Statistics of trajectories in two-state master equations

Andrew D. Jackson and Simone Pigolotti
The Niels Bohr International Academy, The Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark
(Received 2 December 2008; published 18 February 2009)

We derive a simple expression for the probability of trajectories of a master equation. The expression is particularly useful when the number of states is small and permits the calculation of observables that can be defined as functionals of whole trajectories. We illustrate the method with a two-state master equation, for which we calculate the distribution of the time spent in one state and the distribution of the number of transitions, each in a given time interval. These two expressions are obtained analytically in terms of modified Bessel functions.

DOI: 10.1103/PhysRevE.79.021121 PACS number(s): 02.50.Ga, 05.10.Gg

The evolution of many systems in physics, chemistry, and biology is properly described by master equations. This description is adequate when the system under consideration has discrete states and when the rate of jumping from one state to another does not depend on the history of the system, i.e., when the Markov property holds. In recent years, this description has been successfully applied to a plethora of new problems in several fields. As examples, master equations are commonly used in biochemistry to understand the fluctuations of chemical concentrations inside the cell [1]. In statistical physics, they can provide a simple description of nonequilibrium systems, useful for testing the validity of fluctuation relations [2–4].

From a technical point of view, master equations now constitute a well-established field of research, and many techniques have been developed which permit their analytical or numerical treatment [5–7]. In complicated cases, these techniques permit calculation of the steady-state probabilities P_n of being in state n. In simpler cases, it is sometimes possible to solve equations in time in order to determine the propagator $p(n,t|n_0,0)$ that gives the probability of being in state n at time T starting from a state n_0 at time $t=0$.

For many practical purposes, determination of the propagator is sufficient, since many interesting observables can be expressed as a function of the propagator. There are, however, observables that cannot be obtained conveniently from the propagator, including in particular quantities which are more easily expressed as *functionals* of entire trajectories. Examples include the distribution of the time spent in a given state and the probability of observing a given number of transitions, both for a fixed time interval. Functional methods are well known for continuous stochastic process, where techniques have been developed in parallel to those used in quantum mechanics [8]. There are fewer examples of functional methods for discrete processes [9,10]. These methods are often field theoretic in nature and involve complications such as renormalization which one would like to avoid in simple discrete systems.

In this paper, we present a simple way to calculate probabilities of the trajectories of master equations. The method is straightforward, rigorous, and does not require any specific assumptions on the equation. It is particularly useful when the number of states available to the system is small, where it is possible to obtain closed analytical expression for several interesting observables. We study as example of our method general two-state systems that, despite their simplicity, have many nontrivial applications in problems related to single-molecule spectroscopy (see, e.g. [11,12]) and biophysics (see, e.g., [13–15]). Specifically, we calculate the probability of observing a given number of transitions, N, in a time T, and the distribution of time spent in one of the two states in a time T. Each of these quantities can be expressed in terms of modified Bessel functions.

We consider a master equation

$$\frac{d}{dt}P_n = \sum_m W_{mn}P_m - W_{nm}P_n, \tag{1}$$

where $P_n(t)$ is the time-dependent probability of being in state n and W_{mn} is the transition rate from state m to n. For convenience we also define

$$W_n^{\mathrm{out}} = \sum_i W_{ni}, \tag{2}$$

the total out-rate of state n. The probability that, in a time T, the trajectory visits a predetermined sequence of states $n_0, n_1, n_2, \ldots, n_N$ then becomes

$$\mathcal{P}(n_0,n_1,\ldots,n_N;T) = \int_0^T dt_1 \int_{t_1}^T dt_2 \cdots \int_{t_{n-1}}^T dt_N e^{-W_0^{\mathrm{out}} t_1} \times W_{n_0 n_1} e^{-W_1^{\mathrm{out}}(t_2-t_1)} \cdots W_{n_{N-1}n_N} e^{-W_n^{\mathrm{out}}(T-t_n)}. \tag{3}$$

This expression can be understood by noticing that the integrand represents the probability density in time of the N consecutive transitions according to the master equation (see Fig. 1). By summing over all trajectories having predetermined properties, one can reconstruct the full statistics of the stochastic process. An obvious example is the propagator, which can be evaluated as the sum over all trajectories that start in a given state n_0 at time $t=0$ and end in a state n_f at time T:

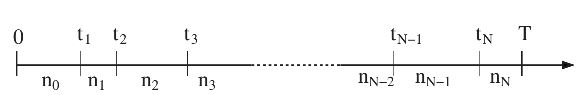

FIG. 1. Trajectory of a master equation as a function of time. The integral in Eq. (3) is over the times of the N transition points.

$$p(n_f,T|n_0,0) = \sum_{N=0}^{\infty} \sum_{\{n_1\cdots n_{N-1}\}} \mathcal{P}(n_0,n_1,\ldots,n_N=n_f;T). \quad (4)$$

Note that all probabilities are properly normalized; in particular, the propagator satisfies the closure relation $\Sigma_{n_F} p(n_f,T|n_0,0)=1$.

