_{j=1}^K \sum_{k=1}^K \sum_{b=1}^{N_i} \frac{N_{q_j} e^{q_j S_b^{q_k}+f_{q_j}} e^{q S_b^{q_k}+f_q}}{\left(\sum_{i=1}^K N_{q_i} e^{q_i S_b^{q_k}+f_{q_i}}\right)^2} \frac{\partial f_{q_j}}{\partial S_a^{q_l}}. \quad \text{(B2)}$$

The necessary derivatives $\frac{\partial f_{q_j}}{\partial S_a^{q_l}}$ are determined by setting $f_q = f_{q_i}$. This turns Eq. (B2) into a set of linear equation that needs to be solved for every data point.

a. Expectation Value The expectation value through reweighting, see Eq. (10), at a fixed value of q is given by

$$\mu(q;\mathcal{N}) = \sum_{i=1}^K \sum_{a=1}^{N_{q_i}} \frac{S_a^{q_i} e^{q S_a^{q_i}+f_q}}{\sum_{k=1}^K N_{q_k} e^{q_k S_a^{q_i}+f_{q_k}}}, \quad \text{(B3)}$$

with the normalization constants f_{q_k} from Eq. (11). The corresponding variance is then calculated via

$$\left(\sigma_{\mu(q;\mathcal{N})}\right)^2 = \sum_{i=1}^K \sum_{a=1}^{N_{q_i}} \left(\frac{\partial \mu(q;\mathcal{N})}{\partial S_a^{q_i}}\right)^2 \left(\sigma_S^{q_i}\right)^2, \quad \text{(B4)}$$

where we naturally assume here and below that all sampled values $S_a^{q_i}$ exhibit the same variance $\left(\sigma_S^{q_i}\right)^2$ given by Eq. (B1). The needed partial derivatives read as

$$\frac{\partial \mu(q;\mathcal{N})}{\partial S_a^{q_l}} = \frac{(1+qS_a^{q_l}) e^{q S_a^{q_l}+f_q}}{\sum_{i=1}^K N_{q_i} e^{q_i S_a^{q_l}+f_{q_i}}} - S_a^{q_l} e^{q S_a^{q_l}+f_q} \frac{\sum_{j=1}^K N_{q_j} q_j e^{q_j S_a^{q_l}+f_{q_j}}}{\left(\sum_{i=1}^K N_{q_i} e^{q_i S_a^{q_l}+f_{q_i}}\right)^2} + \mu(q;\mathcal{N}) \frac{\partial f_q}{\partial S_a^{q_l}} - \sum_{j=1}^K \sum_{k=1}^K \sum_{b=1}^{N_{q_k}} \frac{N_{q_j} S_b^{q_k} e^{q_j S_b^{q_k}+f_{q_j}} e^{q S_b^{q_k}+f_q}}{\left(\sum_{i=1}^K N_{q_i} e^{q_i S_b^{q_k}+f_{q_i}}\right)^2} \frac{\partial f_{q_j}}{\partial S_a^{q_l}}. \quad \text{(B5)}$$

b. SCGF The error-propagation for the SCGF

$$\Psi(q;\mathcal{N}) := \frac{1}{\mathcal{N}} \ln \left\langle e^{qS} \right\rangle_{p(S;\mathcal{N})}, \quad \text{(B6)}$$

at a finite scale $\mathcal{N}$ and fixed bias parameter q, goes in a similar fashion. Resolving the angular brackets in Eq. (B6) through Eq. (10), as shown above, yields Eq. (12). The variance for the SCGF is then

$$\left(\sigma_{\Psi(q;\mathcal{N})}\right)^2 = \sum_{i=1}^K \sum_{a=1}^{N_{q_i}} \left(\frac{\partial \Psi(q;\mathcal{N})}{\partial S_a^{q_i}}\right)^2 \left(\sigma_S^{q_i}\right)^2 = \frac{1}{\left(\mathcal{N} \left\langle e^{qS} \right\rangle_{p(S;\mathcal{N})}\right)^2} \sum_{i=1}^K \sum_{a=1}^{N_{q_i}} \left(\frac{\partial \left\langle e^{qS} \right\rangle_{p(S;\mathcal{N})}}{\partial S_a^{q_i}}\right)^2 \left(\sigma_S^{q_i}\right)^2, \quad \text{(B7)}$$

