

Asymptotics of rare events in birth–death processes bypassing the exact solutions

Charles R Doering¹, Khachik V Sargsyan¹, Leonard M Sander² and
Eric Vanden-Eijnden³

¹ Department of Mathematics and Michigan Center for Theoretical Physics, University of Michigan, Ann Arbor, MI 48109-1043, USA

² Department of Physics and Michigan Center for Theoretical Physics, University of Michigan, Ann Arbor, MI 48109-1120, USA

³ Courant Institute, New York University, New York, NY 10012, USA

E-mail: doering@umich.edu, ksargsya@umich.edu, lsander@umich.edu and eve2@cims.nyu.edu

Received 4 September 2006

Published 22 January 2007

Online at stacks.iop.org/JPhysCM/19/065145

Abstract

We investigate the near-continuum asymptotics of mean first passage times in some one-variable birth–death processes. The particular problem we address is how to extract mean first passage times in the near-continuum limit from their defining finite-difference equations alone. For the simple class of processes we consider here, exact closed-form solutions for the mean first passage time between any two states are available and the near-continuum expansion of these formulae defines the correct limiting behaviour and is used to check the results of asymptotic analysis of the difference equations. We find that in some cases the asymptotic approach does not lead unequivocally to the proper result.

1. Introduction

The theoretical transition from discrete to continuous dynamical descriptions of particle systems is a classical problem of continuing—even increasing—interest in view of current developments in materials science, biology and nanotechnology. Among the fundamental phenomena of importance is the emergence of long ‘macroscopic’ timescales from an ensemble of particles evolving on relatively fast ‘microscopic’ timescales. Relaxation rates and mean first passage times are familiar examples of such quantities [1] that enter into the next level of modelling of the bulk kinetics, often in terms of deterministic rate equations. These quantities are of interest in other contexts as well, for example in population biology and epidemiology where the evolution of a large collection of individuals or, say, an infection within the group, is often more easily studied by modelling the population as a continuum.

These issues have a long history in physics, chemistry and biology. For example Van Kampen [2] introduced the $1/\Omega$ expansion to recover some aspects of the dynamics and

associated fluctuations of models expressed in terms of master equations (continuous-time discrete state space Markov processes), in particular birth–death processes with a focus on chemical reactions. The product of that analysis is a Markov diffusion process (continuous time, continuous state space) characterized by a Fokker–Planck equation or, equivalently, a related stochastic differential equation that provides more detail than purely deterministic ‘mean field’ rate equations that neglect effects due to the underlying discreteness. Such analysis produces a reduced description of local relaxation toward, and fluctuations around, deterministically stable states. It can be modified to capture more global qualitative features like relative stability and extinction [3].

More recently it has been recognized that some quantitative features of the dynamics of microscopically discrete systems are incorrectly captured by the most straightforward quasi-continuum modelling via Markov diffusion process. One of the earliest examples in the physics literature was the analysis of transitions in discrete bistable systems by Hänggi *et al* [4], who observed that in the large- N (particle number) limit the ‘effective potential’ governing the exponentially large mean transition times is *not* the mean-field potential derived from the rate equations. On the other hand, the mean-field potential consistent with the rate equations is the one that governs the Fokker–Planck description. This reveals an irreconcilable discrepancy between the limiting behaviour of the timescale and the timescale predicted by the near-continuum limiting description. Even more recently, a subset of the present authors made a direct comparison of the asymptotic large- N expansion of the exact mean extinction time in a class of birth–death processes, explicitly displaying the discrepancy when the timescale is exponentially large [5].

In this paper we investigate the near-continuum asymptotics of mean first passage times in a class of discrete state space Markov processes, namely one-variable birth–death processes. The particular question we address is of how to extract relevant ‘macroscopic’ timescales in the near-continuum (large- N) limit from the dynamical description alone. For the simple class of processes that we consider here, exact—albeit complicated—closed-form solutions for the mean first passage time between any two states are available and the asymptotic expansion of these formulae defines the correct limiting behaviour. In more general cases, however, no such exact solutions are known and the analysis must proceed from the finite-difference equation for the mean first passage time derived from the defining master equation. Hence we develop the analysis for some simple one-dimensional problems using the exact solutions only as benchmarks for checking the assumptions *a posteriori*. Our fundamental observations are twofold: for some models familiar methods, namely asymptotic expansions and matching, may be employed to derive the proper large- N behaviour of mean first passage times in a relatively routine manner. But these methods applied to other models, including deterministically bistable systems, may be ambiguous in the sense that, while the ‘correct’ solution can be recovered, the analysis does not clearly lead to it as a unique prediction.

The asymptotic methods we use are familiar from the theory of large deviations for discrete state space Markov processes. These methods, in particular a WKB-like ansatz, are similar to those employed in large-deviation analysis for Markov diffusion processes described by Fokker–Planck equations. We work directly with the discrete state space processes, however, because the usual Fokker–Planck equations (and related stochastic differential equations) obtained by truncating Kramer–Moyal expansions of master equations do not always accurately describe the rare events associated with large deviations [6]. The basic problem that we encounter here is connected with the fact that there are not unique critical points for the functionals describing the paths of extremal probability, and the paths contributing most significantly to the mean first passage times between two points need not be the most probable trajectory connecting those two points. Also, as discussed further in the concluding section,

this problem is not limited to discrete state space systems: it can appear as well in the asymptotic analysis of mean first passage times for diffusion processes defined by Fokker–Planck equations.