The above expressions become particularly useful when the number of distinct states visited by the system is small. In this case, it is convenient to rearrange the integrals in Eq. (3) by grouping together all time intervals in which the system is in the same state. If k_i is the number of times the system visits state i on a given trajectory, one finds

$$\mathcal{P}(n_0,n_1,n_2,\ldots,n_N;T) = W_{n_0n_1}W_{n_1n_2}\cdots W_{n_{N-1}n_N}\int_0^T \delta\left(\sum_i t_i - T\right) \times \prod_i \left(\exp(-W_i^{\text{out}}t_i)\frac{t_i^{k_i-1}}{(k_i-1)!}\right)dt_i \quad (5)$$

where the index i runs over all states visited by the system at least once in the given sequence.

As an example, we consider the simple case of a master equation with two states, + and −, with transition rates k_+ (from − to +) and k_- (from + to −). In spite of its simplicity, this case is of interest for many physical and biological problems [11–15]. We will show that Eq. (5) allows analytic calculation of the probabilities of different classes of trajectories. This makes it possible, for example, to obtain closed expressions for the probability of observing a given number of transitions in a time T and for the probability of spending a given time in states ± during a time T. For convenience, we introduce here the total rate $k_T=k_++k_-$ and the equilibrium probabilities $P_+^{\text{eq}}=k_+/k_T$ and $P_-^{\text{eq}}=k_-/k_T$. In this case, we can immediately write the probabilities of all possible trajectories according to Eq. (5). The simplest trajectories are evidently those in which there is no transition in the interval $[0,T]$:

$$\mathcal{P}(+;T) = e^{-k_-T},$$

$$\mathcal{P}(-;T) = e^{-k_+T}. \quad (6)$$

The determination of general trajectories is simplified by having only two states, since trajectories can only alternate between them. It is then convenient to classify trajectories according to (a) the initial state (+ or −), (b) the total time T, (c) the total time spent in state ±, $t_\pm$, and (d) the total number of transitions, N. This is sufficient to characterize a general term in Eq. (5). Note that slightly different expressions are obtained for N even and for N odd. The result is

$$\mathcal{P}(-,T,t_+,N_{\text{even}}) = \frac{[k_+(T-t_+)]^{N/2}(k_-t_+)^{(N/2-1)}}{\frac{N}{2}!\left(\frac{N}{2}-1\right)!}k_-e^{-r},$$

$$\mathcal{P}(-,T,t_+,N_{\text{odd}}) = \frac{[k_+(T-t_+)(k_-t_+)]^{(N-1)/2}}{\frac{N-1}{2}!\frac{N-1}{2}!}k_+e^{-r},$$

$$\mathcal{P}(+,T,t_+,N_{\text{even}}) = \frac{[k_+(T-t_+)]^{N/2-1}(k_-t_+)^{N/2}}{\frac{N}{2}!\left(\frac{N}{2}-1\right)!}k_+e^{-r},$$

$$\mathcal{P}(+,T,t_+,N_{\text{odd}}) = \frac{[k_+(T-t_+)k_-t_+]^{(N-1)/2}}{\frac{N-1}{2}!\frac{N-1}{2}!}k_-e^{-r}, \quad (7)$$

where $r=[k_-t_++k_+(T-t_+)]$. These equations describe all trajectories with $N>0$ while Eqs. (6) describe the two trajectories with $N=0$. Note, however, that Eqs. (6) describes probabilities while Eqs. (7) are probability densities in t_+. To obtain consistent notation, the two expressions in Eqs. (6) should be multiplied by $\delta(t_+-T)$ and $\delta(t_+)$, respectively. This formalism allows us to calculate the distribution of time spent in a state during a time interval T, $g(t_\pm|T)$. Drawing the initial state from the equilibrium distribution $(P_+^{\text{eq}},P_-^{\text{eq}})$, we find

$$g(t_+|T) = P_+^{\text{eq}}\sum_{N=0}^{\infty}\mathcal{P}(+,t_+,N) + P_-^{\text{eq}}\sum_{N=0}^{\infty}\mathcal{P}(-,T,t_+,N). \quad (8)$$

