

# Large Deviations of the Activity in the Contact Process: Finite-Time and Finite- $N_c$ Scalings in the Large- $L$ Limit

Esteban Guevara Hidalgo<sup>1,2,\*</sup>

<sup>1</sup>*Institut Jacques Monod, CNRS UMR 7592, Université Paris Diderot, Sorbonne Paris Cité, F-750205, Paris, France*

<sup>2</sup>*Laboratoire de Probabilités et Modèles Aléatoires, Sorbonne Paris Cité,  
UMR 7599 CNRS, Université Paris Diderot, 75013 Paris, France*

In a recent paper, we studied the finite-time and finite- $N_c$  (population size) scalings of a large deviation function estimator (LDF) by means of the cloning algorithm. Its convergence-speed was used in order to extract its asymptotic behavior in the infinite- $t$  and infinite- $N_c$  limits. For the cases analyzed, this limit resulted to render a better LDF estimation in comparison to the standard estimator. The approach was based on the fact that the scalings of the systematic errors of the estimator behave as  $t^{-1}$  and  $N_c^{-1}$  in the large- $t$  and large- $N_c$  asymptotics. However, the validity of these scalings and the efficiency of the method proposed were proved only in cases for which the number of sites  $L$  (where the dynamics occurs) was small. In this paper, the analysis is extended to a wider range of system sizes. In order to characterize the large- $L$  behavior, we introduce the  $t^{-\gamma_t}$ - and  $N_c^{-\gamma_{N_c}}$ -scalings for the LDF. We show the dependence of the exponents  $\gamma_t$  and  $\gamma_{N_c}$  with  $L$  and how the configuration where  $\gamma_t \approx 1$  and  $\gamma_{N_c} \approx 1$  is restricted to specific regions of the plane  $s - L$ .

PACS numbers: 05.40.-a, 05.10.-a, 05.70.Ln

## I. INTRODUCTION

In order to study the properties of rare events and rare trajectories in stochastic dynamics, a large variety of methods have been developed [1–3]. The numerical approaches range from importance sampling [4], to “go with the winner” algorithms [5, 6], adaptive multilevel splitting [7] and transition path sampling [8]. Through this paper we will give particular attention to the population dynamics algorithms [9–11]. Under this approach, the study of rare trajectories in a system is done by exponentially biasing their probability. The resulting modified dynamics consists in the coupled evolution of a large number of copies  $N_c$  of the original process supplemented with a selection rule according to which a copy of the system is multiplied if it is rare or killed, if it is not. This selection-mutation mechanism favors the occurrence of the atypical trajectories of interest which have become typical under the biased dynamics.

The distribution of the class of rare trajectories in the original dynamics is related with the exponential growth of the population of clones of the system and LDF can be estimated from its growth rate. The numerical determination of this estimator is systematized in a method known as the cloning algorithm which can be performed in a number of ways [2, 9–14]. However, this method introduces two additional parameters into consideration: the population size  $N_c$  and the (simulation) time  $t$ . Both of which affect considerably the accuracy of the LDF estimation which is expected to be high in the infinite- $t$  and infinite- $N_c$  limit. Given that this is not achievable in practice, what is generally done is to choose these parameters large enough such that the average estimator

(over several realizations) does not depend on them.

The finite- $t$  and finite- $N_c$  scalings of the LDF were analyzed recently following two different approaches: an analytical one, in Ref. [13], using a discrete-time version of the population dynamics algorithm [9], and a numerical one, in Ref. [14], using a continuous-time version [10, 11]. The systematic errors of these scalings were found to behave as  $1/t$  and  $1/N_c$  in the large- $t$  and large- $N_c$  asymptotics respectively. Moreover, these scaling properties can be used in order to improve the LDF estimation (as shown in Ref. [14]). This is done considering that the asymptotic behavior of the estimator in the  $t \rightarrow \infty$  and  $N_c \rightarrow \infty$  limits may be interpolated from the data obtained from simulations at finite (and relative small)  $t$  and  $N_c$ . The improvement in the LDF estimation was illustrated on a simple two-states annihilation-creation dynamics (in one site) and on a more complex system, a contact process [15] (with  $L = 6$  sites). However, the validity of these scalings as well as the efficacy of the method as the number of sites  $L$  increases was leaved as pending. This is precisely the purpose of this paper where we complement the problem introduced in Ref. [14] by extending the analysis of the scaling of the LDF to a large- $L$  contact process.

The paper is organized as follows. In Sec. II we introduce the method used in order to estimate large deviations of additive observables. The finite-time and finite- $N_c$  scalings of the LDF are summarized in Sec. III A, and generalized to large- $L$  systems in Sec. IV. We make use of these results in Sec. V where we check the validity of the  $t^{-1}$ - and  $N_c^{-1}$ -scalings (Sec. V A), their behavior in the  $s$ -modified dynamics (Sec. V B) as well as the applicability of the scaling method (Sec. V C) for a contact process with  $L = 100$  sites. This analysis is generalized in Sec. VI where we characterize the finite- $t$  and finite- $N_c$  scalings of the LDF in the plane  $s - L$ . Finally, we present our conclusions in Sec. VII.

\* [esteban\\_guevarah@hotmail.com](mailto:esteban_guevarah@hotmail.com)

## II. LARGE DEVIATIONS OF ADDITIVE OBSERVABLES

In order to analyze the large deviations of the activity in the contact process we will make use of the continuous-time version of the “cloning algorithm” [2, 10, 11]. This approach allows to obtain the LDF from the exponential growth (or decay) rate of a set of copies of a system which evolves following a “ $s$ -modified dynamics”. The procedure followed in order to obtain this estimator is summarized below.

### A. Biased Dynamics and the Cloning Algorithm

We consider a general Markov dynamics which evolves continuously in time. The system jumps from configuration  $C$  to  $C'$  with transition rate  $W(C \rightarrow C')$ . The probability  $P(C, t)$  to find the system at time  $t$  in configuration  $C$  verifies the master equation

$$\partial_t P(C, t) = \sum_{C' \neq C} W(C' \rightarrow C) P(C', t) - r(C) P(C, t), \quad (1)$$

where  $r(C) = \sum_{C' \neq C} W(C \rightarrow C')$  is the escape rate from configuration  $C$ .

A trajectory of  $K$  configurations jumps,  $(C_0, \dots, C_K)$ , can be characterized by some additive observable  $\mathcal{O}$  (extensive in time) which is defined as

$$\mathcal{O} = \sum_{k=1}^{K-1} a(C_k, C_{k+1}) + \int_0^t dt' b(C(t')), \quad (2)$$

where  $a$  and  $b$  describe elementary increments. In that case, the joint distribution  $P(C, \mathcal{O}, t)$  describes the probability of finding the system in the configuration  $C$ , with a value of the observable  $\mathcal{O}$ , and at time  $t$ .