The rest of this paper is organized as follows. In section 2 we illustrate these ideas for systems with one deterministically stable non-zero state but subject to extinction, i.e. with an absorbing state at zero population. In this case, the asymptotic analysis of the finite-difference equation is straightforward and is easily confirmed with the expansion of the exact solution. In section 3 we consider a deterministically bistable system. We first present the correct large- N behaviour of the mean first passage time from an arbitrary initial state to one or the other of the ‘equilibrium’ states by expanding the exact solution. Then we perform the asymptotic analysis of the difference equation finding that there is apparently not enough information on hand to uniquely select between the possible approximate solutions that may be constructed. The concluding discussion summarizes the results and remarks on the outlook for the resolution of this problem.

2. Discrete SIS epidemics model as a birth–death process

In this section we study a class of models that possess a unique deterministically stable steady state but are also subject to extinction, a rare event that does not occur in the continuum limit (i.e. for the population scale $N \rightarrow \infty$). We approach the near-continuum large- N approximate formula for the mean extinction time first via appropriate expansion of the exact formula, and then by asymptotic analysis of the finite-difference equation. We begin with a specific example that sets the stage and illustrates the role of the population scale N .

Consider the susceptible-infected-susceptible (SIS) model of epidemiology [7]. A population of size N within which n individuals suffer from an infection and the rest, $N - n$, are susceptible. Suppose the infection rate per contact is Λ/N and the recovery rate is unity (to fix the unit of time). A recovered individual immediately becomes susceptible. One can formulate this system as a birth–death process:

$$\begin{aligned} \lambda_n & n \rightarrow n + 1, \\ \mu_n & n \rightarrow n - 1, \end{aligned} \tag{1}$$

with rates

$$\begin{aligned} \lambda_n &= \Lambda n(1 - n/N), \\ \mu_n &= n. \end{aligned} \tag{2}$$

We consider the case $\Lambda > 1$, where there is a single stable equilibrium state $n_e = N(1 - 1/\Lambda)$ with $\lambda_{n_e} = \mu_{n_e}$. The results of the following sections apply both to this specific model and others that share these structural features.

2.1. Exact solutions for the mean extinction time

Since $\lambda_0 = \mu_0 = 0$, we have an absorbing state at $n = 0$. Although the state of the system is most likely to be found near the stable equilibrium, there is finite probability of so-called extinction (absorption to the zero state) of the disease (everyone is recovered).

We are interested in the mean extinction time τ_n from any state n . Also we will find large system size asymptotics of mean extinction time. Since we have a unique stable equilibrium, the mean extinction time will be an exponentially large quantity in N . We will try to find the exponent bypassing the exact solution. This technique can be crucial in the absence of exact solutions, e.g. in multi-dimensional systems or even one-dimensional ones with more complex structure than single-step jumps.

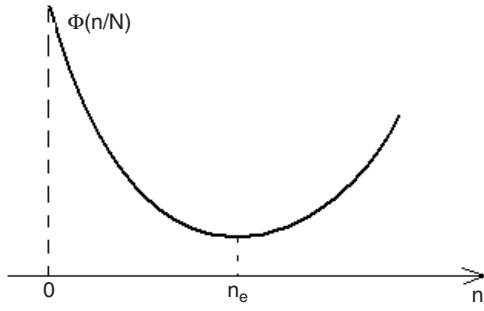


Figure 1. Effective potential for the SIS and related models. The birth and death rates are equal at n_e . Extinction corresponds to $n = 0$, an absorbing state.

The mean extinction time τ_n satisfies the backward inhomogeneous equation

$$-1 = \lambda_n \tau_{n+1} - (\lambda_n + \mu_n) \tau_n + \mu_n \tau_{n-1} \tag{3}$$

with boundary conditions $\tau_0 = 0$ and $\tau_r - \tau_{r-1} = \frac{1}{\mu_r}$ for some large, reflecting state r . The contributions from this reflection in the asymptotics below will be negligible. The equation above can be solved exactly (see, e.g., [5]):

$$\tau_n = \sum_{m=1}^n \left[\frac{1}{\mu_m} + \prod_{i=1}^{m-1} \frac{1}{\rho_i} \sum_{j=m+1}^r \frac{1}{\mu_j} \prod_{k=1}^{j-1} \rho_k \right] \tag{4}$$

where we have introduced the ratio of the rates $\rho_i = \lambda_i / \mu_i$.

2.2. Large system size asymptotics from the exact solution

Before stating the results of asymptotics of the sums above, we assume that both λ and μ involve N in a special way:

$$\begin{aligned} \lambda_n &= N \bar{\lambda}(x), \\ \mu_n &= N \bar{\mu}(x), \end{aligned} \tag{5}$$

where $x = n/N$ and $\bar{\lambda}, \bar{\mu}$ are smooth functions of x .

We also define $\rho_n = \bar{\rho}(n/N)$, where $\bar{\rho}(x)$ is a bounded, smooth, and non-negative function and the ‘effective potential’

$$\Phi(x) = - \int_0^x \log \bar{\rho}(\xi) d\xi \tag{6}$$

which, for $\Lambda > 1$, is a smooth function passing through the origin with a quadratic minimum at $x_e = n_e/N$; see figure 1.

