Unified perspective on exponential tilt and bridge algorithms for rare trajectories of discrete Markov processes

Javier Aguilar [1,2,*] and Riccardo Gatto [3,†]

[1] Investigador ForInDoc del Govern de les Illes Balears en el departamento de Electromagnetismo y Física de la Materia e Instituto Carlos I de Física Teórica y Computacional, Universidad de Granada, Granada E-18071, Spain

[2] Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), Campus UIB, 07122 Palma de Mallorca, Spain

[3] Institute of Mathematical Statistics and Actuarial Science, University of Bern, Alpeneggstrasse 22, 3012 Bern, Switzerland

(Received 24 July 2023; revised 27 November 2023; accepted 6 February 2024; published 8 March 2024)

This article analyzes and compares two general techniques of rare event simulation for generating paths of Markov processes over fixed time horizons: exponential tilting and stochastic bridge. These two methods allow us to accurately compute the probability that a Markov process ends within a rare region which is unlikely to be attained. Exponential tilting is a general technique for obtaining an alternative or tilted sampling probability measure, under which the Markov process becomes likely to hit the rare region at terminal time. The stochastic bridge technique involves conditioning paths towards two endpoints: the terminal point and the initial one. The terminal point is generated from some appropriately chosen probability distribution that covers well the rare region. We show that both methods belong to the class of importance sampling procedures by providing a common mathematical framework of these two conceptually different methods of sampling rare trajectories. We also conduct a numerical comparison of these two methods, revealing distinct areas of application for each Monte Carlo method, where they exhibit superior efficiency. Detailed simulation algorithms are provided.

DOI: 10.1103/PhysRevE.109.034113

I. INTRODUCTION

Rare trajectories account for realizations of stochastic processes that are extremely unlikely to happen, in the sense that conventional methods to sample stochastic trajectories, like the Euler-Maruyama methods in the context of stochastic differential equations [1,2] or the Gillespie algorithm for pure jumping processes [3,4], mostly fail to provide satisfactory statistical characterization within affordable times. Also, rare paths usually occur in particular shapes and unveil spatiotemporal patterns [5,6]. Even if such paths are uncommon, they can have a decisive role in nature: catastrophic events such as extinction of species [7,8], extreme rainfalls [9], or earthquakes [10] manifest themselves through rare fluctuations. The estimation of improbable events holds significance across various scientific domains. These include scenarios such as determining the probability for a medical therapy efficacy falling below a low threshold [11], the level of a dam exceeding a certain (very small or very large) threshold [12], the capital of an insurance company falling below zero [13], or cosmic radiation corrupting memory cells in silicon microchips [14]. Therefore, the challenge of generating unlikely paths by using conventional methods, coupled with their significance in numerous phenomena, motivates the need for sophisticated algorithms for sampling rare paths.

The literature contains numerous Monte Carlo algorithms that thwart the rarity of the event to simulate, thus making it possible to control the simulation error. The original reference in the statistical literature on importance sampling by exponential tilting for first passage times of stochastic processes is [15]; see also [1], pp. 164–166. Some methods of rare event simulation rely on the generation of the bridge process, which conditions paths to the endpoints and associates rare trajectories with the occurrence or rare pairs of endpoints [16–22]. Other algorithms like cloning, also called splitting [23–26], or like Metropolis schemes [27,28] link rare trajectories to the occurrence of rare sample averages of integrated quantities. A survey of simulation of rare events in queueing and reliability models can be found in [29]. Following a theoretical approach, [30] provides a review of central concepts of rare event simulation for light- and heavy-tailed systems. A recent presentation of importance sampling in the context of large deviations theory is given by [31]. Then [32] discusses variational formulations of the thermodynamic free energy within the framework of importance sampling and [33] addresses the problem of automated search for optimal importance sampling schemes by using recent ideas from deep learning.

The large variety of methods makes it difficult to assess which is the best strategy to tackle the generation of rare paths in specific problems. Moreover, it is not always simple to convey to what extent the available Monte Carlo algorithms are fundamentally different or simply different expressions of the same mathematical framework. The present article shows that two families of methods, those relying on exponential tilting and on bridge processes, can be understood within the common framework of importance sampling. This article focuses on their analytical and numerical comparison. These two methods are relevant as they encompass many strategies

*Corresponding author: javieraguilar@ifisc.uib-csic.es

†riccardo.gatto@unibe.ch

for sampling rare paths in nonequilibrium processes within finite time horizons. The present article does not tackle the theory of asymptotic optimality of estimation of rare event probabilities,[1] which relies on the theory of large deviations; see, e.g., [35] for a general reference. Nor does it cover methods that do not use changes of measure, such as the forward flux sampling [36,37], the transition path sampling [38–40], or the saddlepoint technique, which provides the deterministic most likely path in the limit of small noise [41,42].

More precisely, the scope of this article is the following. The first and main objective is to show that these two conceptually different techniques can be re-expressed in terms of the same change-of-measure formula for the expected value. Thus, the two algorithms differ only in their likelihood ratio, namely, the Radon-Nikodym derivative, which is the factor that accounts for the replacement of the original sampling probability measure by a new one. The second objective of this work is to compare the numerical performance of these two techniques in various settings. We provide numerical evidence that in real situations with finite time horizons, the relative errors of backtracking and exponential tilting appear bounded. This provides a surrogate for theoretical results of asymptotic optimality. For both methods, we numerically show that there is an optimal parameter minimizing the relative error. Finally, we find that stochastic bridges provide superior suitability for tackling challenges associated with rare transitions between metastable states.

As a byproduct of demonstrating the precise emergence of these two methods from the technique of change of measure, we establish that stochastic bridge techniques may introduce systematic errors in the estimation of averages, which is a unique theoretical result within the field. Furthermore, we obtain a fairly complete tutorial on these two techniques of rare event simulation, since we offer readily applicable algorithms that streamline the implementation of these concepts into computer programs. For the sake of simplicity, many measure-theoretic details are given separately in footnotes.

The rest of the article has the following structure. Section II reviews the general theory of change-of-measure for stochastic processes and in particular for Markov processes with discrete time and state spaces. In Secs. III and IV we recast exponential tilting and bridge change-of-measure. Numerical applications of these two Monte Carlo techniques to a simple binomial process and to a process with metastable states are provided in Sec. V. Concluding remarks are presented in Sec. VI. Supplemental Material (SM) [43], Sec. SM 1, shows explicitly the Monte Carlo algorithms introduced in this article.

II. CHANGE OF MEASURE AND LIKELIHOOD RATIO PROCESS

This section provides a succinct introduction to the theory of change of measure. We refer readers to [1,34,44,45] for a more in-depth introduction to the topic. We start with the case of the single random variable and we then generalize this to the stochastic process.

A. Random variable

One of the central ideas in rare event simulation is the change of the sampling measure, which allows transforming the problem of estimating averages over some probability measure P to another average estimation over a different measure $\tilde{\mathsf{P}}$. The idea is that suitable transformations of this kind can ease the estimation of averages. Let us present the basic result of the theory. Let $(\Omega, \mathcal{F}, \tilde{\mathsf{P}})$ be a probability space [46]. Assume that the random variable L over this space is nonnegative with $\tilde{\mathsf{P}}$ probability 1, i.e., $\tilde{\mathsf{P}}$-almost surely ($\tilde{\mathsf{P}}$-a.s.), and satisfies $\mathsf{E}_{\tilde{\mathsf{P}}}[L] = 1$. Then one shows that

$$\mathsf{P}[A] = \mathsf{E}_{\mathsf{P}}[I_A] = \mathsf{E}_{\tilde{\mathsf{P}}}[I_A L] = \int_A L\, d\tilde{\mathsf{P}}, \quad \forall A \in \mathcal{F}, \tag{2.1}$$

defines an unique probability measure P on $(\Omega, \mathcal{F})$. In Eq. (2.1) we denote the indicator as

$$I_A(\omega) = I\{\omega \in A\} = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, computing the probability $\mathsf{P}[A]$, which is the expectation of I_A under the measure P, is equivalent to computing the expectation, this time under $\tilde{\mathsf{P}}$, of $I_A L$. The above result is a fairly general one, since the only restriction on the new measure $\tilde{\mathsf{P}}$ is absolute continuity with respect to (w.r.t.) P, denoted $\mathsf{P} \ll \tilde{\mathsf{P}}$ on $\mathcal{F}$. This means $\tilde{\mathsf{P}}[A] = 0 \Rightarrow \mathsf{P}[A] = 0$, $\forall A \in \mathcal{F}$. In other words, any set A allowed by P must be allowed by $\tilde{\mathsf{P}}$ as well. The random variable L, often called likelihood ratio, is the Radon-Nikodym derivative of P w.r.t. $\tilde{\mathsf{P}}$, denoted $d\mathsf{P}/d\tilde{\mathsf{P}}$.

Equation (2.1) gives the following change-of-measure result. Let X be a random variable on $(\Omega, \mathcal{F})$, then

$$z = \mathsf{E}_{\mathsf{P}}[X] = \mathsf{E}_{\tilde{\mathsf{P}}}[XL] \tag{2.2}$$

when $\mathsf{P} \ll \tilde{\mathsf{P}}$.[2]

A simple illustration is the following. Let X be Gaussian with mean 0 and variance equal to 1, under P, let $\mu \in \mathbb{R}$ and let

$$L = \exp\left\{-\mu(X - \mu) - \frac{\mu^2}{2}\right\}. \tag{2.3}$$

We have from Eq. (2.2) that

$$\mathsf{E}_{\mathsf{P}}[e^{vX}] = \mathsf{E}_{\tilde{\mathsf{P}}}[e^{vX} L] = e^{\frac{1}{2}v^2}, \quad \forall v \in \mathbb{R},$$

iff X is Gaussian with mean μ and variance equal to one under $\tilde{\mathsf{P}}$. So this change of measure allows for arbitrary recentering of X, yet without changing the variance of X.