The procedure followed in order to analyze the large deviations of these observables, consists in biasing the statistical weight of histories of the system by a parameter  $s$  [10, 11]. A value of  $s$  different from 0 favors the non-typical values of the observable  $\mathcal{O}$  whose average value has been fixed. The large-time limit of the cumulants of  $\mathcal{O}$  can be recovered from the growth rate  $\psi(s)$  of the dynamical partition function  $Z(s, t) = \langle e^{-s\mathcal{O}} \rangle \sim e^{t\psi(s)}$  for  $t \rightarrow \infty$ , as derivatives of  $\psi(s)$  in  $s = 0$  [1]. This exponent  $\psi(s)$  is called scaled cumulant generating function (CGF) and its Legendre transform, large deviation function.

The CGF can be estimated numerically taking into consideration the relation between the dynamical partition function

$$Z(s, t) = \sum_C \hat{P}(C, s, t), \quad (3)$$

and the Laplace transform of the distribution  $P(C, \mathcal{O}, t)$

$$\hat{P}(C, s, t) = \int d\mathcal{O} e^{-s\mathcal{O}} P(C, \mathcal{O}, t). \quad (4)$$

Thus, the original dynamics have been transformed into a “ $s$ -modified” one, which verifies the time-evolution equation [16]

$$\partial_t \hat{P}(C, s, t) = \sum_{C' \neq C} (\mathbb{W}_s)_{C'C} \hat{P}(C', s, t) + \delta r_s(C) \hat{P}(C, s, t), \quad (5)$$

where  $(\mathbb{W}_s)_{C'C} = W_s(C' \rightarrow C) - r_s(C) \delta_{CC'}$ , and  $\delta r_s(C) = r_s(C) - r(C) - sb(C)$ . The expression

$$W_s(C \rightarrow C') = e^{-sa(C, C')} W(C \rightarrow C'), \quad (6)$$

represent a  $s$ -modified transition rate, whereas

$$r_s(C) = \sum_{C'} W_s(C \rightarrow C') \quad (7)$$

is the corresponding escape rate. Equation (5) can be interpreted as a population dynamics on a large number  $N_c$  of copies of the system which evolves in a coupled way with transition rates  $W_s(C \rightarrow C')$  and with a selection mechanism of rates  $\delta r_s(C)$  [9, 17]. Depending on  $\delta r_s(C)$  a copy of the system is multiplied or killed, so that under this  $s$ -biased dynamics an atypical class of histories of the original process become typical. The CGF estimator, that we will denote as  $\Psi_s^{(N_c)}$ , is then obtained from the exponential growth (or decay) rate of a population of copies of the system evolving with these rules. The method which systematize the numerical determination of this estimator is known as cloning algorithm which can be performed in a number of ways [2, 9–13]. A detailed description of the version used through this paper can be found in Sec. II.C.1. of Ref. [14].

### B. CGF Numerical Estimator

Using the constant-population approach of the continuous-time cloning algorithm on a  $s$ -biased Markov dynamics, the average over  $R$  realizations of the CGF estimator  $\Psi_s^{(N_c)}$  for  $N_c$  clones (or copies of the system) and a final simulation time  $T$ , is defined as

$$\overline{\Psi_s^{(N_c)}} = \frac{1}{R} \sum_{r=1}^R \frac{1}{t_r^{\mathcal{F}}} \log \prod_{i=1}^{\mathcal{K}_r} X_i^r, \quad (8)$$

where  $\mathcal{K}$  is the total number of configuration changes in the full population up to time  $t_r^{\mathcal{F}} \lesssim T$  which is the actual final simulation time [12]. However, as  $|\overline{\Psi_s^{(N_c)}}(T) - \overline{\Psi_s^{(N_c)}}(t_r^{\mathcal{F}})| \approx 0$  for  $t_r^{\mathcal{F}} \gg 0$ , it is possible to set  $t_r^{\mathcal{F}} \approx T$  in Eq. (8). At each configuration change, the population of clones is increased by a factor  $X_i^r$  where  $r \in \{1, \dots, R\}$ .

It is expected that in the infinite- $t$  and infinite- $N_c$  limits, Eq. (8) provides an accurate estimation of the CGF, i.e.,

$$\lim_{N_c \rightarrow \infty} \lim_{t \rightarrow \infty} \overline{\Psi_s^{(N_c)}}(t) = \psi(s). \quad (9)$$

However, as these limits are not achievable in practice, the best estimation of the large deviation function can be obtained considering a large enough simulation time  $T$  and number of clones  $N_c$ . The dependence of the estimator with these parameters is summarized in Sec. III A.

### C. Contact Process

The process of interest throughout this paper consists in a one-dimensional lattice with  $L$  sites and periodic boundary conditions. Each site  $i$  is occupied by a spin which can be in two possible states,  $n_i = 0$  or  $n_i = 1$ . The transition rates between states are given by

$$W(n_i = 1 \rightarrow n_i = 0) = 1, \quad (10)$$

$$W(n_i = 0 \rightarrow n_i = 1) = \lambda(n_{i-1} + n_{i+1}) + h, \quad (11)$$

where  $\lambda$  and  $h$  are positive constants [15]. We will use this system in order to extend the analysis of the large deviations of the activity  $K^1$  introduced in Ref. [14] to the large- $L$  limit.

### III. SCALINGS OF THE LARGE DEVIATION FUNCTION ESTIMATOR

The approach described in Sect. II was followed in Ref. [14] in order to analyze the finite-time and finite- $N_c$  behavior of the large deviations of the activity for two specific models: a simple one-site annihilation-creation dynamics and a contact process (as described in Sec. II C) with  $L = 6$  sites. The resulting scaling behavior was verified to hold in both cases. However, an analysis of the dependence of this behavior with the number of sites  $L$  remained pending. Below, we summarize the finite-time and finite- $N_c$  scalings of the CGF estimator. We will make use of these results in further sections in order to check their validity in the large  $L$ -limit.

#### A. Large-Time and Large- $N_c$ Behavior

The time evolution of the CGF estimator  $\overline{\Psi_s^{(N_c)}(t)}$  is well described by a curve  $f_t^{(N_c)}$  whose behavior indicate the existence of a  $t^{-1}$ -convergence to  $f_\infty^{(N_c)} = \lim_{t \rightarrow \infty} \overline{\Psi_s^{(N_c)}(t)}$  ( $t^{-1}$ -**scaling**). Similarly, this infinite-time limit  $f_\infty^{(N_c)}$  exhibits  $1/N_c$  corrections for large but finite  $N_c \in \vec{N}_c = \{N_c^{(1)}, \dots, N_c^{(j)}\}$  ( $N_c^{-1}$ -**scaling**). These

<sup>1</sup> The activity is defined as the number of configuration changes on the time interval  $[0, t]$ . For this observable,  $a(C, C') = 1$  and  $b \equiv 0$  in Eq. (2).