The large system size expansion is [5]:

$$\tau_n = \left(1 - e^{-n \log \bar{\rho}(0)}\right) \sqrt{\frac{2\pi}{\bar{\lambda}(x_e) \bar{\mu}(x_e) \Phi''(x_e)}} \frac{\sqrt{\bar{\rho}(0)} e^{-N\Phi(x_e)}}{\bar{\rho}(0) - 1} \frac{1}{\sqrt{N}} \times (1 + \mathcal{O}(1/N)), \tag{7}$$

where the first factor is important in the boundary layer of thickness $\mathcal{O}(1)$ in n ($\mathcal{O}(1/N)$ in $x = n/N$). Figure 2 shows the behaviour of the leading-order exponent $V_n = \frac{1}{N} \log \tau_n$. It is constant, i.e. does not depend on n , except in the boundary layer.

2.3. Large system size asymptotics using exponential ansatz

In this example we had the exact solution (4) available. For more general problems, in particular higher-dimensional ones, there might not be a closed form for the mean extinction time. In

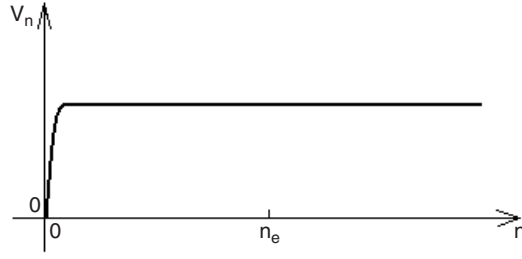


Figure 2. The leading-order exponent $V_n = \frac{1}{N} \log \tau_n$ does not depend on n , except in the boundary layer.

these cases, we can bypass the exact solution by putting the WKB-type exponential ansatz $\tau_n = \sigma_n e^{NV_n}$ with $V_n > 0$ into (3). The motivation for this is the fact that extinction is a rare event with exponentially small likelihood, hence the mean extinction time has to be exponentially large in N .

With this ansatz, (3) becomes

$$-1 = \lambda_n(e^{NV_{n+1}}\sigma_{n+1} - e^{NV_n}\sigma_n) + \mu_n(e^{NV_{n-1}}\sigma_{n-1} - e^{NV_n}\sigma_n). \quad (8)$$

Rewrite the above equation in the form

$$-e^{-NV_n} = \lambda_n(e^{N\Delta V_{n+1}}\sigma_{n+1} - \sigma_n) + \mu_n(e^{-N\Delta V_n}\sigma_{n-1} - \sigma_n), \quad (9)$$

where $\Delta V_n = V_n - V_{n-1} = \mathcal{O}(1/N)$, assuming smooth behaviour of the exponent V_n .

The left-hand side of (9) is exponentially small in N , hence the right-hand side vanishes up to the order $\mathcal{O}(1/N)$:

$$0 = \lambda_n(e^{N\Delta V_{n+1}}\sigma_{n+1} - \sigma_n) - \mu_n e^{-N\Delta V_n}(e^{N\Delta V_n}\sigma_n - \sigma_{n-1}), \quad (10)$$

which implies, up to the order $\mathcal{O}(1/N)$,

$$0 = (\lambda_n - \mu_n e^{-N\Delta V_n})(e^{N\Delta V_n}\sigma_n - \sigma_{n-1}). \quad (11)$$

For $n = \mathcal{O}(N)$, expand σ_n in the powers of N , assuming the leading order is N^α (with the exponent α to be found later): $\sigma_n = N^\alpha \sigma_n^0 + N^{\alpha-1} \sigma_n^1 + \mathcal{O}(N^{\alpha-2})$.

Equation (11) leads us to two possibilities for ΔV_n (one or the other factor must vanish):

$$\Delta V_n = 0 \quad (12)$$

or

$$\Delta V_n = \frac{1}{N} \ln \frac{\mu_n}{\lambda_n}. \quad (13)$$

The mean-field quadratic potential and associated rate equation dynamics suggests that, no matter where we start, the system quickly approaches the equilibrium and stays there before becoming extinct due to rare fluctuations. Hence the exponent in the mean extinction time should not depend on the starting position n and this suggests that we pick the first branch for all n . This selection does constitute a choice rather than a strict deduction, but the physical argument is reasonably compelling in this case. As will be seen in section 3, however, the selection process is much more subtle for the bistable system.

Now $\Delta V_n = 0$ or $V_n = C$, and (9) becomes

$$\lambda_n(\sigma_{n+1} - \sigma_n) + \mu_n(\sigma_{n-1} - \sigma_n) = -e^{-NC} \sim 0 \quad \text{as } N \rightarrow \infty. \quad (14)$$

Using the expansion of σ_n , we will arrive at the leading order $0 = (\lambda_n - \mu_n)(\sigma_{n+1}^0 - \sigma_n^0)$, implying $\sigma_n^0 = C_1 = \text{const}$. C_1 is non-vanishing in general, which is incompatible with the

boundary condition $\tau_0 = 0$. To get rid of this incompatibility, we need to consider the boundary layer $n = \mathcal{O}(1)$.