Let f and $\tilde{f}$ be the densities of the random variable X under P and $\tilde{\mathsf{P}}$, respectively. We then have that

$$L = \frac{f(X)}{\tilde{f}(X)}$$

[1]The two usual criteria of asymptotic optimality of Monte Carlo estimators of rare event probabilities are logarithmic efficiency and bounded relative error, for which we refer to pp. 158–160 of [1] or to [34].

[2]Rigorously, Eq. (2.2) requires $\mathsf{P} \ll \tilde{\mathsf{P}}$ only on the restriction of $\mathcal{F}$ to $\sigma(X) \cap \{X \neq 0\}$, where $\sigma(X) = \{X^{-1}(B) | B \in \mathcal{B}(\mathbb{R})\}$ is the σ-algebra generated by X.

is a valid likelihood ratio for the change of measure

$$\begin{aligned} z &= \mathsf{E}_{\mathsf{P}}[g(X)] \\ &= \mathsf{E}_{\tilde{\mathsf{P}}}[g(X)L] \\ &= \int_{\mathbb{R}} g(x)\frac{f(x)}{\tilde{f}(x)}\tilde{\mathsf{P}}[X \in (x, x+dx)],^{3} \end{aligned}$$

for any Borel $g: \mathbb{R} \to \mathbb{R}$. In this situation, $\mathsf{P} \ll \tilde{\mathsf{P}}$ can be reexpressed as the support of the density f being included into the support of the density $\tilde{f}$. We see directly that the likelihood ratio in Eq. (2.3) is indeed the ratio of the two given Gaussian densities, evaluated at X.

The importance sampling algorithm amounts to select a large number of replication m, to generate $X_1, \ldots, X_m$ independently from $\tilde{f}$ and then to estimate z by

$$\hat{z}_m = \frac{1}{m}\sum_{j=1}^{m} g(X_j)\frac{f(X_j)}{\tilde{f}(X_j)}.$$

B. Stochastic process

In this section we show the extension of Sec. II A to the case of stochastic processes. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geqslant 0}, \tilde{\mathsf{P}})$ be a filtered probability space,[4] where time is either discrete, $t \in \mathbb{N}$, $\mathbb{N} = \{0, 1, \ldots\}$, or continuous, $t \in [0, \infty)$. Assume that the stochastic process $\{L_t\}_{t\geqslant 0}$ over this space is a $\tilde{\mathsf{P}}$-a.s. nonnegative martingale w.r.t. the filtration $\{\mathcal{F}_t\}_{t\geqslant 0}$[5] such that $\mathsf{E}_{\tilde{\mathsf{P}}}[L_t] = 1$, $\forall t \geqslant 0$. Then there exists a unique probability measure P on $(\Omega, \mathcal{F})$ such that

$$\forall t \geqslant 0, \quad \mathsf{P}[A_t] = \mathsf{E}_{\tilde{\mathsf{P}}}[I_{A_t} L_t] = \int_{A_t} L_t d\tilde{\mathsf{P}}, \quad \forall A_t \in \mathcal{F}_t. \tag{2.4}$$

Thus $\mathsf{P} \ll \tilde{\mathsf{P}}$.[6] The martingale $\{L_t\}_{t\geqslant 0}$ is called Radon-Nikodym or likelihood ratio process. At any $t \geqslant 0$, L_t is the density or Radon-Nikodym derivative of P w.r.t. $\tilde{\mathsf{P}}$ on $\mathcal{F}_t$.[7] Proof of Eq. (2.4) can be found, e.g., in [45]. Thus Eq. (2.4) generalizes Eq. (2.1). We have the following change-of-measure result for stochastic processes: for any integrable process $\{X_t\}_{t\geqslant 0}$,[8] it holds that

$$\mathsf{E}_{\mathsf{P}}[X_s] = \mathsf{E}_{\tilde{\mathsf{P}}}[X_s L_s] = \mathsf{E}_{\tilde{\mathsf{P}}}[X_s L_t], \tag{2.5}$$

provided $\mathsf{P} \ll \tilde{\mathsf{P}}$.[9]

C. Discrete Markov process

Throughout this article, we will consider the Markov process with discrete time domain $\mathbb{N}$ and discrete state space $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$. We note that most applied problems can be indeed formulated in this setting through time and space discretization, so this choice should not entail practical restrictions. We obtain the likelihood ratio process of change of measure from the induced probability of the Markov process. The likelihood ratio takes a simple form, depending only on the transition kernels of the Markov process. In this section we consider the time $t \geqslant 1$ and the states $n, n' \in \mathbb{Z}$.

Our Markov process $\{X_t\}_{t\in\mathbb{N}}$ is defined on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geqslant 0}, \mathsf{P})$. Let us define the transition probabilities

$$p_{t,n}(j) = \mathsf{P}[X_t = n + j \mid X_{t-1} = n], \quad \forall j \in \mathbb{Z},$$

together with the probabilities of the initial state

$$p_0(n) = \mathsf{P}[X_0 = n].$$

Let $\tilde{\mathsf{P}}$ denote a second probability measure on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geqslant 0})$, which is unambiguously determined through the change-of-measure kernels $q_0 : \mathbb{Z} \to [0, \infty)$ and $q_t : \mathbb{Z} \times \mathbb{Z} \to [0, \infty)$ as follows:

$$p_0(n) = q_0(n)\tilde{\mathsf{P}}[X_0 = n] \tag{2.6}$$

and

$$\begin{aligned} \mathsf{P}[X_t = n' \mid X_{t-1} = n, X_{t-2} = n_{t-2}, \ldots, X_1 = n_1] \\ = q_t(n, n')\tilde{\mathsf{P}}[X_t = n' \mid X_{t-1} = n, X_{t-2} \\ = n_{t-2}, \ldots, X_1 = n_1], \end{aligned} \tag{2.7}$$

for $n_{t-2}, \ldots, n_1 \in \mathbb{Z}$. Because $\{X_t\}_{t\in\mathbb{N}}$ is a Markov process under P, the values of $n_{t-2}, \ldots, n_1$ on the left side of Eq. (2.7) are irrelevant. They remain irrelevant on the right side and thus $\{X_t\}_{t\in\mathbb{N}}$ remains a Markov process under the new probability measure $\tilde{\mathsf{P}}$. So we can simplify Eq. (2.7) to

$$p_{t,n}(n' - n) = q_t(n, n')\tilde{\mathsf{P}}[X_t = n' \mid X_{t-1} = n]. \tag{2.8}$$

We can define the transition probabilities under $\tilde{\mathsf{P}}$ by

$$\tilde{p}_{t,n}(j) = \tilde{\mathsf{P}}[X_t = n + j \mid X_{t-1} = n] = \frac{p_{t,n}(j)}{q_t(n, n + j)}, \quad \forall j \in \mathbb{Z}, \tag{2.9}$$

and the initial probability under $\tilde{\mathsf{P}}$ by

$$\tilde{p}_0(n) = \tilde{\mathsf{P}}[X_0 = n].$$

We then have, for $n_0, \ldots, n_m \in \mathbb{Z}$,

$$\frac{p_0(n_0)}{\tilde{p}_0(n_0)}\prod_{t=1}^{m}\frac{p_{t,n_{t-1}}(n_t - n_{t-1})}{\tilde{p}_{t,n_{t-1}}(n_t - n_{t-1})} = q_0(n_0)\prod_{t=1}^{m} q_t(n_{t-1}, n_t), \quad \text{for } m = 1, 2, \ldots.$$

---

[3]Precisely, the validity of $L = f(X)/\tilde{f}(X)$ is limited to the restriction of $\mathcal{F}$ to $\sigma(X)$, meaning that $\mathsf{E}_{\mathsf{P}}[Z] = \mathsf{E}_{\tilde{\mathsf{P}}}[ZL]$ would be untrue with Z not $\sigma(X)$-measurable, namely, with any Z that could not take the form $Z = g(X)$, for some Borel function g.

[4]The sequence of σ algebras $\{\mathcal{F}_t\}_{t\geqslant 0}$ is a filtration in the sense that $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}, \forall\, 0 \leqslant s \leqslant t < \infty$ (with inclusion weakly meant).

[5]This means that $\mathsf{E}_{\tilde{\mathsf{P}}}[L_t|\mathcal{F}_s] = L_s, \forall\, 0 \leqslant s \leqslant t < \infty$.

[6]Precisely, $\mathsf{P} \ll \tilde{\mathsf{P}}$ holds on the restriction of $\mathcal{F}$ to $\mathcal{F}_t, \forall t \geqslant 0$.

[7]An alternative commonly used notation is $L_t = d\mathsf{P}/d\tilde{\mathsf{P}}|_t$.

[8]The process $\{X_t\}_{t\geqslant 0}$ must also be $\{\mathcal{F}_t\}_{t\geqslant 0}$-adapted, in the sense that X_t is $\mathcal{F}_t$-measurable, $\forall t \geqslant 0$.

[9]Precisely, $\mathsf{P} \ll \tilde{\mathsf{P}}$ is required on the restriction of $\mathcal{F}$ to $\mathcal{F}_s \cap \{X_s \neq 0\}, \forall 0 \leqslant s \leqslant t$.

This last expression gives us the following general form of the likelihood ratio process:

$$L_0 = q_0(X_0) \text{ and}$$
$$L_m = q_0(X_0)\prod_{t=1}^{m} q_t(X_{t-1}, X_t), \quad \text{for } m = 1, 2, \ldots.$$

Thus L_m is a function of $X_0, \ldots, X_m$, for $m = 0, 1, \ldots$.[10]

When the Markov process is homogeneous under P and the change-of-measure kernel q_t does not depend on $t \geqslant 1$, then the Markov process $\{X_t\}_{t\in\mathbb{N}}$ remains homogeneous under $\tilde{\mathsf{P}}$. In this case, by redenoting the change-of-measure kernels at times $t \neq 0$ in Eq. (2.7) simply by $q_\bullet$, we obtain the likelihood ratio process

$$L_0 = q_0(X_0) \text{ and}$$
$$L_m = q_0(X_0)\prod_{t=1}^{m} q_\bullet(X_{t-1}, X_t), \qquad \text{for } m = 1, 2, \ldots. \tag{2.10}$$

Thus the following change-of-measure formula holds for all events depending on the Markov process over the finite time horizon $[0, t^\dagger]$, for some $t^\dagger \geqslant 1$. For some given function $g: \mathbb{Z}^{t^\dagger+1} \to \mathbb{R}$, define the importance sampling estimator $Z_{t^\dagger} = g(X_0, \ldots, X_{t^\dagger})L_{t^\dagger}$. We then have

$$z_{t^\dagger} = \mathsf{E}_{\mathsf{P}}[g(X_0, \ldots, X_{t^\dagger})] = \mathsf{E}_{\tilde{\mathsf{P}}}[g(X_0, \ldots, X_{t^\dagger})L_{t^\dagger}] = \mathsf{E}_{\tilde{\mathsf{P}}}[Z_{t^\dagger}], \tag{2.11}$$

whenever $\mathsf{P} \ll \tilde{\mathsf{P}}$.[11]

Until now, nothing has been said about the form of the change-of-measure kernels (q). We will focus on two particular choices for these kernels, namely, the exponential tilt and bridge changes of measure.