$t$ - and  $N_c$ -scalings are explicitly described by

$$f_t^{(N_c)} = f_\infty^{(N_c)} + b_t^{(N_c)} t^{-1}, \quad (12)$$

$$f_\infty^{(N_c)} = f_\infty^\infty + b_\infty^{(N_c)} N_c^{-1}, \quad (13)$$

where  $f_\infty^\infty = \lim_{N_c \rightarrow \infty} f_\infty^{(N_c)}$  is the infinite- $N_c$  limit of  $f_\infty^{(N_c)}$ . The parameters  $b_t^{(N_c)}$  and  $b_\infty^{(N_c)}$  can be interpreted as characteristic times and sizes, respectively.<sup>2</sup>

The curve  $\overline{f_t^{(N_c)}}$  (Eq. (12)) is determined from a fit in time over  $\overline{\Psi_s^{(N_c)}(t)}$  up to (the final simulation) time  $T$ . By other hand,  $f_\infty^{(N_c)}$  (Eq. (13)) is obtained from a fit in  $N_c$  over  $f_\infty^{(N_c)}$  computed for several populations sizes  $N_c \in \vec{N}_c = \{N_c^{(1)}, \dots, N_c^{(j)}\}$ . These equations imply that  $\overline{\Psi_s^{(N_c)}(t)}$  converges to its infinite-time and  $-N_c$  limit  $f_\infty^\infty$  proportionally to  $1/t$  and  $1/N_c$ . Moreover, this limit can be extracted from finite-time and finite- $N_c$  data by making use of the **scaling method** also proposed in Ref. [14]. The results obtained for  $f_\infty^\infty$  rendered a better estimation of  $\psi(s)$  than the standard estimator  $\overline{\Psi_s^{(N_c)}(t)}$  evaluated for  $N_c = \max \vec{N}_c$  and for  $t = T$ .

### IV. SCALINGS IN THE LARGE- $L$ LIMIT

In order to prove whether the finite-time and  $-N_c$  scalings observed in small (number of sites  $L$ ) systems are also valid in the large- $L$  limit, we assume that the CGF estimator (and its infinite-time limit) can be described by equations of the form

$$\chi_t^{(N_c)} \equiv \chi_\infty^{(N_c)} + \kappa_t^{(N_c)} t^{-\gamma_t}, \quad (14)$$

$$\chi_\infty^{(N_c)} \equiv \chi_\infty^\infty + \kappa_\infty^{(N_c)} N_c^{-\gamma_{N_c}}, \quad (15)$$

redefining in a more general way the scalings (12) and (13).<sup>3</sup> We will refer to Eq. (14) as  $t^{-\gamma_t}$ -**scaling** whereas Eq. (15) as  $N_c^{-\gamma_{N_c}}$ -**scaling**.

The problem reduces in determining the exponents  $\gamma_t$  and  $\gamma_{N_c}$  in order to verify if effectively  $\gamma_t \approx 1$  and  $\gamma_{N_c} \approx 1$

<sup>2</sup> Additionally, the behavior of the CGF estimator at final time  $T$  (the standard estimator)  $\overline{\Psi_s^{(N_c)}(T)}$  as a function of the population size  $N_c$  is well described by a curve of the form

$$g_{N_c}^{(T)} = g_\infty^{(T)} + \tilde{b}_{N_c}^{(T)} N_c^{-1}$$

indicating that  $\overline{\Psi_s^{(N_c)}(T)}$  also converges to its infinite- $N_c$  limit  $g_\infty^{(T)} = \lim_{N_c \rightarrow \infty} \overline{\Psi_s^{(N_c)}(T)}$  with an error proportional to  $1/N_c$ .

<sup>3</sup> Similarly, the  $N_c$ -behavior of  $\overline{\Psi_s^{(N_c)}(T)}$  can be described by the equation

$$g_{N_c}^{(T)} = \chi_\infty^{(T)} + \tilde{\kappa}_{N_c}^{(T)} N_c^{-\gamma_{N_c}^T}, \quad (16)$$

where  $\chi_\infty^{(T)} = \lim_{N_c \rightarrow \infty} \overline{\Psi_s^{(N_c)}(T)}$ . Here it is important to remark that both  $\chi_\infty^{(N_c)}$  and  $\overline{\Psi_s^{(N_c)}(T)}$  scale in the same way in  $N_c$ . In other words,  $\gamma_{N_c} \approx \gamma_{N_c}^T$ .

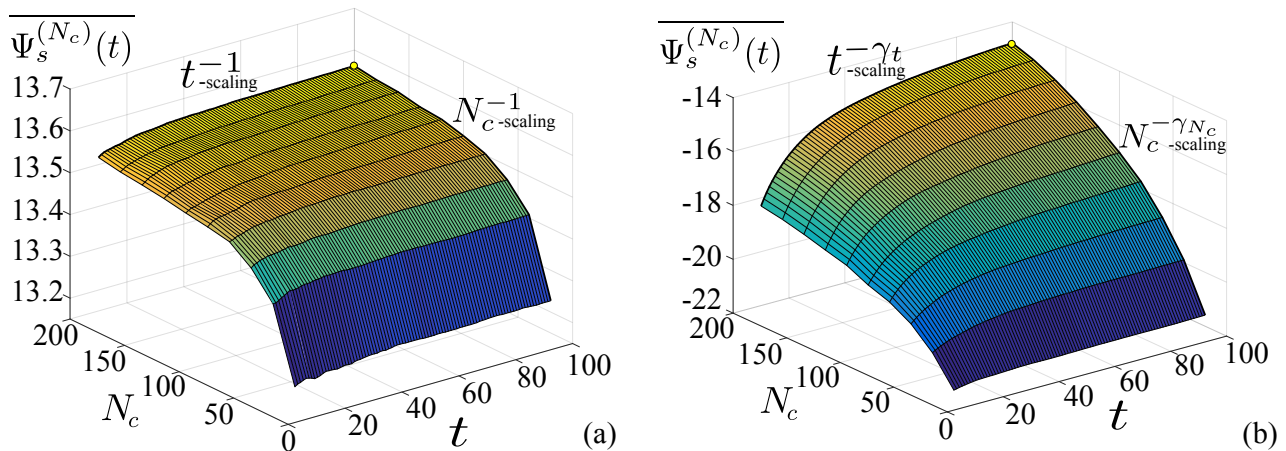


Figure 1. Large deviations of the activity for a contact process with  $L = 100$ ,  $\lambda = 1.75$  and  $h = 0.1$ . The CGF estimator  $\overline{\Psi_s^{(N_c)}(t)}$  (Eq. (8)) is presented as a function of time  $t$  and the number of clones  $N_c$  for (a)  $s = -0.1$  and (b)  $s = 0.2$ . These surfaces were computed using the continuous-time cloning algorithm up a final simulation time  $T = 100$ ,  $\vec{N}_c = \{20, \dots, 200\}$  and  $R = 500$  realizations. The  $N_c^{-1}$ -scaling observed in small  $L$  systems holds only for  $s = -0.1$  whereas for  $s = 0.2$  a  $N_c^{-\gamma_{N_c}}$ -scaling is observed ( $\gamma_{N_c}(s = 0.2) \approx -0.16$ ). Similarly, for the time-scaling for which  $\gamma_t(s = 0.2) \approx 0.7$ .