We can now solve $0 = \lambda_n(\sigma_{n+1} - \sigma_n) + \mu_n(\sigma_{n-1} - \sigma_n)$ to get

$$\sigma_n^{\text{inner}} = C_2 \sum_{m=1}^n \prod_{k=1}^{m-1} \frac{1}{\rho_k} \quad (15)$$

for some constant C_2 and $n \geq 2$, while $\sigma_1^{\text{inner}} = C_2$ and the ‘empty’ product (for $m = 1$) is taken to be equal to 1, here and further in this paper.

Next, the matching $\sigma_{n \rightarrow \infty}^{\text{inner}} = N^\alpha C_1$ leads to the relation between two constants C_1 and C_2 :

$$C_2 \sum_{m=1}^{\infty} \prod_{k=1}^{m-1} \frac{1}{\rho_k} = N^\alpha C_1. \quad (16)$$

In order to find one of the constants, let us get back to the equation for mean extinction time (3). We will multiply both sides by the ‘discrete integrating factor’ β_k (to be determined below) and sum over $k = 1$ to $r - 1$:

$$-\sum_{k=1}^{r-1} \beta_k = \sum_{k=1}^{r-1} \beta_k (\lambda_k(\tau_{k+1} - \tau_k) + \mu_k(\tau_{k-1} - \tau_k)). \quad (17)$$

In order to get cancellations in the right-hand side, we require $\beta_k \lambda_k = \gamma_{k+1}$ and $\beta_k \mu_k = \gamma_k$. This leads to $\beta_1 = 1$ and $\beta_k = \frac{\mu_1}{\mu_k} \prod_{i=1}^{k-1} \rho_i$, up to a constant, non-important factor.

Now (17) becomes, using the boundary conditions,

$$-\sum_{k=1}^{r-1} \beta_k = \sum_{k=1}^{r-1} \gamma_{k+1}(\tau_{k+1} - \tau_k) - \gamma_k(\tau_k - \tau_{k-1}) = \beta_r - \mu_1 \tau_1 \quad (18)$$

or, sending β_r to the left-hand side and dividing through by $-\mu_1$,

$$\sum_{k=1}^r \frac{1}{\mu_k} \prod_{i=1}^{k-1} \rho_i = \tau_1 = \sigma_1^{\text{inner}} e^{NC} = C_2 e^{NC} = \frac{N^\alpha C_1 e^{NC}}{\sum_{m=1}^{\infty} \prod_{k=1}^{m-1} \frac{1}{\rho_k}}. \quad (19)$$

Rewrite the above equation

$$N^\alpha C_1 e^{NC} = \left(\sum_{k=1}^r \frac{1}{\mu_k} \prod_{i=1}^{k-1} \rho_i \right) \sum_{m=1}^{\infty} \prod_{k=1}^{m-1} \frac{1}{\rho_k}. \quad (20)$$

Using Riemann sum approximation, as well as Gaussian approximation on exponential integrals, we get

$$\begin{aligned} \sum_{k=1}^r \frac{1}{\mu_k} \prod_{i=1}^{k-1} \rho_i &= \sum_{k=1}^r \frac{1}{\mu_k} \frac{1}{\sqrt{\bar{\rho}(0)\bar{\rho}(k/N)}} e^{-N\Phi(k/N)} \times (1 + \mathcal{O}(1/N)) \\ &= \int_{0^+}^{r/N} \frac{dz}{\sqrt{\bar{\rho}(0)\bar{\rho}(z)}\bar{\mu}(z)} e^{-N\Phi(z)} \times (1 + \mathcal{O}(1/N)) \\ &= \frac{1}{\sqrt{\bar{\rho}(0)\bar{\rho}(x_e)}\bar{\mu}(x_e)} \sqrt{\frac{2\pi}{N\Phi''(x_e)}} e^{-N\Phi(x_e)} \times (1 + \mathcal{O}(1/N)) \end{aligned} \quad (21)$$

and

$$\begin{aligned} \sum_{m=1}^{\infty} \prod_{k=1}^{m-1} \frac{1}{\rho_k} &= \sum_{m=1}^{\infty} \sqrt{\bar{\rho}(0)\bar{\rho}(m/N)} e^{N\Phi(m/N)} \times (1 + \mathcal{O}(1/N)) \\ &= \frac{\bar{\rho}(0)}{\bar{\rho}(0) - 1} \times (1 + \mathcal{O}(1/N)). \end{aligned} \quad (22)$$

In the last step we used $n = \mathcal{O}(N)$, $\Phi'(0) = -\ln \bar{\rho}(0) < 0$, $\Phi(0) = 0$ and the smoothness of $\bar{\rho}$ to compute the geometric sum after expanding $\Phi(m/N)$ around 0.