D. Absolute continuity and simulation

In the context of simulation, P represents the original measure and replications of XL [see Eq. (2.2)] or of XL_s [see Eq. (2.5)] are drawn under the importance sampling measure $\tilde{\mathsf{P}}$. It may appear weird to state the existence of the original measure P (which we already have) through Eq. (2.1) and Eq. (2.4), but the important aspect here is the unambiguous relationship between P and $\tilde{\mathsf{P}}$: if either Eq. (2.1) or Eq. (2.4) can be established, then the importance sampling algorithm with $\tilde{\mathsf{P}}$ is valid.

From the theoretical perspective, the only restriction for choosing the importance sampling measure ($\tilde{\mathsf{P}}$) is absolute continuity ($\mathsf{P} \ll \tilde{\mathsf{P}}$), which $\tilde{\mathsf{P}}$-a.s. guarantees the existence of the likelihood process ($\{L_t\}_{t\geqslant 0}$). A sample path with probability zero under the original measure (P) may thus receive positive probability under the importance sampling measure ($\tilde{\mathsf{P}}$). However, since paths can be important observables themselves, one can be interested in a new measure ($\tilde{\mathsf{P}}$) that samples only the paths that have positive probability under the original probability (P). For example, in the context of stochastic thermodynamics, random paths have a prominent role in the characterization of entropy production [47,48]. As another example, rare paths can be measurable objects with important biological implications [6,49]. Thus, it can also be useful to have the stronger constraint of equivalence of measures ($\mathsf{P} \ll \tilde{\mathsf{P}}$ and $\tilde{\mathsf{P}} \ll \mathsf{P}$).[12] In Sec. III we will see that the exponential tilting always satisfies this condition, whereas in the bridge process, the equivalence of measures depends on the choice of the terminal distribution (see Sec. IV).

Last, we note that when only absolute continuity of the new measure w.r.t. the original one ($\tilde{\mathsf{P}} \ll \mathsf{P}$) holds, an importance sampling algorithm may still be used but introduce systematic errors (also called bias errors) in the estimation of the quantity of interest z, given in Eq. (2.11). Such systematic errors can be small, relative to the Monte Carlo variability, if the forbidden region by $\tilde{\mathsf{P}}$ is irrelevant for the estimation of Z. Nevertheless, these systematic errors can also hold significance as they have the potential to surpass Monte Carlo (statistical) errors, a point we will illustrate through a numerical example in Sec. V A.

III. EXPONENTIAL TILT FOR MARKOV PROCESSES

This section provides the analytical formulation of importance sampling by exponential tilting. The technique is first introduced for a single random variable, then for the simple process of partial sums of i.i.d. random variables, and finally for discrete Markov processes.

Exponential tilting is a fairly general change-of-measure procedure that can be applied whenever the underlying distribution decays sufficiently fast at its extremities, namely, when the distribution is "light-tailed." This procedure embeds the original probability measure into a new one, which renders likely specific trajectories that would have been otherwise rare, under the original probability. This technique is sometimes called Esscher transformation. It was suggested by [50,51] for local applications of the central limit theorem, in order to obtain a very accurate analytical approximation to the distribution of the sum. It was then shown by [52] that Esscher's approximation can be reformulated in terms the saddlepoint approximation of asymptotic analysis [53]. Theoretically, both saddlepoint approximation and optimal exponential tilting belong to the class of large deviations approximations [54]. These approximations are adequate for obtaining the very small probabilities of rare events; see, e.g., Chapter 3 of [44].

Exponential tilting is introduced in Sec. III A for a single random variable. It is then given for the sum of i.i.d. random variables, in Sec. III B. The likelihood ratio process of exponential tilt for Markov processes is provided in Sec. III C. We conclude with two remarks in Sec. III D: Sec. III D 1 concerns the choice of the tilting parameter, and Sec. III D 2 presents a closely related method, called s-ensemble.

[10]In other terms, the likelihood ratio process is adapted to the filtration generated by the Markov process.

[11]Precisely, $\mathsf{P} \ll \tilde{\mathsf{P}}$ is required on $\mathcal{F}_{t^\dagger}$.

[12]Precisely, one assumes $\mathsf{P} \ll \tilde{\mathsf{P}}$ and $\tilde{\mathsf{P}} \ll \mathsf{P}$ on $\mathcal{F}_t$, $\forall t \geqslant 0$.

A. Exponential tilt for random variable

Consider now the random variable X with cumulant generating function (c.g.f.)

$$K(\theta) = \log \mathsf{E}_{\mathsf{P}}[e^{\theta X}], \quad \text{for } \theta \in \mathbb{R},$$

where P represents the present probability measure, and we consider values of θ such that $K(\theta)$ is finite. The exponentially tilted measure P_θ is the measure $\tilde{\mathsf{P}}$ of Eq. (2.4) obtained by the Radon-Nikodym derivative or likelihood ratio

$$L_\theta = \frac{d\mathsf{P}}{d\mathsf{P}_\theta} = \exp[-\theta X + K(\theta)], \tag{3.1}$$

over $\sigma(X)$. Thus P_θ is equivalent to P, in the sense that $\mathsf{P}_\theta \ll \mathsf{P}$ and $\mathsf{P} \ll \mathsf{P}_\theta$.[13] The measure P_θ, called exponential tilt of P, is a practical importance sampling measure. We note that if F is the distribution function (d.f.) of X under P, then

$$dF_\theta(y) = \exp[\theta y - K(\theta)]dF(y) \tag{3.2}$$

provides the d.f. under P_θ.

Although our focus lies on univariate processes, we can briefly mention that exponential tilting generalizes directly to the multivariate setting. When $\boldsymbol{X}$ is a random vector in $\mathbb{R}^d$, for some $d \geqslant 2$, with c.g.f. $K(\boldsymbol{v}) = \log \mathsf{E}_{\mathsf{P}}[e^{\langle \boldsymbol{v}, \boldsymbol{X}\rangle}]$, for $\boldsymbol{v} \in \mathbb{R}^d$, then the likelihood ratio of Eq. (3.1) becomes

$$L_{\boldsymbol{\theta}} = \frac{d\mathsf{P}}{d\mathsf{P}_{\boldsymbol{\theta}}} = \exp[-\langle \boldsymbol{\theta}, \boldsymbol{X}\rangle + K(\boldsymbol{\theta})], \tag{3.3}$$

for $\boldsymbol{\theta} \in \mathbb{R}^d$.

B. Exponential tilt for random walk

Let us introduce the exponential tilt likelihood ratio process for the simple random walk, which is the process of partial sums of independent random variables $Y_1, Y_2, \ldots$ with common d.f. F and c.g.f. K, under some probability measure P. Consider thus

$$X_t = \sum_{j=1}^{t} Y_j, \quad \text{for } t = 1, 2, \ldots.$$

We can assume the fixed initial state 0, viz. define $X_0 = 0$. The exponentially tilted measure P_θ over $\sigma(X_1, \ldots, X_t)$ is obtained from the likelihood ratio process

$$L_t(\theta) = \exp[-\theta X_t + tK(\theta)], \quad \text{for } t = 1, 2 \ldots,$$

with θ such that $K(\theta)$ is finite. We have $\mathsf{P} \ll \mathsf{P}_\theta$ and $\mathsf{P}_\theta \ll \mathsf{P}$.[14]

For a given time horizon $t^\dagger \geqslant 1$ and for a given function $g\colon \mathbb{Z}^{t^\dagger} \to \mathbb{R}$, we are generally interested in computing $z_{t^\dagger} = \mathsf{E}_{\mathsf{P}}[g(X_1, \ldots, X_{t^\dagger})]$. The importance sampling estimator of exponential tilting is given by

$$Z_{t^\dagger}(\theta) = g(X_1, \ldots, X_{t^\dagger})L_{t^\dagger}(\theta) \tag{3.4}$$

and we have

$$\begin{aligned} z_{t^\dagger} &= \mathsf{E}_{\mathsf{P}}[g(X_1, \ldots, X_{t^\dagger})] \\ &= \mathsf{E}_{\mathsf{P}_\theta}[g(X_1, \ldots, X_{t^\dagger})L_{t^\dagger}(\theta)] \\ &= \mathsf{E}_{\mathsf{P}_\theta}[Z_{t^\dagger}(\theta)]. \end{aligned} \tag{3.5}$$

Given the multidimensional likelihood ratio formula of Eq. (3.3), the generalization of the above one-dimensional exponential tilting to the random walk $\{\boldsymbol{X}_t\}_{t\geqslant 0}$ with individual values in $\mathbb{R}^d$ with $d \geqslant 2$ is straightforward.

Note that sampling under P_θ amounts to generating i.i.d. summands from the exponentially tilted d.f. of Eq. (3.2). We thus obtain Algorithm SM1.1 of SM [43], Sec. SM 1, for importance sampling by exponential tilt for random walks.