and whether the terms  $\chi_\infty^{(N_c)}$  and  $\chi_\infty^\infty$  represent the limits in  $t \rightarrow \infty$  and  $N_c \rightarrow \infty$  of the CGF estimator. Thus, a value of the exponent  $\gamma_t \approx 1$ , verifies  $\chi_\infty^{(N_c)} \approx f_\infty^{(N_c)}$  and  $\gamma_{N_c} \approx 1$ , verifies  $\chi_\infty^\infty \approx f_\infty^\infty$ . This is done in the section below on a contact process with  $L = 100$  sites.

## V. CONTACT PROCESS ( $L = 100$ SITES)

### A. Finite-Time and Finite- $N_c$ Scalings

In Fig. 1, we have considered the behavior of  $\overline{\Psi_s^{(N_c)}(t)}$  as function of  $t$  and  $N_c$ , for two representative values of the parameter  $s$ ,  $s = -0.1$  (left) and  $s = 0.2$  (right). Each point of these surfaces was obtained using the cloning algorithm (Eq. (8)) up to time  $T = 100$ , for  $\vec{N}_c = \{20, 40, \dots, 180, 200\}$  and for  $R = 500$  realizations. The best possible CGF estimation (i.e.,  $\overline{\Psi_s^{(N_c)}(t)}$  evaluated at largest  $T$  and number of clones) in both cases is shown with solid circles. According to Ref. [14], these estimations could be improved by using the  $t^{-1}$  and  $N_c^{-1}$ -scalings of the CGF estimator (i.e., the scaling method).

Although the exponents  $\gamma_t$  and  $\gamma_{N_c}$  can be in principle computed for any value of  $N_c \in \vec{N}_c$  and for any  $t \leq T$  (as could be intuited from Fig. 1), from now on, we will consider these exponents defined at the highest number of clones and at final simulation time, i.e.,

$$\gamma_t := \gamma_t(N_c = \max \vec{N}_c), \quad (17)$$

$$\gamma_{N_c} := \gamma_{N_c}(t = T). \quad (18)$$

Thus, the exponent  $\gamma_t$  is obtained by adjusting Eq. (14) to  $\overline{\Psi_s^{(N_c)}(t)}$  for  $N_c = \max \vec{N}_c = 200$  and  $\gamma_{N_c}$  using the

scaling (15) at  $T = 100$ . Alternatively, the exponent  $\gamma_{N_c}$  can be determined from a fit in  $N_c$  over  $\overline{\Psi_s^{(N_c)}(t = T)}$  for  $N_c \in \vec{N}_c$  (Eq. (16)). In simple words, these exponents can be obtained from an adequate fit over the thick curves in Fig. (1). They characterize the finite- $t$  and finite- $N_c$  scalings of the large deviations of the activity  $K$ .

Following this approach, we found that the  $t^{-1}$ -scaling (12) is satisfied only for  $s = -0.1$ . This means that the exponent  $\gamma_t$  was found to be  $\gamma_t \approx 1$ . As a consequence, the parameter  $\chi_\infty^{(N_c)}$  obtained from Eq. (14) effectively represents the limit in  $t \rightarrow \infty$  of the CGF estimator, i.e.,  $\chi_\infty^{(N_c)} \approx f_\infty^{(N_c)}$ . This is not the case for  $s = 0.2$  for which  $\gamma_t(s = 0.2) \approx 0.7$ . Similarly, a  $N_c^{-\gamma_{N_c}}$ -scaling is observed for  $s = 0.2$ , whereas for  $s = -0.1$ , the  $N_c^{-1}$ -scaling (13) holds. It is important to remark that a value of exponent  $\gamma_{N_c} > 0$  could still guaranty the convergence of the CGF estimator in the infinite- $N_c$  limit. However, even though  $\gamma_{N_c}(t = 10) > 0$  at initial times, at final time  $T$ , the exponent is negative ( $\gamma_{N_c}(t = T) \approx -0.16$ ), which would imply that  $\chi_\infty^\infty$  does not represent the infinite- $t$  infinite- $N_c$  limit  $f_\infty^\infty$  of the CGF estimator. Below, we present how the change in the scalings is produced depending on  $s$ .

### B. Dependence on $s$ : $\gamma_t(s)$ & $\gamma_{N_c}(s)$

The exponents  $\gamma_t$  and  $\gamma_{N_c}$  were determined (as described above) for values of  $s$  ranging in the interval  $s \in [-0.1, 0.2]$ . Two values of system size,  $L = 6$  (circles) and  $L = 100$  (squares), are compared in Fig. 2(a) for  $\gamma_t(s)$ . As can be seen, independently of  $L$ , for  $s < 0$ , the exponent  $\gamma_t(s)$  varies around 1. However for  $s > 0$ ,