Finally, plug (21) and (22) into (20) to get

$$N^\alpha C_1 e^{NC} = \sqrt{\frac{2\pi}{\Phi''(x_e)\bar{\lambda}(x_e)\bar{\mu}(x_e)}} \frac{\sqrt{\bar{\rho}(0)} e^{-N\Phi(x_e)}}{\bar{\rho}(0) - 1} \times (1 + \mathcal{O}(1/N)) \quad (23)$$

for $n = \mathcal{O}(N)$. The left-hand side is the mean extinction time,

$$\tau_n = \sigma_n e^{NV_n} \sim N^\alpha C_1 e^{NC}, \quad (24)$$

so we deduce asymptotic formulae for the exponent $V_n = C$ and the prefactor $\sigma_n = N^\alpha C_1$:

$$C = -\Phi(x_e) > 0 \quad (25)$$

and

$$N^\alpha C_1 = \sqrt{\frac{2\pi}{\Phi''(x_e)\bar{\lambda}(x_e)\bar{\mu}(x_e)}} \frac{\sqrt{\bar{\rho}(0)}}{\bar{\rho}(0) - 1} \times \frac{1}{\sqrt{N}}. \quad (26)$$

This also fixes $\alpha = -1/2$.

In order to get a complete agreement with the asymptotics (7) of the exact solution, notice that, for $n = \mathcal{O}(1)$, the geometric sum in (22) is cut at n , resulting in an extra factor $1 - e^{n\Phi'(0)} = 1 - e^{-n \log \bar{\rho}(0)}$. The upshot is that we see that the analysis proceeding from the difference equation alone is capable of recovering the exact asymptotic behaviour bypassing the exact formula.

3. Discrete Schlögl model as a birth–death process

In this section we extend the analyses to birth–death processes describing deterministically bistable systems. Now there are three values of the population n where the birth and death rates coincide: two corresponding to deterministically stable steady states and one intermediate point corresponding to an unstable steady state. Again we begin by presenting a specific example, followed by a near-continuum expansion of the exact expression for the mean first passage time between any two points. We then turn to the asymptotic analysis of the mean first passage time finite-difference equation, in this case finding that this approach does not uniquely produce the correct result.

Consider the following chemical reaction, introduced by Schlögl [8] as a catalysis model:



Let n be the number of particles X . Then, keeping the concentrations of A and B fixed and of the order of the system volume N , we are interested in the evolution of n . k_i are the rates of the corresponding reactions in (27).

The reaction (27) also can be rewritten as a birth–death process with the following rates:

$$\begin{aligned} \lambda_n & \quad n \rightarrow n + 1, \\ \mu_n & \quad n \rightarrow n - 1. \end{aligned} \quad (28)$$

Here the birth and death rates are, correspondingly,

$$\lambda_n = \frac{k_1}{N} n(n-1) + k_4 N \quad \text{and} \quad \mu_n = \frac{k_2}{N^2} n(n-1)(n-2) + k_3 n. \quad (29)$$

Notice that, without loss of generality, we absorbed the concentrations of particles A and B into the system volume N .

For a particular choice of parameters, there are two stable equilibrium values n_1, n_2 , and an unstable one, n_3 . For these states, $\lambda_{n_i} = \mu_{n_i}, i = 1, 2, 3$.

3.1. Exact solutions of mean first passage times

In this bistable system, the question of relative stability arises—namely, which stable equilibrium state is ‘more’ stable: that is, if we take a ‘snapshot’ of the system at any particular time, near which of the stable equilibria it is more likely to be found.

The relative stability question can be answered by comparing the mean first passage times from n_1 to n_2 and vice versa. Without loss of generality, we analyse the first passage to the first equilibrium, n_1 . Denote by τ_n the mean first passage time from a state n to n_1 .

As before, τ_n solves the equation

$$-1 = \lambda_n \tau_{n+1} - (\lambda_n + \mu_n) \tau_n + \mu_n \tau_{n-1} \tag{30}$$

with modified rates $\lambda_{n_1} = \mu_{n_1} = 0$ and the boundary conditions $\tau_{n_1} = 0, \tau_r - \tau_{r-1} = \frac{1}{\mu_r}$. Eventually, we will take r to be very large, and the contributions from this reflection will vanish.

The solution of (30) is

$$\tau_n = \sum_{m=n+1}^{n_1} \left[\frac{1}{\lambda_0} \prod_{k=1}^{m-1} \frac{1}{\rho_k} + \sum_{i=2}^{m-1} \frac{1}{\mu_i} \prod_{k=i}^{m-1} \frac{1}{\rho_k} \right] \tag{31}$$

for $n < n_1$, and

$$\tau_n = \sum_{m=n+1}^n \left[\frac{1}{\mu_m} + \prod_{i=1}^{m-1} \frac{1}{\rho_i} \sum_{j=m+1}^r \frac{1}{\mu_j} \prod_{k=1}^{j-1} \rho_k \right] \tag{32}$$

for $n > n_1$.

3.2. Large system size asymptotics from the exact solution

As in the previous section, we will introduce the potential function (6), which in this case is a double-well function with quadratic extrema at $x_i = n_i/N$, for $i = 1, 2, 3$, the only three solutions of $\bar{\lambda}(x) = \bar{\mu}(x)$. The asymptotic $N \rightarrow \infty$ behaviour of the exact solutions (31) and (32) is derived using standard techniques [2, 5, 9]:

$$\tau_n \sim \begin{cases} \int_{n/N}^{n_1/N} \frac{\bar{\rho}(\xi)}{\bar{\mu}(\xi) \log \bar{\rho}(\xi)} d\xi \times (1 + \mathcal{O}(1/N)) & \text{for } n \in [0, n_1] \\ \frac{(\frac{\pi}{2})^{\frac{3}{2}}}{\sqrt{\bar{\lambda}'(n_1/N) \bar{\mu}'(n_1/N) \Phi''(n_1/N)}} \sqrt{N} + \mathcal{O}(\log N) & \text{for } n \in [n_1, n_s] \\ \frac{1}{\bar{\lambda}(n_2/N)} \sqrt{\frac{2\pi}{\Phi''(n_2/N)}} \frac{\sqrt{\bar{\rho}(n/N)} e^{N(\Phi(n/N) - \Phi(n_2/N))}}{e^{-\Phi'(n/N)} - 1} \frac{1}{\sqrt{N}} \times (1 + \mathcal{O}(1/N)) & \text{for } n \in [n_s, n_3] \\ \frac{2\pi}{\bar{\lambda}(n_2/N) |\Phi''(n_2/N) \Phi''(n_3/N)|} \frac{e^{N(\Phi(n_3/N) - \Phi(n_2/N))}}{N} \times (1 + \mathcal{O}(1/N)) & \text{for } n \in [n_3, r]. \end{cases} \tag{33}$$

Here n_s is, if any, the point in the first well characterized by $\Phi(n_s/N) = \Phi(n_2/N)$ (if the second well is deeper than the first one, we take $n_s = n_1$ and are left with three regions only).

The leading order of the exponent $V_n = \frac{1}{N} \log \tau_n$ is not constant anymore—there is a layer $[n_s, n_3]$, non-dependent on N , where V_n is not constant; see figure 4.

3.3. Large system size asymptotics using exponential ansatz

As can be seen from (33), the mean first passage time is exponential in the last two regions. Hence, similar to the SIS case, we would like to find the correct exponents and the prefactors from an ansatz

$$\tau_n = \sigma_n e^{NV_n} \quad (34)$$

with $V_n > 0$.

Plugging the exponential ansatz into the equation for the mean first passage time (3), we arrive again at the two possibilities

$$\Delta V_n = 0 \quad (35)$$

or

$$\Delta V_n = \frac{1}{N} \ln \frac{\mu_n}{\lambda_n}. \quad (36)$$

The selection here is not as trivial as in the SIS case. As can be seen from the exact solutions' asymptotics (33), for $n \in [n_s, n_3]$ we need to pick

$$\Delta V_n = \frac{1}{N} \ln \frac{\mu_n}{\lambda_n} \quad (37)$$

in order to get a non-constant exponent that varies with n . This is not really satisfactory, since the main reason for using the exponential ansatz (34) was to avoid the exact solution. In section 4 we will discuss this ambiguity problem in more detail.

Let us rewrite (10) one order further than in (11):

$$0 = (\lambda_n - \mu_n e^{-N\Delta V_n})(e^{N\Delta V_n} \sigma_n - \sigma_{n-1}) + [\lambda_n(e^{N\Delta V_{n+1}} \sigma_{n+1} - \sigma_n) - \lambda_n(e^{N\Delta V_n} \sigma_n - \sigma_{n-1})]. \quad (38)$$

The first summand vanishes because of the chosen ΔV_n , hence the second summand in square brackets vanishes too. This implies that $e^{N\Delta V_n} \sigma_n - \sigma_{n-1} = C = \text{const}$.

Again, expand σ_n in the powers of N : $\sigma_n = N^\alpha \sigma_n^0 + N^{\alpha-1} \sigma_n^1 + \mathcal{O}(N^{\alpha-2})$. Then the equation for the leading order is

$$\sigma_n^0 = \frac{N^{-\alpha} C}{e^{N\Delta V_n} - 1} = \frac{N^{-\alpha} C}{\frac{\mu_n}{\lambda_n} - 1}. \quad (39)$$

The exponent is

$$V_n = D + \sum_{m=n_s+1}^n \Delta V_m = D + \frac{1}{N} \sum_{m=n_s+1}^n \ln \frac{\mu_m}{\lambda_m} \quad (40)$$

with a constant $D = V_{n_s}$ to be found. Now the leading order for the mean first passage time τ_n is

$$\frac{N^{-\alpha} C}{\frac{\mu_n}{\lambda_n} - 1} \exp\left(ND + \sum_{m=n_s+1}^n \ln \frac{\mu_m}{\lambda_m}\right) = \frac{N^{-\alpha} C}{\frac{\mu_n}{\lambda_n} - 1} \sqrt{\frac{\rho_n}{\rho_{n_s}}} e^{ND+N(\Phi(n/N)-\Phi(n_s/N))}. \quad (41)$$

In order to find the constants C and D , we multiply both sides by the ‘discrete integrating factor’ β_k and sum over $k = n_s + 1$ to $r - 1$:

$$- \sum_{k=n_s+1}^{r-1} \beta_k = \sum_{k=n_s+1}^{r-1} \beta_k (\lambda_k (\tau_{k+1} - \tau_k) + \mu_k (\tau_{k-1} - \tau_k)). \quad (42)$$

To get cancellations in the right-hand side, we require $\beta_k \lambda_k = \gamma_{k+1}$ and $\beta_k \mu_k = \gamma_k$. This leads to $\beta_{n_s+1} = 1$ and $\beta_k = \frac{\mu_{n_s+1}}{\mu_k} \prod_{i=n_s+1}^{k-1} \rho_i$, up to a constant, non-important factor.