with derivatives

$$\frac{\partial \left\langle e^{qS} \right\rangle_{p(S;\mathcal{N})}}{\partial S_a^{q_l}} = + \left\langle e^{qS} \right\rangle_{p(S;\mathcal{N})} \frac{\partial f_q}{\partial S_a^{q_l}} + \frac{q e^{q S_a^{q_l}+f_0}}{\sum_{i=1}^K N_{q_i} e^{q_i S_a^{q_l}+f_{q_i}}} - e^{q S_a^{q_l}+f_0} \frac{\sum_{j=1}^K N_{q_j} q_j e^{q_j S_a^{q_l}+f_{q_j}}}{\left(\sum_{i=1}^K N_{q_i} e^{q_i S_a^{q_l}+f_{q_i}}\right)^2} - \sum_{j=1}^K \sum_{k=1}^K \sum_{b=1}^{N_{q_k}} \frac{N_{q_j} e^{q_j S_b^{q_k}+f_{q_j}} e^{q S_b^{q_k}+f_0}}{\left(\sum_{i=1}^K N_{q_i} e^{q_i S_b^{q_k}+f_{q_i}}\right)^2} \frac{\partial f_{q_j}}{\partial S_a^{q_l}}. \quad \text{(B8)}$$

Appendix C: Fit Parameters for Figure 3 and Figure 7

The fit parameters are showon in Tables I and II.

TABLE I. Fit parameters C, A and B corresponding to Eq. (13) for the curves depicted in Figure 3. These belong to the case study for binomial distributed data.

$g(\mathcal{N};q)$	q	C	A	B
$\Psi(q;\mathcal{N})$	1.997	0.931(2)	6.71(6)	0.777(2)
	0.495	0.1473(2)	2.94(1)	0.928(1)
	-0.996	-0.1755(6)	-9.1(1)	1.133(3)
$\mu(q;\mathcal{N})$	1.997	0.694(1)	1.45(2)	0.617(4)
	0.495	0.3487(5)	5.16(4)	0.877(2)
	-0.996	0.1129(6)	9.5(2)	1.226(6)

[1] F. den Hollander, *Large Deviations* (American Mathematical Society, Providence, 2000).

[2] H. Touchette, Physics Reports **478**, 1 (2009).

[3] H. Touchette, "A basic introduction to large deviations: Theory, applications, simulations," (2012), arXiv:1106.4146 [cond-mat.stat-mech].

TABLE II. Fit parameters C, A and B corresponding to Eq. (13) for the curves depicted in Figure 7. These belong to the case study of the Erdős-Rényi random graph ensemble at connectivity $c = 0.5$.

$g(\mathcal{N};q)$	q	C	A	B
$\Psi(q;\mathcal{N})$	0.316	0.0097(3)	0.99(2)	0.908(4)
	0.251	0.0032(3)	0.59(2)	0.803(5)
	0.101	0.0008(1)	0.230(5)	0.807(5)
$\mu(q;\mathcal{N})$	0.316	0.13(1)	6366(5640)	3.1(5)
	0.251	0.037(1)	2.75(7)	0.845(6)
	0.101	0.0011(6)	1.78(3)	0.712(3)
$\mu(q;\mathcal{N})$ $\mathcal{N} \geq 200$	0.316	0.20(1)	-8(1)	0.84(3)