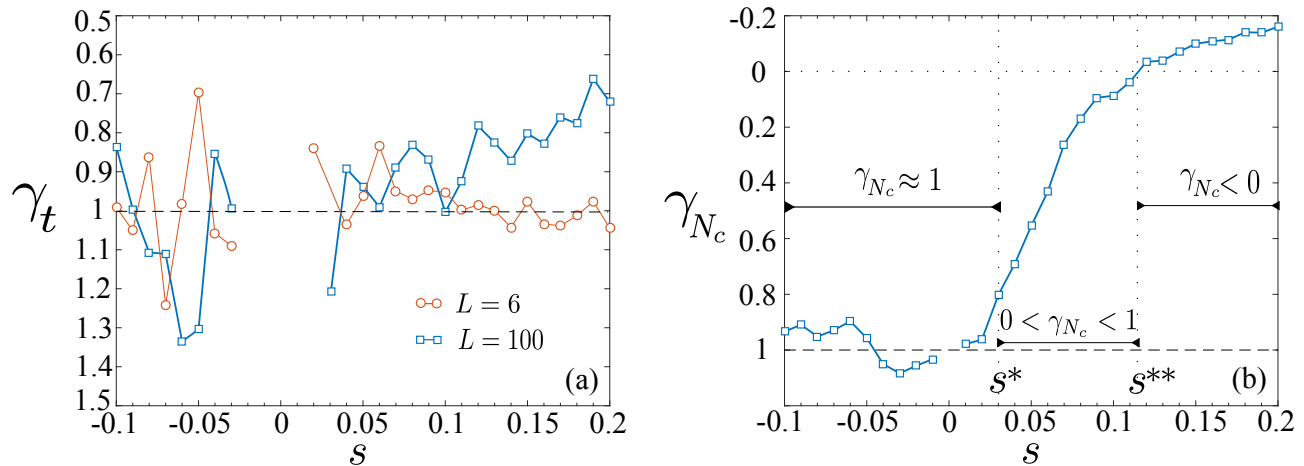


Figure 2. Dependence of the  $t^{-\gamma_t}$ - and  $N_c^{-\gamma_{N_c}}$ -scalings with the parameter  $s \in [-0.1, 0.2]$ . **(a)** The exponent  $\gamma_t$  is determined adjusting  $\overline{\Psi_s^{(N_c)}}(t)$  (up to  $T = 100$ ) to Eq. (14) at  $\max \vec{N}_c = 200$ . For  $s < 0$ , this exponent takes values around 1 for both number of sites  $L$  considered:  $L = 6$  (circles) and  $L = 100$  (squares). However, for  $s > 0$ , the value corresponding to  $L = 100$  deviates from 1 decreasing with  $s$ . **(b)** The exponent  $\gamma_{N_c}(s)$  is obtained by fitting  $\overline{\Psi_s^{(N_c)}}(T = 100)$  as function of  $N_c \in \vec{N}_c = \{20, 40, \dots, 180, 200\}$  by Eq. (16). Three stages of  $\gamma_{N_c}(s)$  can be clearly distinguish for  $L = 100$ : i)  $\gamma_{N_c}(s < s^*) \approx 1$ , ii)  $0 < \gamma_{N_c}(s^* < s < s^{**}) < 1$ , and iii)  $\gamma_{N_c}(s > s^{**}) < 0$ .

this is true only for  $L = 6$  (for which the amplitudes of the fluctuations are considerably smaller). For  $L = 100$ , we observe that  $\gamma_t$  deviates slightly from 1 decreasing with  $s$  up to  $\gamma_t \approx 0.7$  at  $s = 0.2$ .

In order to describe the behavior of this exponent, results convenient to define the quantity  $s'$  as the value of the parameter  $s \in [s_a, s_b]$  such that  $\gamma_t(s < s') \approx 1$ , i.e., until which the  $t^{-1}$ -scaling holds. If the scaling holds  $\forall s \in [s_a, s_b]$  (given some  $L$ ), then  $s' = s_b$ . Thus,

$$\begin{cases} \gamma_t(s) \approx 1, & \text{for } s < s' \\ 0 < \gamma_t(s) < 1, & \text{otherwise.} \end{cases} \quad (19)$$

By other hand, the value of  $s \in [s_a, s_b]$  which signals the validity of the  $N_c^{-1}$ -scaling is denoted by  $s^*$ . From this point,  $\gamma_{N_c}$  decreases until eventually it becomes negative, as can be seen in Fig. 2(b). Here, we introduce  $s^{**}$  such that  $\gamma_{N_c}(s = s^{**}) = 0$ . Thus,  $\gamma_{N_c} < 0$  for  $s > s^{**}$ . This behavior was not observed for  $L = 6$  for which the  $N_c^{-1}$ -scaling was valid independently of  $s$  [14]. In those cases,  $s^* = s_b$  and  $\nexists s^{**}$ . Instead of confirming for  $L = 100$  the  $N_c^{-1}$ -scalings of the CGF estimator presented in Ref. [14], here we have been able to distinguish clearly three stages for the exponent  $\gamma_{N_c}(s)$ :

$$\begin{cases} \gamma_{N_c}(s) \approx 1, & \text{for } s < s^* \\ 0 < \gamma_{N_c}(s) < 1, & \text{for } s^* < s < s^{**} \\ \gamma_{N_c}(s) < 0, & \text{for } s > s^{**}. \end{cases} \quad (20)$$

The possibility of extracting the infinite- $t$  and infinite- $N_c$  limit of the CGF estimator relied on the validity of the  $t^{-1}$ - and  $N_c^{-1}$ -scalings. How the results obtained from the application of the scaling method are affected by  $\gamma_t$  and  $\gamma_{N_c}$  are shown below.

### C. The Scaling Method

The scaling method allows to determine the asymptotic limit to which the CGF estimator (8) converges in the infinite- $t$  and  $-N_c$  limits. This limit, that we have denoted  $f_\infty^\infty$  (Eq. (13)), was proved to render a better estimation of the analytical LDF  $\psi(s)$  than the standard estimator  $\overline{\Psi_s^{(N_c)}}(t)$  evaluated at  $\max \vec{N}_c$  and at  $t = T$ , at least for the cases analyzed in Ref. [14]. However, the results we just presented would suggest that the determination of  $f_\infty^\infty$  could be affected depending whether  $\gamma_t \approx \gamma_{N_c} \approx 1$  or not. If this is true, the scaling method could render valid results in our example only for  $s < 0$ . Solely in this region the extracted  $\chi_\infty^\infty$  (obtained from Eq. (15)) would represent the infinite- $t$  and  $-N_c$  limit of the CGF estimator. Indeed, this can be observed in Fig. 3 where we have applied the scaling method to our example.

The method can be performed following two different approaches: i)  $(t^{-1}, N_c^{-\gamma_{N_c}})$ : First, imposing a  $t^{-1}$ -scaling for  $\overline{\Psi_s^{(N_c)}}(t)$  (setting  $\gamma_t = 1$  in Eq. (14)) and then, considering a  $N_c^{-\gamma_{N_c}}$ -scaling (15) for the extracted  $\chi_\infty^{(N_c)}$ , or alternatively, ii)  $(t^{-\gamma_t}, N_c^{-\gamma_{N_c}})$ : Leaving  $\gamma_t$  and  $\gamma_{N_c}$  as free parameters in Eqs. (14) and (15). Both resulting estimators  $\chi_\infty^\infty(i)$  and  $\chi_\infty^\infty(ii)$  are shown in Fig. 3 with squares and circles, respectively. Additionally, the infinite- $N_c$  limit  $\chi_\infty^T$  (16) is also presented with diamonds.

The standard CGF estimator  $\overline{\Psi_s^{(\max \vec{N}_c)}}(T)$  (dots) serves as reference of the effectiveness of the method.