With the chosen β_k , (42) becomes

$$- \sum_{k=n_s+1}^{r-1} \beta_k = \beta_r - \mu_{n_s+1} (\tau_{n_s+1} - \tau_{n_s}) \quad (43)$$

or

$$- \sum_{k=n_s+1}^r \beta_k = -\mu_{n_s+1} e^{NV_{n_s}} (\sigma_{n_s+1} e^{N\Delta V_{n_s+1}} - \sigma_{n_s}). \quad (44)$$

Plugging in the expressions for β_k , as well as for the leading-order prefactor (39) and the exponent (40), we get

$$- \sum_{k=n_s+1}^r \frac{\mu_{n_s+1}}{\mu_k} \prod_{i=n_s}^{k-1} \rho_i = -\mu_{n_s+1} e^{ND} N^{-\alpha} C. \quad (45)$$

The asymptotics of the left-hand side

$$- \sum_{k=n_s+1}^r \frac{\mu_{n_s+1}}{\mu_k} \prod_{i=n_s}^{k-1} \rho_i = -\frac{\mu_{n_s+1}}{\mu_k} \sum_{k=n_s+1}^r \sqrt{\frac{\rho_{n_s}}{\bar{\rho}(k/N)}} e^{N\Phi(n_s/N) - N\Phi(k/N)} \times (1 + \mathcal{O}(1/N)) \quad (46)$$

$$= -\mu_{n_s+1} \sqrt{\rho_{n_s}} \int_{n_s/N}^{r/N} \frac{dz}{\sqrt{\bar{\rho}(z)} \bar{\mu}(z)} e^{N\Phi(n_s/N) - N\Phi(z)} \times (1 + \mathcal{O}(1/N)) \quad (47)$$

$$= -\frac{\mu_{n_s+1} \sqrt{\rho_{n_s}}}{\mu_{n_2} \sqrt{\rho_{n_2}}} \sqrt{\frac{2\pi}{N\Phi''(n_2/N)}} \times (1 + \mathcal{O}(1/N)) \quad (48)$$

leads to the expressions for the constants (note that $\mu_{n_2} = \lambda_{n_2}$)

$$C = \frac{\sqrt{\rho_{n_s}}}{\lambda_{n_2}} \sqrt{\frac{2\pi}{\Phi''(n_2/N)}}, \quad \alpha = 1/2 \quad \text{and} \quad D = 0. \quad (49)$$

Plugging the values of C and D into the mean first passage time leading order (41), we get

$$\frac{1}{\lambda_{n_2}} \sqrt{\frac{2\pi}{N\Phi''(n_2/N)}} \frac{\sqrt{\rho_n}}{\frac{\mu_n}{\lambda_n} - 1} e^{N(\Phi(n/N) - \Phi(n_2/N))} \quad (50)$$

which coincides with the asymptotics (33) of the exact solution.

Similarly, for the last region $n \in [n_3, r]$ we pick the other increment $\Delta V_n = 0$, which will lead to a constant prefactor and a constant exponent (see the calculations for the SIS case) that coincide with the exact solution’s asymptotics (33).

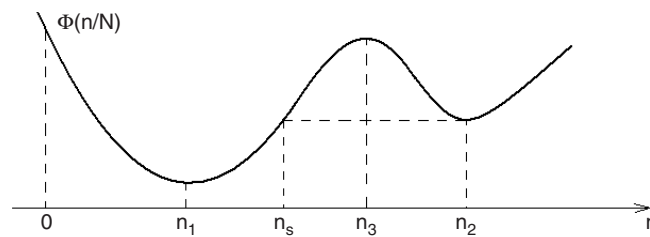


Figure 3. Effective potential for a deterministically bistable system. The state at $n = 0$ is reflecting in the case of interest here.

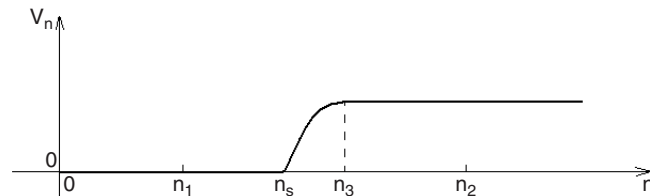


Figure 4. The leading-order exponent $V_n = \frac{1}{N} \log \tau_n$ depends on n . It vanishes for $n \leq n_s$ and is a constant for $n \geq n_3$. The variation between n_s and n_3 is *not* a boundary layer, i.e. its thickness is $\mathcal{O}(N)$ as $N \rightarrow \infty$.

4. Discussion

We have studied two classes of birth–death models defined by one-step Markov jump processes, the SIS model from epidemiology and the Schlögl model from chemical kinetics. For these benchmark models, near-continuum (large system size) asymptotics of certain rare events’ mean waiting times are found exactly as well as using large deviation techniques, namely a WKB-type exponential ansatz. For non-single step processes, as well as multi-dimensional models, the exact solutions may not be available, highlighting the importance of the WKB method that bypasses the exact solution.