For some large $x > \mathsf{E}_{\mathsf{P}}[Y_1]$, let $\mathsf{I}_x = (t^\dagger x, \infty)$, for some time horizon $t^\dagger \geqslant 1$. We are now interested in the rare event probability $z_{t^\dagger}(\mathsf{I}_x) = \mathsf{P}[X_{t^\dagger} \in \mathsf{I}_x]$, which a small upper tail probability of the sample mean. The importance sampling estimator is thus given by

$$Z_{t^\dagger}(\theta, \mathsf{I}_x) = I\{X_{t^\dagger} \in \mathsf{I}_x\}L_{t^\dagger}(\theta) = I\{X_{t^\dagger} > t^\dagger x\}L_{t^\dagger}(\theta) \tag{3.6}$$

and we have $z_{t^\dagger}(\mathsf{I}_x) = \mathsf{E}_{\mathsf{P}_\theta}[Z_{t^\dagger}(\theta, \mathsf{I}_x)]$.

But not every choice of tilting parameter θ reduces the variability, and inadequate choices may also increase it, substantially. Let $\theta(x)$ the solution w.r.t. v of

$$K'(v) = \frac{d}{dv}K(v) = x, \quad \text{i.e.,} \quad \mathsf{E}_{\mathsf{P}_{\theta(x)}}[Y_1] = x, \quad \text{i.e.,}$$
$$\mathsf{E}_{\mathsf{P}_{\theta(x)}}[X_{t^\dagger}] = t^\dagger x. \tag{3.7}$$

It is shown that $\theta(x)$ exists and it is unique, for any x within the interior of the range of K'; see, e.g., [52]. Moreover, [52] shows also that $\theta(x)$ appears as a saddlepoint on the surface of the real part of the complex exponent of the Fourier transform of the density. It is shown at pp. 168–169 of [1] that the importance sampling estimator given in Eq. (3.6) with $\theta = \theta(x)$ is optimal, in the sense of logarithmic efficiency, under $\mathsf{P}_{\theta(x)}$. A less rigorous but simple justification of the optimality of this choice of titling parameter is given in Sec. III D 1.

Logarithmic efficiency is a slightly weaker criterion than bounded relative error; see footnote 1. These two usual optimality criteria of rare event simulation are asymptotic for vanishing probabilities like $z_{t^\dagger}(\mathsf{I}_x)$, as $x \to \infty$. For many important accurate estimators of small probabilities, only logarithmic efficiency can be established [1]. However, these two criteria can hardly be distinguished in most practical situations.

A detailed presentation of importance sampling in the context of large deviations theory is given by [31]. In particular, an analysis of the joint large deviations behavior of the random process of interest (called "observable") and of the (logarithmically rescaled) likelihood ratio process is presented. The article provides necessary and sufficient conditions, for any general change-of-measure procedure (not necessarily exponential tilt), in order to have logarithmic efficiency. Interestingly, this result motivates further research concerning the existence and the characterization of logarithmically efficient change-of-measure methods other than exponential tilt.

[13]Precisely, $\mathsf{P}_\theta \ll \mathsf{P}$ and $\mathsf{P} \ll \mathsf{P}_\theta$ hold on on the restriction of $\mathcal{F}$ to $\sigma(X)$.

[14]Precisely, $\mathsf{P}_\theta \ll \mathsf{P}$ and $\mathsf{P} \ll \mathsf{P}_\theta$ hold on $\sigma(X_1, \ldots, X_t)$, which is the σ-algebra generated by $X_1, \ldots, X_t$, for $t = 1, 2, \ldots$.

Note finally the following property: the exponentially tilted distribution $\mathsf{P}_{\theta(x)}$ is the closest one to the original distribution P, under all distributions that are centered according to Eq. (3.7), where closeness is in terms of Kullback-Leibler divergence; cf., e.g., [54].

C. Exponential tilt for homogeneous Markov processes

The discrete time Markov process, its change-of-measure kernels, and and its likelihood ratio process are all introduced in Sec. II C. We showed that the change-of-measure kernels allow us to obtain an alternative probability measure $\tilde{\mathsf{P}}$. The objective is to choose $\tilde{\mathsf{P}}$ so to reorient sample paths towards a specific region of interest, which is rarely reached under the original measure P. We show here how $\tilde{\mathsf{P}}$ is obtained through exponential tilt, for the homogeneous Markov process. We only need to obtain the change-of-measure kernels of Eq. (2.6) and Eq. (2.7) of exponential tilting. In this section we consider the time $t \geqslant 1$ and the states $n, n' \in \mathbb{Z}$.

Let

$$z(n,\theta) = \exp[\theta n - K_0(\theta)] \tag{3.8}$$

and

$$z(n,n',\theta) = \exp[\theta(n'-n) - K_{\bullet n}(\theta)], \tag{3.9}$$

where

$$K_0(\theta) = \log \sum_{j\in\mathbb{Z}} \exp(\theta j) p_0(j) \tag{3.10}$$

and

$$K_{\bullet n}(\theta) = \log \sum_{j\in\mathbb{Z}} \exp(\theta j) p_{\bullet n}(j), \tag{3.11}$$

at any $\theta \in \mathbb{R}$ where the two sums above converge, are the c.g.f. of the probability of the initial state, denoted p_0, and the c.g.f. of the homogeneous transition probabilities, denoted $p_{\bullet n} = p_{t,n}$ and independent of the time index $t \geqslant 1$. The exponentially tilted probability measure, $\tilde{\mathsf{P}} = \mathsf{P}_\theta$, is characterized by

$$\begin{aligned} p_0(n,\theta) &= \mathsf{P}_\theta[X_0 = n] \\ &= \exp[\theta n - K_0(\theta)]\mathsf{P}[X_0 = n] = z(n,\theta)p_0(n) \end{aligned} \tag{3.12}$$

and

$$\begin{aligned} p_{\bullet n}(j,\theta) &= \mathsf{P}_\theta[X_t = n+j \mid X_{t-1} = n] \\ &= \exp[\theta j - K_{\bullet n}(\theta)]\mathsf{P}[X_t = n+j \mid X_{t-1} = n] \\ &= z(n,n+j,\theta)p_{\bullet n}(j), \quad \forall j \in \mathbb{Z}. \end{aligned} \tag{3.13}$$

Thus exponential tilt corresponds to the particular choice of the general change-of-measure kernels Eqs. (2.6) and (2.7), respectively, given by

$$q_0(n) = \frac{1}{z(n,\theta)} \quad \text{and} \quad q_\bullet(n,n') = \frac{1}{z(n,n',\theta)}.$$

Thus, the likelihood ratio process of Eq. (2.10) for the case of exponential tilt becomes

$$\begin{aligned} L_0(\theta) &= q_0(X_0) = [z(X_0,\theta)]^{-1} \quad \text{and} \\ L_t(\theta) &= \left[z(X_0,\theta) \prod_{k=1}^t z(X_{k-1},X_k,\theta) \right]^{-1} \\ &= \Bigg\{ \exp\Bigg[\theta X_0 - K_0(\theta) \\ &\quad + \sum_{k=1}^t \theta(X_k - X_{k-1}) - K_{\bullet X_{k-1}}(\theta) \Bigg]\Bigg\}^{-1} \\ &= \exp\left\{ -\theta X_t + \left[K_0(\theta) + \sum_{k=1}^t K_{\bullet X_{k-1}}(\theta) \right]\right\} \\ &= \mathrm{e}^{-\theta X_t} M_0(\theta) \prod_{k=1}^t M_{\bullet X_{k-1}}(\theta), \end{aligned} \tag{3.14}$$

where the argument θ has been added to the likelihood ratio for convenience and where $M_0 = \mathrm{e}^{K_0}$ and $M_{\bullet n} = \mathrm{e}^{K_{\bullet n}}$ are the moment generating functions of p_0 and $p_{\bullet n}$, respectively.

Let us now give a couple of remarks. The required absolute continuity is clearly satisfied, because of the positivity of the change-of-measure kernels of exponential tilt. In fact, both $\mathsf{P} \ll \mathsf{P}_\theta$ and $\mathsf{P}_\theta \ll \mathsf{P}$ hold, namely, P and P_θ are equivalent. In contrast with the likelihood ratio for the bridge process, presented in Sec. IV, the likelihood ratio of Eq. (3.14) is not restricted to problems with finite time horizons.

Consider the time horizon $t^\dagger \geqslant 1$ and the function g: $\mathbb{Z}^{t^\dagger+1} \to \mathbb{R}$. We want to evaluate $z_{t^\dagger} = \mathsf{E}_\mathsf{P}[g(X_0,\dots,X_{t^\dagger})]$. The estimator of exponential tilting is given by

$$Z_{t^\dagger}(\theta) = g(X_0,\dots,X_{t^\dagger}) L_{t^\dagger}(\theta), \tag{3.15}$$

for $L_{t^\dagger}(\theta)$ given in Eq. (3.14), and we have

$$\begin{aligned} z_{t^\dagger} &= \mathsf{E}_\mathsf{P}[g(X_0,\dots,X_{t^\dagger})] \\ &= \mathsf{E}_{\mathsf{P}_\theta}[g(X_0,\dots,X_{t^\dagger}) L_{t^\dagger}(\theta)] \\ &= \mathsf{E}_{\mathsf{P}_\theta}[Z_{t^\dagger}(\theta)]. \end{aligned} \tag{3.16}$$

Exponential tilting can be also obtained for the multidimensional homogeneous Markov process $\{\boldsymbol{X}_t\}_{t\geqslant 0}$ taking individual values in $\mathbb{Z}^d$, at any $d \geqslant 2$, essentially by replacing Eqs. (3.8) and (3.9) by

$$\begin{aligned} z(\boldsymbol{n},\boldsymbol{\theta}) &= \exp[\langle \boldsymbol{\theta},\boldsymbol{n}\rangle - K_0(\boldsymbol{\theta})] \quad \text{and} \\ z(\boldsymbol{n},\boldsymbol{n}',\boldsymbol{\theta}) &= \exp[\langle \boldsymbol{\theta},\boldsymbol{n}'-\boldsymbol{n}\rangle - K_{\bullet \boldsymbol{n}}(\boldsymbol{\theta})], \end{aligned}$$

where $\boldsymbol{n},\boldsymbol{n}' \in \mathbb{Z}^d$, $K_0(\boldsymbol{\theta}) = \log \sum_{\boldsymbol{j}\in\mathbb{Z}^d} \exp[\langle \boldsymbol{\theta},\boldsymbol{j}\rangle] p_0(\boldsymbol{j})$ and $K_{\bullet \boldsymbol{n}}(\boldsymbol{\theta}) = \log \sum_{\boldsymbol{j}\in\mathbb{Z}^d} \exp[\langle \boldsymbol{\theta},\boldsymbol{j}\rangle] p_{\bullet \boldsymbol{n}}(\boldsymbol{j})$, at any $\boldsymbol{\theta} \in \mathbb{R}^d$ for which the two sums above converge.