As can be seen in Fig. 3, the different estimators correspond to each others up to  $s = s^*$ . From this point, their distance with respect to  $\overline{\Psi_s^{(\max \vec{N}_c)}}(T)$  increases rapidly

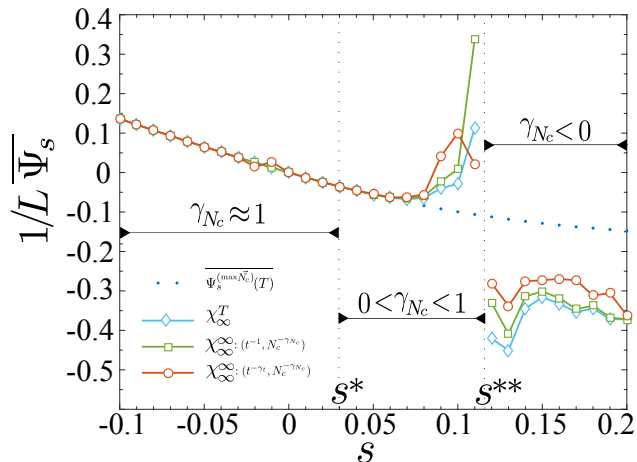


Figure 3. Estimators of the large deviations of the activity as function of the parameter  $s \in [-0.1, 0.2]$  for a contact process with  $L = 100$  sites. The standard CGF estimator  $\overline{\Psi_s^{(N_c)}(t)}$  (evaluated at  $N_c = \max \bar{N}_c = 200$ ,  $t = T = 100$  and for  $R = 500$  realizations) is shown with dots meanwhile the ones obtained from the scaling method  $\chi_\infty$  are presented in squares and circles, and  $\chi_\infty^T$  in diamonds. The legend  $(t^{-1}, N_c^{-\gamma_{N_c}})$  refers to the assumption of a  $t^{-1}$ -scaling for  $\overline{\Psi_s^{(N_c)}(t)}$  (setting  $\gamma_t = 1$  in Eq. (14)) and a  $N_c^{-\gamma_{N_c}}$ -scaling (15) for  $\chi_\infty^{(N_c)}$ . By other hand,  $(t^{-\gamma_t}, N_c^{-\gamma_{N_c}})$  refers to the fact that we have left  $\gamma_t$  and  $\gamma_{N_c}$  as free parameters. The different estimators correspond to each others up to  $s = s^*$  from which they diverge up to  $s = s^{**}$  where there is a discontinuity and they become negative for  $s > s^{**}$ . This is directly related with the behavior of the exponent  $\gamma_{N_c}$  observed in Fig. 2(b).

with  $s$  up to  $s = s^{**}$  where a discontinuity occurs. In fact, the behavior observed in Fig. 3 keeps correspondence with the  $N_c^{-\gamma_{N_c}}$ -scaling of the CGF estimator. Specifically, with the stages of the exponent  $\gamma_{N_c}$  that were presented in Sect. VB and Fig. 2(b). Thus, the discontinuity in  $\chi_\infty$  is related precisely with the change in sign of  $\gamma_{N_c}$  in  $s = s^{**}$  in the same way as the divergence of the estimators from the standard one at  $s = s^*$  is related with the fact that from this point,  $\gamma_{N_c} \neq 1$ .

The example presented through this section related the effectiveness of the scaling method as proposed in Ref. [14] with the actual scaling of the CGF estimator in large  $L$  systems. Depending on the value of the exponents  $(\gamma_t, \gamma_{N_c})$  it is possible to extract or not the infinite- $t$  and infinite- $N_c$  limit of the CGF estimator. Moreover, we also showed the dependence of these exponents with the parameter  $s$ . Below we extend our analysis by considering the scaling behavior on a wider range of values of  $L$ . This will provide a complete overview of how the CGF estimator behaves and how the change in scaling is given.

## VI. PLANE $s - L$ BEHAVIOR

In this section we detail the behavior of the finite- $t$  and  $-N_c$  scalings of the CGF estimator in the plane  $s - L$ . The exponents  $\gamma_t$  and  $\gamma_{N_c}$  were computed for values of the parameters  $s > 0$  and number of sites  $L$  ranging in the interval  $L \in [3, 100]$ .

### A. $\gamma_t(s, L)$

The contour plot in Fig. 4 shows the value of the exponent  $\gamma_t$  as it changes depending on the parameters  $L$  and  $s$ . We have focused in the region for  $s \in [0.02, 0.2]$  as for  $s < 0$ ,  $\gamma_t \approx 1$  and thus, the  $t^{-1}$ -scaling (12) holds. The values closest to 1 are presented with the darkest tone while smaller values are shown with clearer tones. As can be seen, the exponent  $\gamma_t$  decreases gradually as  $L$  and  $s$  increase.

Given a particular value of  $L$ , qualitatively speaking, the behavior of  $\gamma_t$  with respect to  $s$  is similar to what we described for  $L = 100$  in Sec. VB (except for small values of  $L$ , for which the  $t^{-1}$ -scaling holds even for  $s > 0$  [14]). In order to extend that description into the plane  $s - L$ , we introduce a number of sites dependency of the bound  $s'$ . We denote by  $s'(L)$  the value of  $s$  until which the  $t^{-1}$ -scaling is valid given a particular  $L$ . Similarly,  $\gamma_t^o(L)$  is the lower bound of  $\gamma_t^{(L)}(s)$ . Thus, the exponent  $\gamma_t$  which characterizes the  $t^{-\gamma_t}$ -scaling (14) of the CGF estimator

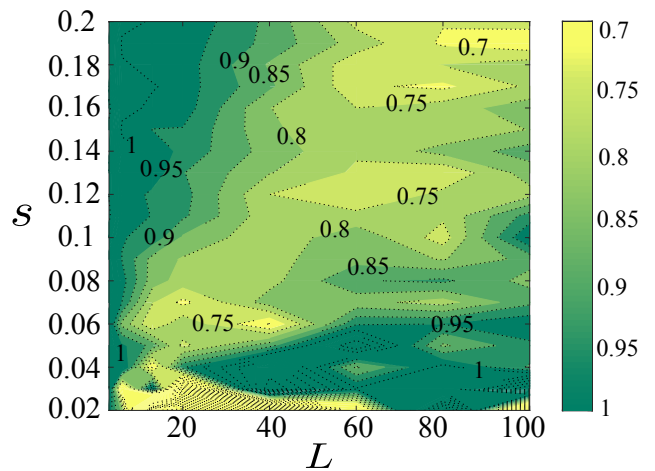


Figure 4. Behavior of the exponent  $\gamma_t$  in a region of the plane  $s - L$  given by the parameters  $s \in [0.02, 0.2]$  and  $L \in [3, 100]$ . This exponent characterizes the finite-time scaling of the large deviations of the activity in the contact process ( $t^{-\gamma_t}$ -scaling). The values of  $\gamma_t$  closest to 1 are presented with the darkest tones whereas smaller values are shown with clearer ones. Although for small values of  $L$  the  $t^{-1}$ -scaling holds independently of  $s$ , in general the exponent  $\gamma_t$  decreases gradually as  $L$  and  $s$  increase.