[4] A. Engel, R. Monasson, and A. K. Hartmann, Journal of Statistical Physics **117**, 387 (2004).
[5] A. K. Hartmann, Eur. Phys. J. B **84**, 627 (2011).
[6] A. K. Hartmann, The European Physical Journal Special Topics **226**, 567 (2017).
[7] C. Giardinà, J. Kurchan, and L. Peliti, Phys. Rev. Lett. **96**, 120603 (2006).
[8] V. Lecomte and J. Tailleur, Journal of Statistical Mechanics: Theory and Experiment **2007**, P03004 (2007).
[9] E. Guevara Hidalgo, T. Nemoto, and V. Lecomte, Phys. Rev. E **95**, 062134 (2017).
[10] M. Körner, H. G. Katzgraber, and A. K. Hartmann, J. Stat. Mech. **2006**, P04005 (2006).
[11] F. Cérou and A. Guyader, Stochastic Analysis and Applications **25**, 417 (2007), https://doi.org/10.1080/07362990601139628.
[12] T. Agranov, P. Zilber, N. R. Smith, T. Admon, Y. Roichman, and B. Meerson, Phys. Rev. Res. **2**, 013174 (2020).
[13] A. K. Hartmann, S. N. Majumdar, and A. Rosso, Phys. Rev. E **88**, 022119 (2013).
[14] A. K. Hartmann and B. Meerson, Phys. Rev. E **109**, 014146 (2024).
[15] N. R. Smith and S. N. Majumdar, Journal of Statistical Mechanics: Theory and Experiment **2022**, 053212 (2022).
[16] W. Staffeldt and A. K. Hartmann, Phys. Rev. E **100**, 062301 (2019).
[17] A. K. Hartmann, P. L. Doussal, S. N. Majumdar, A. Rosso, and G. Schehr, Europhys. Lett. **121**, 67004 (2018).
[18] A. K. Hartmann, B. Meerson, and P. Sasorov, Phys. Rev. Research **1**, 032043 (2019).
[19] A. K. Hartmann, A. Krajenbrink, and P. Le Doussal, Phys. Rev. E **101**, 012134 (2020).
[20] N. R. Smith, Phys. Rev. E **106**, L042202 (2022).
[21] A. K. Hartmann, Phys. Rev. E **89**, 052103 (2014).
[22] P. Werner and A. K. Hartmann, Phys. Rev. E **104**, 034407 (2021).
[23] P. Werner, A. K. Hartmann, and S. N. Majumdar, Phys. Rev. E **110**, 024115 (2024).
[24] J. A. Bucklew, *Introduction to rare event simulation* (Springer-Verlag, New York, 2004).
[25] F. Bouchet, J. Rolland, and J. Wouters, Chaos **29**, 080402 (2019).
[26] J. Tailleur and V. Lecomte, AIP Conference Proceedings **1091**, 212 (2009), https://pubs.aip.org/aip/acp/article-pdf/1091/1/212/11954017/212_1_online.pdf.
[27] C.-E. Bréhier, Latin American journal of probability and mathematical statistics **12** (2015).
[28] F. Cérou, A. Guyader, and M. Rousset, Chaos: An Interdisciplinary Journal of Nonlinear Science **29**, 043108 (2019), https://pubs.aip.org/aip/cha/article-pdf/doi/10.1063/1.5082247/14620716/043108_1_online.pdf.
[29] F. Coghi and H. Touchette, Phys. Rev. E **107**, 034137 (2023).
[30] C. Hartmann, O. Kebiri, L. Neureither, and L. Richter, Chaos: An Interdisciplinary Journal of Nonlinear Science **29**, 063107 (2019), https://pubs.aip.org/aip/cha/article-pdf/doi/10.1063/1.5090271/13534861/063107_1_online.pdf.
[31] T. Grafke and E. Vanden-Eijnden, Chaos **29**, 063118 (2019), https://pubs.aip.org/aip/cha/article-pdf/doi/10.1063/1.5084025/13536944/063118_1_online.pdf.
[32] M. Alqahtani and T. Grafke, Journal of Physics A: Mathematical and Theoretical **54**, 175001 (2021).
[33] P. Glasserman and Y. Wang, The Annals of Applied Probability **7**, 731 (1997).
[34] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).
[35] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[36] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comp. Chem. **13**, 1011 (1992).
[37] A. M. Ferrenberg, D. P. Landau, and R. H. Swendsen, Phys. Rev. E **51**, 5092 (1995).
[38] T. Bereau and R. H. Swendsen, J. Comp. Phys. **228**, 6119 (2009).
[39] T. Nemoto, E. Guevara Hidalgo, and V. Lecomte, Phys. Rev. E **95**, 012102 (2017).
[40] E. G. Hidalgo, Journal of Statistical Mechanics: Theory and Experiment **2018**, 083211 (2018).
[41] A. Edelman, A. Guionnet, and S. Péché, The Annals of Applied Probability **26**, 1659 (2016).
[42] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Ann. Rev. Phys. Chem. **53**, 291 (2002), pMID: 11972010.
[43] A. K. Hartmann, Phys. Rev. E **65**, 056102 (2002).
[44] C. M. Rohwer, F. Angeletti, and H. Touchette, Phys. Rev. E **92**, 052104 (2015).
[45] S. Bhamidi, J. Hannig, C. Y. Lee, and J. Nolen, Electronic Journal of Probability **20**, 1 (2015).
[46] L. Peliti and S. Pigolotti, *Stochastic Thermodynamics: An Introduction* (Princeton University Press, 2021).
[47] A. K. Hartmann, *Big Practical Guide to Computer Simulations* (World Scientific, Singapore, 2015).
[48] G. E. Crooks and D. Chandler, Phys. Rev. E **64**, 026109 (2001).
[49] N. O'Connell, Probability Theory and Related Fields **110**, 277 (1998).