We note the exponential tilting requires the existence of the c.g.f. of the transition probabilities; see Eqs. (3.10) and (3.11). This is a light-tail requirement on the transition distributions. In several problems of physics the state space is a finite set and so these c.g.f. do always exist. Thus there are no restrictions on the value of the tilting parameter θ. But there are situations of physics in which the existence of the c.g.f. of the transition probabilities cannot be guaranteed. This is often the case when the Markov process refers to quantities such as the time-averaged current or activity in, e.g., exclusion processes or to quantities subject to kinetic constraints.

From the above derivations, we can compute the desired expectation in Eq. (3.16) with Algorithm SM1.2 in SM [43], Sec. SM 1, of importance sampling by exponential tilting. We consider the practical case with fixed initial initial state.

D. Remarks: Optimality and s-ensemble

This section presents two general remarks related to exponential tilting. The first one concerns the optimal choice of the tilting parameter and is given in Sec. III D 1. The second remark concerns the closely related technique called s-ensemble in the physical literature, which is another approach to exponential tilting and which is the subject of Sec. III D 2.

1. Optimal tilting parameter under time and space homogeneity

The methodology introduced so far does not address the question of the selection of the tilting parameter associated with the rare event under consideration. In fact, through the numerical examples in Sec. V, we will provide evidence that there is usually an optimal tilting parameter minimizing the sampling error of the numerical estimations. We next show the computation of the optimal tilting parameter for a particular problem as an illustration. In particular, we are now interested in the probability of reaching, at some final time $t^\dagger$, the interval of states

$$\mathsf{I}_c(a) = [c-a, c+a] \cap \mathbb{Z}, \quad \text{for some integers } \ a \geqslant 0 \text{ and } c.$$

Consider also that the initial state is fixed $X_0 = n_0$, for some $n_0 \in \mathbb{Z}$ much smaller than $c - a$. The quantity of interest is thus

$$z_{t^\dagger}(\mathsf{I}_c(a)) = \mathsf{P}[X_{t^\dagger} \in \mathsf{I}_c(a)]. \tag{3.17}$$

We further assume that the process is homogeneous in the state space: the transition probability $p_{\bullet n}$ does not depend on n and we denote $p_{\bullet\bullet} = p_{\bullet n}$. This is the random walk of Sec. III B. Denote by $K_{\bullet\bullet}$ the c.g.f. of $p_{\bullet\bullet}$. We want to determine the exponential tilting parameter θ for which $\mathrm{var}_{\mathsf{P}_\theta}(Z_{t^\dagger}(\theta, \mathsf{I}_c(a)))$ is small, i.e., such that $\mathsf{E}_{\mathsf{P}_\theta}[Z^2_{t^\dagger}(\theta, \mathsf{I}_c(a))]$ is small. Then Eq. (3.14) leads to

$$L_{t^\dagger}(\theta) = \exp[-\theta(X_{t^\dagger} - n_0) + t^\dagger K_{\bullet\bullet}(\theta)].$$

We thus have

$$\begin{aligned}
&\mathsf{E}_{\mathsf{P}_\theta}[\{Z_{t^\dagger}(\theta, \mathsf{I}_c(a))\}^2] \\
&\quad= \mathsf{E}_{\mathsf{P}_\theta}[\,(I\{c-a \leqslant X_{t^\dagger} \leqslant c-a\}L_{t^\dagger}(\theta)\,)^2] \\
&\quad\leqslant \mathsf{E}_{\mathsf{P}_\theta}[I\{X_{t^\dagger} \geqslant c-a\}\{\exp[-\theta(X_{t^\dagger} - n_0) + t^\dagger K_{\bullet\bullet}(\theta)]\}^2] \\
&\quad\leqslant \{\exp[-\theta(c-a-n_0) + t^\dagger K_{\bullet\bullet}(\theta)]\}^2 \mathsf{E}_{\mathsf{P}_\theta}[I\{X_{t^\dagger} \geqslant c-a\}] \\
&\quad\leqslant \exp\{-2[\theta(c-a-n_0) - t^\dagger K_{\bullet\bullet}(\theta)]\},
\end{aligned}$$

given that $\theta > 0$ whenever $a > 0$. Strict convexity of $K_{\bullet\bullet}$ implies that the above exponent is minimized for

$$t^\dagger K'_{\bullet\bullet}(\theta) = c - a - n_0 > 0,$$

namely, for

$$\mathsf{E}_{\mathsf{P}_\theta}[X_{t^\dagger} - X_0 \mid X_0 = n_0] = c - a - n_0, \tag{3.18}$$

which thus recenters the average of $X_{t^\dagger}$ towards the lower bound of the target interval $\mathsf{I}_c(a)$.

The problem of finding the optimal parameter for arbitrary expectations is not yet solved [e.g., for the case of computing the same estimator in Eq. (3.17) but for a process with state-dependent transition probabilities]. Nevertheless, expressions like Eq. (3.18) can be used in an intuitive way to reduce the sampling errors. For example, in the considered situation where the target interval $\mathsf{I}_c(a)$ is well above the starting point n_0, any value $\theta > 0$ that redrifts the process sufficiently upwards is expected to reduce the Monte Carlo variability. We will elaborate more on this point in Sec. V A.

2. s-ensemble

This section briefly summarizes the alternative closely related importance sampling estimator called s-ensemble. It is an ancillary section that is not required for the comprehension of this article. The s-ensemble change-of-measure [55–57] is directly defined at the level of path measures as follows:[15]

$$\begin{aligned}
&\mathsf{P}_s[X_0 = n_0, \ldots, X_{t^\dagger} = n_{t^\dagger}] \\
&\quad= \mathsf{P}[X_0 = n_0, \ldots, X_{t^\dagger} = n_{t^\dagger}]\frac{\mathrm{e}^{s(n_{t^\dagger} - n_0)}}{\mathcal{Z}_{t^\dagger}(s)},
\end{aligned} \tag{3.19}$$

$\forall n_0, \ldots, n_{t^\dagger} \in \mathbb{Z}$, where

$$\mathcal{Z}_{t^\dagger}(s) = \sum_{n_0, n_{t^\dagger} \in \mathbb{Z}} \mathsf{P}[X_0 = n_0, X_{t^\dagger} = n_t]\mathrm{e}^{s(n_{t^\dagger} - n_0)},$$

at any $s \in \mathbb{R}$ where the sum converges. Thus $\mathcal{Z}_{t^\dagger}$ is the moment generating function[16] of the increment $X_{t^\dagger} - X_0$. The parameter s is used to fix the first moment of the increment to some desired value c through

$$\mathsf{E}_{\mathsf{P}_s}[X_{t^\dagger} - X_0] = \frac{d}{ds}\log \mathcal{Z}_{t^\dagger}(s) = c. \tag{3.20}$$

From the definition of the s-ensemble probability measure in Eq. (3.19), we can readily derive the likelihood ratio process of this change of measure in the following form:

$$L_{t^\dagger}(s) = \mathrm{e}^{-s(X_{t^\dagger} - X_0)}\mathcal{Z}_{t^\dagger}(s). \tag{3.21}$$

The s-ensemble change of measure is similar in form to our exponential tilting, as both techniques bias the original measure by exponential factors. Also, in a process with homogeneous transition probabilities that do not depend on the state of the process, nor on time, both changes of measure are identical (see Sec. V A 3). Also, a more general definition of exponential tilting, with one tilting parameter per unit of time, can include both the exponential tilt for the homogeneous Markov processes of Sec. III C and the s-ensemble. However,

[15]Usually the s-ensemble is defined in a more general manner through processes called "integrated observables." Here we have chosen to work with a specific case of integrated observable, namely, the increment $X_t - X_0$, and we refer to [55–57] for a more general definition.

[16]The specific notation $\mathcal{Z}$ (instead of M used in other sections) for the moment generating function is typical in the s-ensemble literature, as is s (instead of θ) for the tilting parameter.

the s-ensemble and exponential tilt as given in Sec. III C exhibit relevant differences: the s-ensemble draws paths with fixed mean global increment through Eq. (3.20) [27,28]. Also, the asymptotic properties of the s-ensemble make it a useful tool to compute large deviation rates [5,23–26,55–59] and constrained paths in the limit of large times ($t^\dagger \to \infty$) [60,61]. On the other side, while transition probabilities in our exponential tilting are obtained through a simple transformation of the transition probabilities of the original process, obtaining the transition probabilities of the s-ensemble is not a trivial task; see, e.g., [5,62]. Since the focus of this work lies on rare paths within finite time horizons, we do not explore applications of the s-ensemble change of measure.

IV. STOCHASTIC BRIDGES FOR MARKOV PROCESSES

The bridge process provides a practical alternative technique to exponential tilting. It also constructs a sampling probability measure $\tilde{\mathsf{P}}$ that makes frequent a given event of interest, which is rare under P. The main idea is to generate a bridge process with fixed boundary points or endpoints. The bridge process is then used for sampling the rare paths that possess unlikely pairs of endpoints under the original probability. In fact, as we explain below, the technique can be readily extended to address arbitrary initial and final distributions. There are various recent applications of this methodology for sampling rare events in the context of intrinsically out-of-equilibrium systems [16–22,63]. The bridge methods are similar in spirit to the transition-path-sampling algorithms [38–40], extensively used in equilibrium molecular dynamics, where paths constrained to both ends are sampled using a Metropolis-Hastings scheme. However, transition path sampling is based on a proposal-rejection scheme like the Metropolis-Hastings algorithm, meaning that generated paths are accepted with some probability. Contrary, all transition paths generated with a bridge process will end in the desired regions by construction. The generator of the bridge process is obtained conditioning the transition probabilities [60,61], which is explained in the following section. As before, we consider processes with discrete state and time spaces. With the methods of this section, we always need a fixed time horizon $t^\dagger \geqslant 1$. We consider the time $t < t^\dagger$, in $\mathbb{N}$, and the states $n_0, n_{t^\dagger}, n, n' \in \mathbb{Z}$.