The mean-field potential for the first (SIS) model is a quadratic function, while the second (Schlögl model) example displays a double-well potential. In the SIS model, the mean first passage time from initial state n to the extinct 0-state, asymptotically proportional to e^{NV_n} , has a constant (as regards n) exponent V_n in the large- N asymptotic limit [10]. On the other hand, in the Schlögl model, the mean first passage time to one of the deterministically stable states (say, n_1) has interesting behaviour as a function of n . It is exponentially large in N if there is a potential maximum to overcome, namely from the region $[n_3, +\infty)$ in figure 3, but it is also exponentially large in the region $[n_s, n_3]$ even though there is a deterministic path ‘downhill’ from $n \in [n_s, n_3]$ to n_1 . Such a deterministic path connecting the initial and final states usually coincides with the most probable path connecting these states. We find in this case, however, that it does *not* coincide with the path that contributes most to the mean first passage time. The direct deterministic path adds nothing to the exponent of the mean first passage time asymptotics, while a trajectory that first goes to the second well (with probability exponent $N(\Phi(n/N) - \Phi(n_3/N))$) makes an exponentially large contribution (with exponent $N(\Phi(n_3/N) - \Phi(n_2/N))$) which results in the correct, positive exponent $N(\Phi(n/N) - \Phi(n_2/N))$.

The exact solution’s asymptotic expansion (33) reveals this fact, but the method that uses the exponential ansatz has an ambiguity built into it: it is not obvious *a priori* how to pick

the correct branch of ΔV_n . The various branches evidently correspond to critical points of the relevant large-deviation functional (the ‘action’) but, unlike the problem of determining the most probable path connecting two states, there is no discernable selection criteria for mean first passage times analogous to the ‘minimum action’ principle. Indeed, it has generally been assumed that the correct mean first passage time may be recovered by restricting attention to the most probable path(s) [11]. (On the other hand, it has recently been realized that the most probable path connecting two states, even in one dimension, may not be the most direct path [12].) In order to resolve this quandary, an appropriate selection rule must be identified. At the present time we do not know what criterion generally determines the most dominant path(s) that determine the asymptotic behaviour of the mean first passage times.

We close by noting that the same issues regarding the ambiguity of the standard WKB-type asymptotic analysis arise for the mean first passage time problems associated with the Fokker–Planck operator. Suppose that $U(x)$ is a bistable potential, as in figure 3, with minima at x_1 and x_2 and a local maximum in between at x_3 . Let the diffusion process $X(t)$ satisfy the stochastic differential equation

$$dX = -U'(X) dt + \sqrt{2\epsilon} dW \quad (51)$$

with reflection at $x = 0$. Then the mean first passage time $\tau(x, x_f)$ from x to x_f satisfies

$$-1 = -U'(x) \frac{\partial \tau}{\partial x} + \epsilon \frac{\partial^2 \tau}{\partial x^2} \quad (52)$$

and $\tau(x_f, x_f) = 0$. The exact solution by quadrature of $\tau(x, x_f)$ is straightforward [1, 9] and its routine $\epsilon \rightarrow 0$ asymptotic expansion shows that $\tau(x, x_1)$ may be $\sim e^{C/\epsilon}$ with $C \neq 0$ for some $x \in (x_1, x_3)$, indicating that, in these cases too, the most direct path is not necessarily the one that determines the mean first passage time.

Acknowledgments

This work was supported in part by US National Science Foundation awards DMS-0244419 and DMS-0553487 (CRD, KVS, LMS), and DMS-0209959 and DMS-0239625 (EVE). Three of the authors (CRD, KVS, EVE) acknowledge the hospitality of the Geophysical Fluid Dynamics Program at Woods Hole Oceanographic Institution, where part of this work was initiated.

References

- [1] Redner S 2001 *A Guide to First-Passage Processes* (Cambridge: Cambridge University Press)
- [2] van Kampen N G 1992 *Stochastic Processes in Physics and Chemistry* (Amsterdam: North-Holland)
- [3] Horsthemke W and Brenig L 1977 Non-linear Fokker–Planck equation as an asymptotic representation of the master equation *Z. Phys. B* **27** 341–8
- [4] Hänggi P, Grabert H, Talkner P and Thomas H 1984 Bistable systems: master equation versus Fokker–Planck modelling *Phys. Rev. A* **29** 371–8
- [5] Doering C R, Sargsyan K V and Sander L M 2005 Extinction times for birth–death processes: exact results, continuum asymptotics, and the failure of the Fokker–Planck approximation *SIAM J. Multiscale Modeling Sim.* **3** 283–99
- [6] Shwartz A and Weiss A 1995 *Large Deviations for Performance Analysis* (London: Chapman and Hall)
- [7] Jacquez J A and Simon C P 1993 The stochastic SI model with recruitment and deaths.1. Comparison with the closed SIS model *Math. Biosci.* **117** 77–125
- [8] Schlögl F 1972 Chemical reaction models for non-equilibrium phase transitions *Z. Phys.* **253** 147–61
- [9] Gardiner C W 1983 *Handbook of Stochastic Methods for Physics, Chemistry* (Berlin: Springer)
- [10] Nasell I 2001 Extinction and quasi-stationarity in the Verhulst logistic model *J. Theor. Biol.* **21** 11–27
- [11] Ludwig D 1975 Persistence of dynamical systems under random perturbations *SIAM Rev.* **17** 605–40
- [12] Soskin S M 2006 Most probable transition path in an overdamped system for a finite transition time *Phys. Lett. A* **353** 281–90