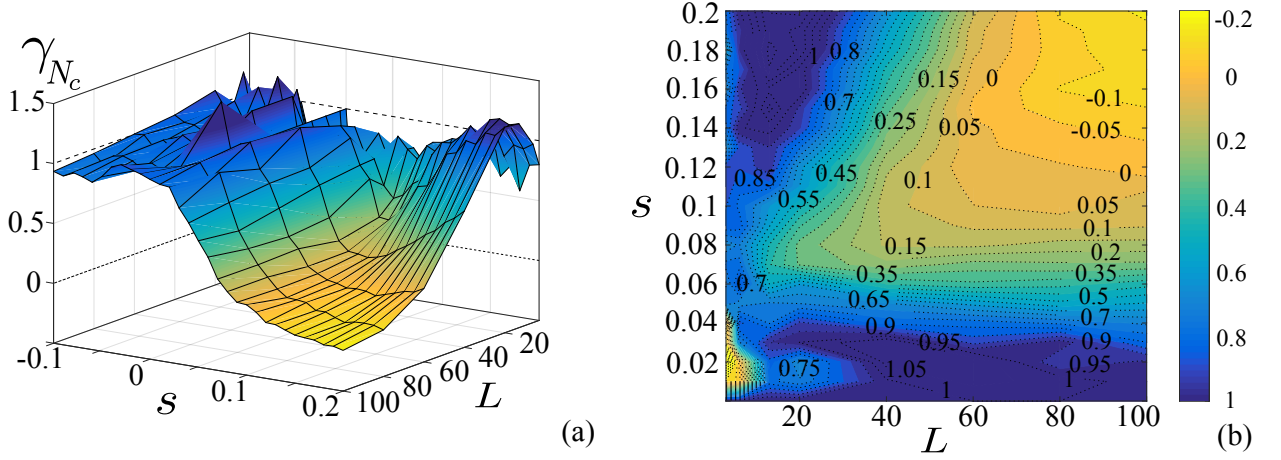


Figure 5. Behavior of the exponent  $\gamma_{N_c}$  in a region of the plane  $s-L$  given by the parameters  $s \in [-0.1, 0.2]$  and  $L \in [3, 100]$ . This exponent characterizes the finite- $N_c$  scaling of the large deviations of the activity in the contact process ( $N_c^{-\gamma_{N_c}}$ -scaling). (a) The surface  $\gamma_{N_c}(s, L)$  illustrates the change in the scaling of the CGF estimator. The exponent  $\gamma_{N_c}(s, L) \approx 1$  up to some value of  $s$  (which depends of  $L$ ), from which it decreases as  $s$  and  $L$  increases, even becoming negative for some values of  $L$ . (b) Projection of the surface in (a) on the plane  $s-L$ . The values of  $\gamma_{N_c}(s, L)$  closer to 1 are shown in dark tones while the smaller tones with clearer ones. The  $N_c^{-\gamma_{N_c}}$ -scaling can be characterized depending on the number of zeros of the exponent  $\gamma_{N_c}(s)$  for a given  $L$  in three regions:  $\mathcal{L}_I$  if has no zeros,  $\mathcal{L}_{II}$ : if have two zeros and  $\mathcal{L}_{III}$ : if have one zero.

is given by

$$\begin{cases} \gamma_t^{(L)}(s) \approx 1, & \text{for } s < s'(L) \\ \gamma_t^\circ(L) < \gamma_t^{(L)}(s) < 1, & \text{otherwise} \end{cases} \quad (21)$$

where  $s'(L) > 0$ ,  $\gamma_t^\circ(L) > 0$  and  $L$  is large. In fact, for this case,  $\gamma_t^\circ(L) > 1/2$ , for all  $L$ .

### B. $\gamma_{N_c}(s, L)$

Similarly as above, in Fig. 5 we present the exponent  $\gamma_{N_c}$  as it changes depending of some particular choice of the parameters  $(s, L)$  within the intervals considered. The surface in Fig. 5(a) illustrates clearly the change in the  $N_c$ -scaling of the CGF estimator. For every value of  $L$  considered, the exponent  $\gamma_{N_c}$  is approximately 1 up to some value of  $s$ , denoted as  $s^*(L)$  (Sec. VB) which is different for each  $L$ . However, from this point, its value decreases as  $s$  and  $L$  increases, becoming, in some cases, negative. This change in the  $N_c^{-\gamma_{N_c}}$ -scaling is also shown in the contour plot in Fig. 5(b) where we have focus specially in the region for  $s > 0$ . The values of  $\gamma_{N_c}$  closer to 1 are shown in dark tones.

In Sec. VB, we also defined  $s^{**}$  such that  $\gamma_{N_c}^{(L)}(s^{**}) = 0$ . This value of course depends on  $L$  and in some cases it does not even exists. However, for some particular values of  $L$  (large), the exponent  $\gamma_{N_c}^{(L)}$  changes sign twice (as can be seen in Fig. 5(b)). We will use this fact in order to characterize the  $N_c^{-\gamma_{N_c}}$ -scaling depending on the number of zeros of the exponent  $\gamma_{N_c}^{(L)}(s)$  for a given  $L$ .

We define  $\mathcal{L}_I$  as the set of values of  $L$ , for which the exponent  $\gamma_{N_c}^{(L)}(s)$  has no zeros,  $\mathcal{L}_{II}$ : if has two zeros ( $s_1^{**}(L)$  and  $s_2^{**}(L)$ , with  $s_2^{**}(L) > s_1^{**}(L)$ ) and  $\mathcal{L}_{III}$ : if has one zero ( $s^{**}(L)$ ). These regions are bounded by  $L_{inf}$  and/or by  $L_{sup}$ , where  $L_{inf}$  is the smallest value of  $L$  such that the curve  $L = L_{inf}$  is tangent to  $\gamma_{N_c}(s, L) = 0$  in one single point. By other hand,  $L_{sup}$  is the largest  $L$  such that the curve  $L = L_{sup}$  cuts  $\gamma_{N_c}(L, s) = 0$  in two points. Thus, the region  $\mathcal{L}_I$  groups the values of  $L$  such that  $L < L_{inf}$ ,  $\mathcal{L}_{II}$  the values of  $L$  within the interval  $L_{inf} < L < L_{sup}$  and  $\mathcal{L}_{III}$ , the values of  $L$  such that  $L > L_{sup}$ . Thus, the exponent  $\gamma_{N_c}$  which characterizes the  $N_c^{-\gamma_{N_c}}$ -scaling (15) of the CGF estimator is given by

$$\begin{cases} \mathcal{L}_I : \begin{cases} \gamma_{N_c}^{(L)}(s) \approx 1, & \text{for } s < s^*(L) \\ 0 < \gamma_{N_c}^{(L)}(s) \lesssim 1, & \text{otherwise.} \end{cases} \\ \mathcal{L}_{II} : \begin{cases} \gamma_{N_c}^{(L)}(s) \approx 1, & \text{for } s < s^*(L) \\ 0 < \gamma_{N_c}^{(L)}(s) < 1, & \text{for } s^*(L) < s < s_1^{**}(L) \\ & \text{and } s > s_2^{**}(L) \\ \gamma_{N_c}^{(L)}(s) < 0, & \text{for } s_1^{**}(L) < s < s_2^{**}(L) \end{cases} \\ \mathcal{L}_{III} : \begin{cases} \gamma_{N_c}^{(L)}(s) \approx 1, & \text{for } s < s^*(L) \\ 0 < \gamma_{N_c}^{(L)}(s) < 1, & \text{for } s^*(L) < s < s^{**}(L) \\ \gamma_{N_c}^{(L)}(s) < 0, & \text{for } s > s^{**}(L) \end{cases} \end{cases} \quad (22)$$