A. Conditioned Markov process

The bridge process is obtained upon conditioning the original Markov process on passing through particular states at given times. It is possible to sample the stochastic bridges both backward or forward in time, giving rise to two possible generators that we describe below.

1. Forward generator

We can derive transition probabilities that are conditional on some fixed final state $X_{t^\dagger} = n_{t^\dagger}$ through the relation

$$
\begin{aligned}
&\mathsf{P}[X_{t+1} = n' | X_t = n, X_{t^\dagger} = n_{t^\dagger}]\\
&\quad = \frac{\mathsf{P}[X_{t+1} = n', X_t = n, X_{t^\dagger} = n_{t^\dagger}]}{\mathsf{P}[X_t = n, X_{t^\dagger} = n_{t^\dagger}]}\\
&\quad = \mathsf{P}[X_{t+1} = n' | X_t = n] \frac{\mathsf{P}[X_{t^\dagger} = n_{t^\dagger} | X_{t+1} = n', X_t = n]}{\mathsf{P}[X_{t^\dagger} = n_{t^\dagger} | X_t = n]}\\
&\quad = \mathsf{P}[X_{t+1} = n' | X_t = n] \frac{\mathsf{P}[X_{t^\dagger} = n_{t^\dagger} | X_{t+1} = n']}{\mathsf{P}[X_{t^\dagger} = n_{t^\dagger} | X_t = n]}.
\end{aligned} \tag{4.1}
$$

We can re-express Eq. (4.1) with specific notation for transition probabilities and change-of-measure kernels as

$$
\tilde{p}_{t+1,n}(n'-n; t^\dagger, n_{t_\dagger}) = u_{t+1}(n, n'; t^\dagger, n_{t^\dagger})\, p_{t+1,n}(n'-n), \tag{4.2}
$$

where

$$
u_{t+1}(n, n'; t^\dagger, n_{t^\dagger}) = \frac{\mathsf{P}[X_{t^\dagger} = n_{t^\dagger} | X_{t+1} = n']}{\mathsf{P}[X_{t^\dagger} = n_{t^\dagger} | X_t = n]} \tag{4.3}
$$

and

$$
\tilde{p}_{t+1,n}(n'-n; t^\dagger, n_{t_\dagger}) = \mathsf{P}[X_{t+1} = n' | X_t = n, X_{t^\dagger} = n_{t^\dagger}].
$$

Thus, Eq. (4.2) is a special case of Eq. (2.9) (with $q_t = 1/u_t$). By using the transition probabilities of Eq. (4.2) and by considering the initial state as fixed, we obtain paths that necessarily cross the boundary points $X_0 = n_0$ and $X_{t^\dagger} = n_{t^\dagger}$. We have thus generated a bridge process. Since the transition probabilities in Eq. (4.2) operate forward in time, we call this procedure the forward generator of the bridge.

2. Backward generator

Now we construct bridges with transition probabilities that fix the initial state and operate backward in time. Using manipulations similar to those of Eq. (4.1), we obtain

$$
\begin{aligned}
&\mathsf{P}[X_t = n' | X_{t+1} = n, X_0 = n_0]\\
&\quad = \mathsf{P}[X_{t+1} = n | X_t = n'] \frac{\mathsf{P}[X_t = n' | X_0 = n_0]}{\mathsf{P}[X_{t+1} = n | X_0 = n_0]},
\end{aligned}
$$

namely,

$$
\begin{aligned}
&\mathsf{P}[X_t = n' | X_{t+1} = n, X_0 = n_0]\\
&\quad = w_t(n', n; n_0) \mathsf{P}[X_{t+1} = n | X_t = n']\\
&\quad = w_t(n', n; n_0) p_{t+1,n'}(n - n'),
\end{aligned} \tag{4.4}
$$

where

$$
w_t(n', n; n_0) = \frac{\mathsf{P}[X_t = n' | X_0 = n_0]}{\mathsf{P}[X_{t+1} = n | X_0 = n_0]} \tag{4.5}
$$

is the backward change-of-measure kernel. Considering fixed final states ($X_{t^\dagger} = n_{t^\dagger}$), the transition probabilities in Eq. (4.4) draw stochastic bridges connecting $X_0 = n_0$ and $X_{t^\dagger} = n_{t^\dagger}$. The kernels of Eq. (4.5) have different functionality than the change-of-measure kernel of Eq. (2.8). This distinction arises from the backward nature of the generator for the bridge process.

Both forward Eq. (4.1) and backward Eq. (4.4) generators are obtained upon multiplying the original transition probabilities by the change-of-measure kernels u_{t+1} and w_t. These kernels take the form of Doob's h-transform; cf. pp. 190–195 of [45]. Nevertheless, the probabilities in numerator and denominator of u_{t+1} of Eq. (4.3) are efficiently computed through a backward Kolmogorov equation, whereas the probabilities in numerator and denominator of w_t of Eq. (4.5) are usually computed with a forward Kolmogorov equation.

Applications that involve sampling bridges with a common initial state (at $t = 0$) but multiple final destinations (at $t = t^\dagger$), thus with the Kronecker delta initial distrbution $\mathsf{P}[X_0 = n] = \delta_{n,n_0}$, are better addressed by the backward generator, as described in [16]. The reason is that with the backward generator we need to iterate the forward Kolmogorov equation only once, in order to compute the quantities $\mathsf{P}[X_t = n|X_0 = n_0]$, for all relevant values of n and t, that are necessary for obtaining the backward change-of-measure kernel (4.5) and thus for sampling the bridges. On the other hand, sampling bridges with fixed initial state and multiple final destinations using the forward generator with its change-of-measure kernel in Eq. (4.3) would require iterating the backward Kolmogorov equation once for each final point of the bridge. For the same reason, the forward generator is more practical than the backward generator when investigating ensembles of bridges with a fixed final position but varying initial conditions.

Since we focus on problems with fixed initial condition, we will work in the following with the backward generator, also called backtracking method [16]. However, it's worth noting that all the derivations related to the backward generator have their analogous counterparts for the forward generator, so we can make this choice without any loss of generality.

B. Stochastic bridge change of measure and likelihood ratio

Now, we want to use one of the bridge generators of Sec. IV A to define the stochastic bridge measure, an alternative version of the generic change-of-measure formulas derived in Sec. II C, which we note as $\tilde{\mathsf{P}} = \mathsf{P}_{n_0}$. To completely define such a path measure, we have to specify the statistics of the initial and final conditions. We focus on paths with fixed initial condition, $X_0 = n_0$, so that $\mathsf{P}_{n_0}[X_0 = n] = \delta_{n,n_0}$. We have considerable freedom in selecting the final distribution, subject to constraints of admissibility associated with absolute continuity (see Sec. IV C for a detailed discussion of this aspect). We call $w_{t^\dagger}$ the probability function of the final state,

$$\mathsf{P}_{n_0}[X_{t^\dagger} = n] = w_{t^\dagger}(n; n_0). \tag{4.6}$$

We chose to use the backward generator to draw realizations of the Markov process under P_{n_0} backward in time, following the transition probabilities specified in Eq. (4.4),

$$\mathsf{P}_{n_0}[X_t = n'|X_{t+1} = n] = w_t(n', n; n_0)\mathsf{P}[X_{t+1} = n'|X_t = n]. \tag{4.7}$$

The corresponding likelihood ratio is obtained by multiplication of these backward change-of-measure kernels and it is thus given by

$$L_{t^\dagger}(n_0) = \{w_0(n_0, X_1; n_0) \cdots w_{t^\dagger-1}(X_{t^\dagger-1}, X_{t^\dagger}; n_0) w_{t^\dagger}(X_t^\dagger; n_0)\}^{-1}. \tag{4.8}$$

Interestingly, all terms in Eq. (4.8) cancel out excepting those that depend on the final state at time $t^\dagger$. Therefore, the expression of the likelihood ratio reduces to

$$L_{t^\dagger}(n_0) = \frac{h(X_{t^\dagger}; n_0)}{w_{t^\dagger}(X_{t^\dagger}; n_0)}, \tag{4.9}$$

where

$$h(n; n_0) = \mathsf{P}[X_{t^\dagger} = n|X_0 = n_0].$$

Thus, for some given function $g\colon \mathbb{Z}^{t^\dagger+1} \to \mathbb{R}$, the importance sampling estimator of backtracking is given by

$$Z_{n_0,t^\dagger} = g(n_0, X_1, \dots, X_{t^\dagger}) L_{t^\dagger}(n_0) \tag{4.10}$$

and we have

$$\begin{aligned} z_{t^\dagger} &= \mathsf{E}_{\mathsf{P}}[g(n_0, X_1, \dots, X_{t^\dagger})] \\ &= \mathsf{E}_{\mathsf{P}_{n_0}}[g(n_0, X_1, \dots, X_{t^\dagger}) L_{t^\dagger}] = \mathsf{E}_{\mathsf{P}_{n_0}}\left[Z_{n_0,t^\dagger}\right]. \end{aligned}$$

We conclude this section with two remarks. We first note that with backtracking it is necessary to fix the time horizon $t^\dagger$ in advance and that there will be only one likelihood ratio random variable, to be used at all intermediate times $t \in [0, t^\dagger]$, instead of a complete likelihood ratio process over the time horizon $[0, t^\dagger]$. The second remark concerns the extension of backtracking to the multidimensional state space. It turns out that there is no conceptual difference when considering a Markov process $\{\boldsymbol{X}_t\}_{t\geqslant 0}$ taking individual values in $\mathbb{Z}^d$, for some $d \geqslant 2$. All formulas of Sec. IV A 2 and of the present section remain valid in their given form when the states $\boldsymbol{n}_0, \boldsymbol{n}, \boldsymbol{n}'$ represent points of $\mathbb{Z}^d$.