Inserting Eqs. (6) and (7) into this expression and summing the series, we obtain

$$g(t_+|T) = P_-^{\text{eq}}e^{-k_+T}\delta(t_+) + P_+^{\text{eq}}e^{-k_+T}\delta(T-t_+) + e^{-r}\left[\left(\frac{k_-}{t_+}+\frac{k_+}{(T-t_+)}\right)\frac{z}{k_T}I_1(2z) + \frac{2k_+k_-}{k_T}I_0(2z)\right], \quad (9)$$

where r is the same as in Eqs. (7), $z=\sqrt{k_-t_+k_+(T-t_+)}$, and $I_0(z)$ and $I_1(x)$ are modified Bessel functions. Notice that this result can be obtained in a less direct way by means of the Anderson formalism [16]. Note also that the propagators can be obtained by an integration over t_+ of the various terms contributing to $g(t_\pm|T)$.

In Fig. 2 we plot the function $g(t_+|T)$ for several values of the parameters, and we compare it with simulations of the master equations. In all cases studied, there is perfect agreement between the simulations and the present analytic result.

An interesting limit of Eq. (9) is that of large T. Using the asymptotic expression $\lim_{x\to\infty}I_\nu(x)=\exp(x)/\sqrt{2\pi x}$, we see that the leading term in $1/T$ is

$$g(t_+|T) \approx \sqrt{\frac{2}{\pi z}}\frac{k_+k_-}{k_T}e^{-[\sqrt{(k_-t_+)}-\sqrt{k_+(T-t_+)}]^2}, \quad (10)$$

which has exponential tails expected from large-deviation arguments [17].

Another issue that can be addressed in this framework is the probability $h(N)$ of observing precisely N transitions in a time interval of T. This is simply

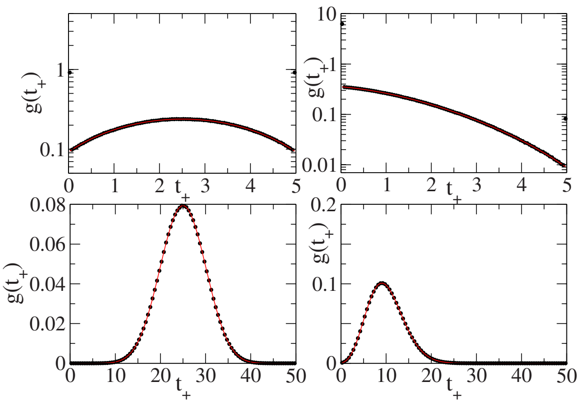

FIG. 2. (Color online) Comparison of a simulation and Eq. (9) for the function $g(t_+|T)$. The parameters are $T=5$ (top figures) and 50 (bottom figures). The rates are $k_+=k_-=0.5$ (left figures) and $k_-=0.8$ and $k_+=0.2$ (right figures). Lines (red online) are the analytic curves; the black points are averages over 10^7 simulations of the master equation. Notice the effect of the δ functions (first and last points) in the top figures, where the probabilities are on a logarithmic scale. (We do not plot the δ functions in the analytic curves.)

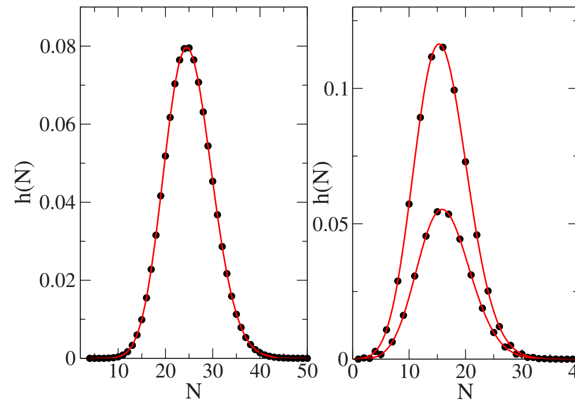

FIG. 3. (Color online) Probabilities $h(N)$ of observing N transitions in a time $T=50$. Transition rates are (left) $k_+=k_-=0.5$ and $k_+=0.2$ and $k_-=0.8$ (right). The points represent statistics collected over 10^7 simulations. The solid lines (red online) are the analytic results of Eqs. (12) and (13). The left figure corresponds to the symmetric limit with $k_+=k_-=0.5$ in which both distributions collapse into a Poisson distribution. Notice the even-odd asymmetry in the right figure.

$$h(N) = \int_0^T dt_+ [P_-^{\mathrm{eq}} \mathcal{P}(-,T,t_+,N) + P_+^{\mathrm{eq}} \mathcal{P}(+,T,t_+,N)]. \tag{11}$$