## VII. CONCLUSION

In this paper we analyzed the large deviations of the activity in the contact process. We used the continuous-time version of the cloning algorithm from which the LDF can be estimated from the exponential growth (or decay) rate of a set of  $N_c$  copies of the system which evolves following a modified dynamics up to a time  $t = T$ . It is expected that in the infinite- $t$  and  $-N_c$  limits, this method provides an accurate LDF estimation. However, in practice, the best estimation is obtained is by considering a large enough simulation time  $T$  and number of clones  $N_c$ .

The dependence of this estimator (and of its accuracy) with these two parameters was studied in Refs. [13, 14]. The finite- $t$  and finite- $N_c$  scalings of the systematic errors of the LDF were found to behave as  $1/N_c$  and  $1/t$  in the large- $N_c$  and large- $t$  asymptotics, respectively. By making use of these convergence-speeds, it was proposed a (scaling) method which allowed to extract the asymptotic behavior of the CGF estimator in the  $t \rightarrow \infty$  and  $N_c \rightarrow \infty$  limits. At least for the cases analyzed in Refs. [13, 14], this infinite-time and infinite- $N_c$  limit resulted to render a better LDF estimation in comparison with the standard estimator. However, the validity of the method and of these scalings were proved only for a simple one-site annihilation-creation dynamics and for a contact process with  $L = 6$  sites, leaving an analysis of the dependence with the number of sites  $L$  pending.

In order to do so, in this paper, we redefined these scalings in a more general way. We assume the behavior of the CGF estimator described by a  $t^{-\gamma_t}$ -scaling (Eq. (14)) and a  $N_c^{-\gamma_{N_c}}$ -scaling (Eq. (15)). This redefinition allowed us to verify in large  $L$  systems if effectively  $\gamma_t \approx 1$  and  $\gamma_{N_c} \approx 1$  and whether the terms  $\chi_\infty^{(N_c)}$  and  $\chi_\infty$  represent the limits in  $t \rightarrow \infty$  and  $N_c \rightarrow \infty$  of the CGF estimator.

This was done at first in Sec. V A where we considered a contact process with  $L = 100$  sites and two representa-

tive values of the parameter  $s$ . Although the  $t^{-1}$ -scaling and  $N_c^{-1}$ -scaling were proved to hold for  $s = 0.1$ , this was not the case for  $s = 0.2$ . How this change in the scaling is produced depending on the parameter  $s$  was presented in detail in Sect. V B where the exponents  $\gamma_t(s)$  and  $\gamma_{N_c}(s)$  were characterized. Particularly, for  $\gamma_{N_c}(s)$ , we were able to distinguish three stages in its behavior, where, the  $N_c^{-1}$ -scaling was valid up to  $s = s^*$ , then  $\gamma_{N_c}$  decreases to 0 at  $s = s^{**}$  and finally, it becomes negative for  $s > s^{**}$ .

In Sec. V C we showed how these scalings affect the determination of the infinite- $t$  and infinite- $N_c$  limit of the CGF estimator. This occurs because the scaling method relied on the validity of the  $t^{-1}$ - and  $N_c^{-1}$ -scalings. As for  $L = 100$  this was not the case, it was possible to see how the different estimators corresponded to each others up to  $s = s^*$  from which they diverge up to  $s = s^{**}$  where there is a discontinuity.

This analysis was extended to the plane  $s-L$  in Sec. VI where the exponents  $\gamma_t$  and  $\gamma_{N_c}$  were computed for a grid of values of the parameters  $(s, L)$ . Their characterization was done introducing a number-of-sites dependency of the bounds  $s'$ ,  $s^*$  and  $s^{**}$  previously defined in Sec. V as well as the use of the number of zeros of the exponent  $\gamma_{N_c}^{(L)}(s)$  in order to characterize the different groups of  $L$ . Whether the results presented through this paper are restricted only to the contact process or not is leaved as a pending problem and a possible direction for future research.

## ACKNOWLEDGMENTS

E. G. thanks Khashayar Pakdaman for his support and discussions. Special thanks to the Ecuadorian Government and the Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación, SENESCYT.

- 
- [1] H. Touchette, *Physics Reports* **478**, 1 (2009).
  - [2] C. Giardinà, J. Kurchan, V. Lecomte, and J. Tailleur, *J. Stat. Phys.* **145**, 787 (2011).
  - [3] J. Bucklew, *Introduction to Rare Event Simulation* (Springer Science & Business Media, 2013).
  - [4] H. Kahn and T. E. Harris, *National Bureau of Standards applied mathematics series* **12**, 27 (1951).
  - [5] D. Aldous and U. Vazirani, in *Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on* (IEEE, 1994) pp. 492–501.
  - [6] P. Grassberger, *Computer Physics Communications* **147**, 64 (2002), proceedings of the Europhysics Conference on Computational Physics Computational Modeling and Simulation of Complex Systems.
  - [7] F. Cérou and A. Guyader, *Stochastic Analysis and Applications* **25**, 417 (2007).
  - [8] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annual Review of Physical Chemistry* **53**, 291 (2002).
  - [9] C. Giardinà, J. Kurchan, and L. Peliti, *Phys. Rev. Lett.* **96**, 120603 (2006).
  - [10] V. Lecomte and J. Tailleur, *J. Stat. Mech.* **2007**, P03004 (2007).
  - [11] J. Tailleur and V. Lecomte, *AIP Conf. Proc.* **1091**, 212 (2009).
  - [12] E. Guevara Hidalgo and V. Lecomte, *J. Phys. A: Math. Theor.* **49**, 205002 (2016).
  - [13] T. Nemoto, E. Guevara Hidalgo, and V. Lecomte, *Phys. Rev. E* **95**, 012102 (2017).
  - [14] E. Guevara Hidalgo, T. Nemoto, and V. Lecomte, *Phys. Rev. E* **95**, 062134 (2017).
  - [15] T. E. Harris, *Ann. Probability* **2**, 969 (1974).
  - [16] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, *J. Phys. A* **42**, 075007 (2009).
  - [17] J. Tailleur and J. Kurchan, *Nat Phys* **3**, 203 (2007).