With the above results, we can provide Algorithm SM1.3 in Sec. SM 1 in SM [43] for importance sampling by backtracking, for the computation of $z_{t^\dagger}$.

C. Choice of terminal distribution

The efficiency of the backtracking method depends on the choice of the final distribution of the new process ($w_{t^\dagger}$). This distribution has an analog role to the tilting parameter (θ) of the exponentially tilted measure. Let $n, n_0 \in \mathbb{Z}$. For example, if we choose $w_{t^\dagger}$ to be equal to the distribution of states at time $t^\dagger$ with the original process ($w_{t^\dagger}(n; n_0) = \mathsf{P}[X_{t^\dagger} = n|X_0 = n_0]$), then the new and original measures assign the same weights to paths ($L_{t^\dagger} = 1$), and therefore the change of measure will not result in improved efficiency for sampling rare events.

The only restriction concerning the choice of the final distribution $w_{t^\dagger}$ is the absolutely continuity $\mathsf{P} \ll \mathsf{P}_{n_0}$.[17] This condition is fulfilled if

$$w_{t^\dagger}(n; n_0) = 0 \Longrightarrow \mathsf{P}[X_{t^\dagger} = n|X_0 = n_0] = 0. \tag{4.11}$$

As discussed in Sec. II D, many applications require that all paths sampled with the new probability measure are also accessible with the original measure (e.g., the equivalence between measures). This stronger constraint is fulfilled when

$$w_{t^\dagger}(n; n_0) = 0 \Longleftrightarrow \mathsf{P}[X_{t^\dagger} = n|X_0 = n_0] = 0. \tag{4.12}$$

Choices of the distribution $w_{t^\dagger}$ fulfilling Eqs. (4.12) and (4.11) generate unbiased changes of measure, in the sense

[17]Rigorously, it is $\mathsf{P} \ll \mathsf{P}_{n_0}$ on $\mathcal{F}_{t^\dagger} \cap \{X_0 = n_0\}$.

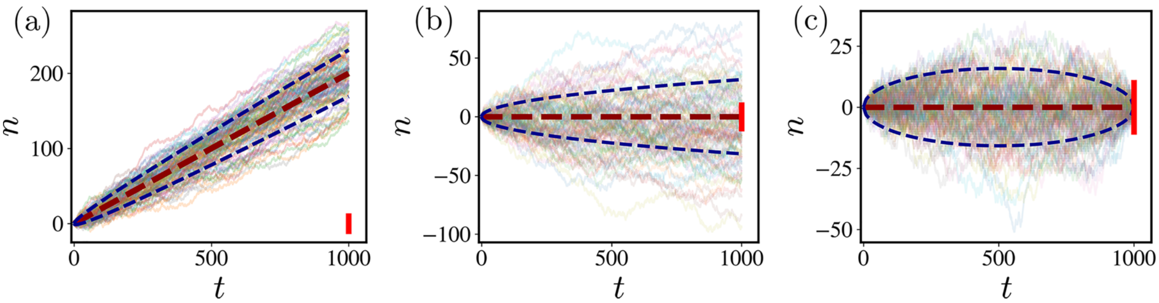

FIG. 1. In each of panels (a)–(c) are shown ensembles of 100 sample paths over the time interval [0,1000]. The original binomial process defined through Eq. (5.9) with $r = 0.6$ is drawn in (a), the exponentially tilted process obtained by Eq. (5.3) with $\rho = 0$ is drawn in (b), and the backtrack processes obtained by Eq. (5.10) with a Kronecker delta distribution for the last state is drawn in (c). Dashed red lines indicate μ_t, and dashed blue lines indicate $\mu_t \pm \sigma_t$, where μ_t denotes expected value and σ_t standard deviation, at time $t \in [0, 1000]$, both defined in Eq. (SM 3) of SM [43], Sec. SM2. The target interval $\mathsf{I}_0(10)$ from Sec. V A 3 is shown as a vertical red segment.

that the computed expected values are not affected by systematic errors. On the other side, if there are forbidden states under the importance sampling measure that were accessible with the original measure, then the errors do not tend to zero as the number of Monte Carlo replications increases. Nevertheless, such errors could be smaller than the sampling errors in cases for which the forbidden areas under the importance sampling measure have little relevance for the estimator. This applies to problems that involve transition paths between metastable states, where choices such as Kronecker delta distributions ($w_{t^\dagger}(n; n_0) = \delta_{n,n_{t^\dagger}}$), violating absolute continuity, can nevertheless be employed for computing estimators with sufficiently small errors, as described in [16].

V. EXAMPLES AND NUMERICAL STUDY

In this section, numerical comparison between backtracking and exponential tilting are presented through the following examples: the binomial process, in Sec. V A, and a process with state-dependent transition probabilities exhibiting metastable states, in Sec. V B. In this section we consider times $s < t \leqslant t^\dagger$, all in $\mathbb{N}$, and states $n_0, n_{t^\dagger}, n, n' \in \mathbb{Z}$.

A. Binomial Markov process

Random walks are prototypical toy models to test methods in nonequilibrium statistical physics. Furthermore, the extreme statistics of random walks have recently become a subject of intense research due to their wide-ranging applications in finance; see, e.g., [64,65]. Our work utilizes this random walk example as a basis for applying the derivations discussed in previous sections. Through a simple and analytically calculable example, we can better understand the two Monte Carlo methods. The process is defined by the transition probabilities $p_{\bullet\bullet} = p_{t,n}$, thus not depending on the state $n \in \mathbb{Z}$, nor on $t \in [0, t^\dagger]$, the time. In particular, they are given by

$$p_{\bullet\bullet}(j) = \begin{cases} 1 - r, & \text{if } j = -1, \\ r, & \text{if } j = 1, \end{cases} \tag{5.1}$$

for some $r \in (0, 1)$. The position of the walker follows a binomial distribution (see, e.g., [47]),

$$\mathsf{P}[X_t = n | X_s = n'] = B\left(\frac{n - n' + t - s}{2}, r, t - s\right),$$

where $\forall p \in (0, 1)$, $k \in \{1, 2, \dots\}$,

$$B(j, p, k) = \begin{cases} \binom{k}{j} p^j (1-p)^{k-j}, & \text{if } j = 0, 1, \dots, k, \\ 0, & \text{otherwise,} \end{cases} \tag{5.2}$$

are binomial probabilities. Figure 1(a) shows instances of trajectories generated with these transition probabilities, with $r = 0.6$ and over the time interval [0,1000] together with the first and second cumulants of the binomial process. In the following examples, we will consider the problem of estimating the probability of a process departing from the fixed state $n_0 \in \mathbb{Z}$ at time 0 and its subsequent passage through specific domains in the time-state space, within the time interval [0,1000]. Thus, the probability for the first state is the Kronecker delta given by $\mathsf{P}[X_0 = n] = \delta_{n,n_0}$.

Before presenting the numerical comparisons, in Secs. V A 3 and V A 4, let us summarize in Secs. V A 1 and V A 2 how the two importance sampling procedures apply to this particular scenario.

1. Likelihood ratio of exponential tilting

Exponential tilting simplifies substantially when considering the binomial process. We find directly the c.g.f.

$$K_{\bullet\bullet}(\theta) = \log[r\mathrm{e}^{\theta} + (1 - r)\mathrm{e}^{-\theta}]$$

and thus the exponential tilting transition probabilities of Eq. (3.13) become

$$p_{\bullet\bullet}(j, \theta) = z(n, n + j, \theta) p_{\bullet\bullet}(j) = \begin{cases} \frac{1-r}{1-r+r\mathrm{e}^{2\theta}}, & \text{if } j = -1, \\ \frac{r}{r+(1-r)\mathrm{e}^{-2\theta}}, & \text{if } j = 1. \end{cases} \tag{5.3}$$

Thus, the process is stable under exponential tilting, in the sense that it remains binomial under P_θ. In consistency with the rest of the text, we consider the fixed initial state $n_0 \in \mathbb{Z}$. Similar to what is done in Eq. (3.18), we obtain the tilting parameter by setting the conditional expectation of the total run through distance equal to $n_{t^\dagger} - n_0$, namely, by solving

$$\mathsf{E}_{\mathsf{P}_\theta}[X_{t^\dagger} - X_0 \mid X_0 = n_0] = t^\dagger\left[2\frac{r}{r+(1-r)\mathrm{e}^{-2\theta}} - 1\right] = n_{t^\dagger} - n_0, \tag{5.4}$$

where the above expectation is the one of the binomial distribution. Note that the capability of fixing the first moment of $X_{t^\dagger} - X_0$, which is always true in the s-ensemble [cf. Sec. III D 2 and in particular Eq. (3.20)], holds only for the case of exponential tilting because of the simplicity of this particular process (i.e., transition probabilities independent of the state of the process). In Sec. V B we will treat another example in which Eq. (5.4) does not hold.

By defining

$$\rho = \frac{n_{t^\dagger} - n_0}{t^\dagger} \tag{5.5}$$

and by inverting Eq. (5.4), we obtain

$$\theta = -\frac{1}{2}\log\left(\frac{r}{1-r}\frac{1-\rho}{1+\rho}\right), \tag{5.6}$$

which is well defined when $|n_{t^\dagger} - n_0| < t^\dagger$ (thus with exclusion of the two monotone sample paths).

Equation (5.6) implies that it is equivalent to fix the tilting parameter (θ), the average current (ρ), or the average final destination ($n_{t^\dagger}$) of the walker. This map between the tilting parameter and the average current ρ makes it easier to find intuitively values of θ that will bias paths towards desired regions of the space. Furthermore, this expression makes it explicit the analogy with the canonical ensemble, in which the temperature parameter fixes the energy of the system on average, thus explaining why exponential tilt methods are also referred to as canonical methods [61]. Still, the choice of the optimal tilting parameter (or average current) that minimizes the relative error will depend to the specific problem to solve (see, e.g., Sec. V A 3).