Using the above expressions for the various terms and performing the integral, we find two different expressions, one for N odd,

$$h(N) = \frac{2\sqrt{\pi}(k_+k_-)^{(N+1)/2}}{[(N-1)/2]!k_T}\left(\frac{T}{k_- - k_+}\right)^{N/2} e^{-(k_+ + k_-)T/2} I_{N/2}(\zeta), \tag{12}$$

and one for N even,

$$h(N) = \frac{\sqrt{\pi}T(k_-k_+)^{(N/2)}}{2k_T(N/2)!}\left(\frac{T}{k_- - k_+}\right)^{(N-1)/2} e^{-(k_+ + k_-)T/2} \times [(k_- + k_+)I_{(N-1)/2}(\zeta) + (k_- - k_+)I_{(N+1)/2}(\zeta)], \tag{13}$$

with $\zeta=(k_- - k_+)T/2$.

In evaluating the two expressions above, care must be taken to pick up the proper branch of the half-integer powers according to the requirement that the function $h(N)$ should be real and positive. In Fig. 3 we compare the distribution $h(N)$ with simulations of the master equation. Here, too, perfect agreement is found. The left panel shows a symmetric case with $k_+=k_-=0.5$ for which Eqs. (12) and (13) each have as limit a Poisson distribution, $h(N)=\lambda^N e^{-\lambda}/N!$ with $\lambda=T/k_+ =T/k_-$. The right panel, for the case $k_+=0.2$ and $k_-=0.8$, is less trivial. The asymmetry in the rates is reflected in a difference between the distributions for N even and N odd. This corresponds to the physical fact that one of the states is short lived and the other long lived, so that is more likely to observe an even number of transitions. In the asymmetric case, accurate numerical studies indicate that the average number of transitions is $\bar{N}=T/\tilde{k}$ with $\tilde{k}=k_+k_-/[2(k_+ + k_-)]$.

In summary, we have shown that the probability distributions associated with the trajectories of master equations can be expressed in general as a product over single-state properties. This can be particularly useful for systems composed of a few states as demonstrated by the exact determination of several statistical quantities of two-state master equations for which results can be expressed simply in terms of modified Bessel functions. While the methods presented here can be applied to the evaluation of individual trajectories in more complex problems, summation over all trajectories becomes increasingly difficult as the number of states increases. If, however, almost all rates are small, the dynamics of the system can be dominated by a relatively small number of trajectories. For example, this is often the case in chemical kinetics, where average reaction paths may be well defined even for high-dimensional dynamics [18]. In such cases, our methods could provide a way to detect these dominant trajectories and to assess their probabilities.

We would like to thank E. Barkai for pointing out relevant references. S.P. wishes to thank J. Ferkinghoff-Borg, J. Fonslet, M. H. Jensen, and S. Krishna for stimulating discussion.

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
[2] J. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).
[3] M. Esposito, U. Harbola, and S. Mukamel, Phys. Rev. E **76**, 031132 (2007).
[4] R. J. Harris and G. M. Schutz, J. Stat. Mech.: Theory Exp. (2007) P07020.
[5] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin 1983).
[6] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[7] S. Pigolotti and A. Vulpiani, J. Chem. Phys. **128**, 154114 (2008).
[8] See, e.g., C. Itzykson and J. M. Drouffe, *Statistical Field Theory* (Cambridge University Press, Cambridge, U.K., 1989), Vol. 1.
[9] L. Peliti, J. Phys. (Paris) **46**, 1469 (1985).
[10] J. Cardy, in *The Mathematical Beauty of Physics*, edited by J. Drouffe and J. B. Zuber (World Scientific, Singapore, 1997).
[11] G. Margolin and E. Barkai, Phys. Rev. Lett. **94**, 080601 (2005).
[12] F. Shikerman and E. Barkai, Phys. Rev. Lett. **99**, 208302 (2007).
[13] G. Bonnet, O. Krichevsky, and A. Libchaber, Proc. Natl. Acad. Sci. U.S.A. **95**, 8602 (1998).
[14] R. Zwanzig, Proc. Natl. Acad. Sci. U.S.A. **94**, 148 (1997).
[15] F. Ritort, C. Bustamante, and I. Tinoco, Proc. Natl. Acad. Sci. U.S.A. **99**, 13544 (2002).
[16] A. M. Berezhkovskii, A. Szabo, and G. H. Weiss, J. Chem. Phys. **110**, 9145 (1999).
[17] H. Touchette, e-print arXiv:0804.0327v1.
[18] D. M. Zuckermann and T. B. Woolf, J. Chem. Phys. **111**, 9475 (1999).