Upon inserting Eq. (5.6) into Eq. (5.3), we obtain the exponentially tilted transition probabilities in the form

$$p_{\bullet\bullet}(j,\theta) = \begin{cases} \frac{1-\rho}{2}, & \text{if } j = -1,\\ \frac{1+\rho}{2}, & \text{if } j = 1. \end{cases}$$

We note that these tilted probabilities do not depend on r. In Fig. 1(b) we show instances of trajectories generated with these transition probabilities using $\rho = 0$. The homogeneity of the process simplifies the expression of the likelihood ratio in Eq. (3.14) to

$$L_t(\theta) = \mathrm{e}^{-\theta(X_t - n_0)}\prod_{k=1}^{t} M_{\bullet\bullet}(\theta) = \mathrm{e}^{-\theta(X_t - n_0)}[r\mathrm{e}^{\theta} + (1-r)\mathrm{e}^{-\theta}]^t, \quad \text{for } t = 1,\dots,t^\dagger, \tag{5.7}$$

where $M_{\bullet\bullet}$ is the moment generating function of $p_{\bullet\bullet}$. We can simplify further Eq. (5.7) by introducing the tilting parameter as given in Eq. (5.6) together with the change of variables $q = (1+\rho)/2$ and $N_t = (X_t - n_0 + t)/2$, yielding

$$L_t(q) = \left(\frac{r}{q}\right)^{N_t}\left(\frac{1-r}{1-q}\right)^{t-N_t}, \quad \text{for } t = 1,\dots,t^\dagger, \tag{5.8}$$

where the selected tilting parameter θ is encoded in q.

2. Likelihood ratio of backtracking

The binomial distribution of the process simplifies the application of backtracking as well, since the probabilities of the backward change-of-measure kernels given in Eq. (4.5) are obtained directly from

$$\mathsf{P}[X_t = n|X_0 = n_0] = B\left(\frac{t+n-n_0}{2}, r, t\right). \tag{5.9}$$

Therefore, insertion of Eq. (5.9) into Eq. (4.5) leads to the transition probabilities under the backtracking measure P_{n_0} through Eq. (4.4) as

$$\mathsf{P}[X_t = n \mid X_{t+1} = n+j, X_0 = n_0] = \begin{cases} \frac{1}{2}\left(1 - \frac{n_0-n}{t}\right), & \text{if } j = -1,\\ \frac{1}{2}\left(1 + \frac{n_0-n}{t}\right), & \text{if } j = 1. \end{cases} \tag{5.10}$$

We can freely choose the probability distribution of the states at final time $t^\dagger$, i.e., $w_{t^\dagger}$ defined at Eq. (4.6). When the final distribution is a Kronecker delta, $w_{t^\dagger}(n; n_0) = \delta_{n,n_{t^\dagger}}$, the process evolving through Eq. (5.10) has fixed initial (n_0) and final ($n_{t^\dagger}$) positions. Therefore, since the backtracking can fix the current ρ exactly [cf. Eq. (5.5)] and not on average as the exponential tilting method, it is also referred to as a microcanonical method [61] (in analogy with the microcanonical ensemble in which the energy is fixed exactly). In Fig. 1(c) we show instances of trajectories generated with these transition probabilities using $w_{t^\dagger}(n; n_0) = \delta_{n,n_0}$. Further, the likelihood ratio of Eq. (4.9) in this case reads

$$L_{t^\dagger}(n_0) = \frac{B\left(\frac{t^\dagger + X_{t^\dagger} - n_0}{2}, r, t^\dagger\right)}{w_{t^\dagger}(X_{t^\dagger}; n_0)}. \tag{5.11}$$

3. Numerical comparisons I: Homogeneity and local time target

Our first numerical experiment compares exponential tilting and backtracking for obtaining the probability that the process departing from state $n_0 = 0$ hits the target interval $\mathsf{I}_0(10) = [-10, 10]$ at terminal time $t^\dagger = 1000$ with a positive bias obtained by setting $r = 0.6$. This is a simple example that can be solved analytically and we indeed obtain

$$z_{t^\dagger}(\mathsf{I}_0(10)) = \mathsf{P}[X_{t^\dagger} \in \mathsf{I}_0(10)] = \sum_{n=-10}^{10}\mathsf{P}[X_{t^\dagger} = n|X_0 = 0] = 7.543 \times 10^{-10}. \tag{5.12}$$

The objective is thus to benchmark the two Monte Carlo methods when the quantity of interest is available. Moreover, this example enables us to assess the importance of appropriate selecting the tilting parameter and the terminal distribution, for exponential tilting and backtracking, respectively. We illustrate that their careful selections are crucial for the efficiency of these methods.

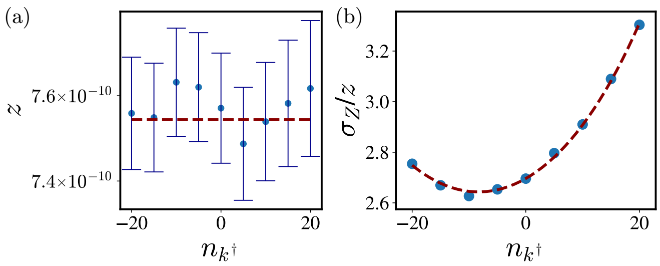

FIG. 2. (a) Exponential tilt estimators of $z_{t^\dagger}(\mathsf{I}_0(10))$ based on 10^5 realizations and for different values of the average arrival state $\langle n_{t^\dagger}\rangle$. Fixing the arrival state on average is equivalent to fixing of the tilting parameter θ in Eq. (5.6), ρ in Eq. (5.5) and q. The dashed line shows the exact value. The error bars show the 95% asymptotic normal confidence intervals. (b) Monte Carlo relative error in blue points together with the theoretical values in dashed red line. The minimal relative error is obtained by fixing the mean arrival state equal to the extreme of the target interval $\langle n_{t^\dagger} = -10\rangle$.

The homogeneity of the process and the simplicity of the target make that we can actually avoid the generation of the full trajectory and only generate, by importance sampling, the last state of the process $(X_{t^\dagger})$. We can then define the importance sampling estimators of exponential tilting and backtracking respectively as

$$Z_{\theta,t^\dagger}(\mathsf{I}_0(10)) = I_{\mathsf{I}_0(10)}(X_{t^\dagger})L_{t^\dagger}(\theta) \quad\text{and}$$
$$Z_{n_0,t^\dagger}(\mathsf{I}_0(10)) = I_{\mathsf{I}_0(10)}(X_{t^\dagger})L_{t^\dagger}(n_0). \qquad (5.13)$$

Then the probability of ending in $\mathsf{I}_0(10)$ is the expectation of these two Monte Carlo estimators w.r.t. their importance sampling distributions, precisely

$$z = z_{t^\dagger}(\mathsf{I}_0(10)) = \mathsf{E}_{\mathsf{P}_\theta}[Z_{\theta,t^\dagger}(\mathsf{I}_0(10))] = \mathsf{E}_{\mathsf{P}_{n_0}}[Z_{n_0,t^\dagger}(\mathsf{I}_0(10))]. \qquad (5.14)$$

For *exponential tilting*, we generate random arrival states from the binomial distribution and then compute their mean weighted by the likelihood ratios in Eq. (5.8). Algorithm SM1.4 in SM [43], Sec. SM 1, generally given for $\mathsf{I}_c(a)$, $a \geqslant 0$ and c integers, follows directly from Eqs. (5.13) and (5.14). In Fig. 2(a) we show that, for various values of the tilting parameter, the importance sampling estimator with the exponential tilted measure is in close agreement with the true analytical value.

We can also compute the second moment of the exponential tilt estimator as

$$z_2 = \mathsf{E}_{\mathsf{P}_\theta}\left[I_{\mathsf{I}_0(10)}(X_{t^\dagger})[L_{t^\dagger}(\theta)]^2\right]$$
$$= \sum_{n=n_0-t^\dagger}^{n_0+t^\dagger} I_{\mathsf{I}_0(10)}(n)[L_{t^\dagger}(\theta)]^2 B\left(\frac{t^\dagger+n-n_0}{2}, r, t^\dagger\right). \qquad (5.15)$$

Equation (5.15) together with Eq. (5.12) allow us to compute the relative error σ_Z/z, where $\sigma_Z = \sqrt{z_2 - z^2}$. This relative error is proportional to the square of the number of Monte Carlo replications required for target precision in the calculations; cf. pp. 158–159 [1]. Minimizing the relative error enhances efficiency, in the sense as fewer realizations are necessary in order to reach a desired level of precision.

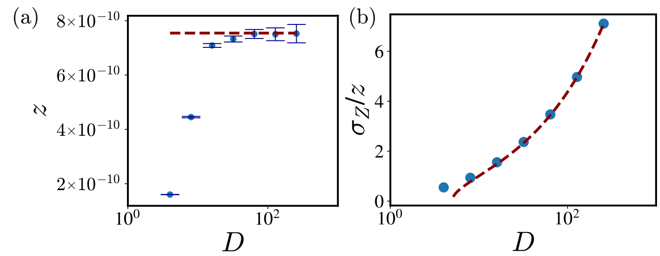

FIG. 3. (a) Results of the estimation of $z = \mathsf{P}[X_{t^\dagger} \in \mathsf{I}_0(10)]$ by using the backtracking measure, with 10^5 generations and with different radii D of the support of the uniform distribution of the final state, given in Eq. (5.16). The red dashed line indicates the exact value. The error bars provide 95% asymptotic normal confidence intervals. (b) Monte Carlo relative error, with blue dots, together with the theoretical value, with the dashed red line.

It is shown in Fig. 2(b) that empirical and theoretical values for the relative error are in agreement. Furthermore, Fig. 2(b) shows that the minimal relative error is obtained by the tilting parameter that places the average final position at the lower bound of the target interval $\langle n_{t^\dagger} = -10\rangle$. This is in agreement with the derivations of Sec. III D 1.

For *backtracking*, we need to choose the distribution of the last states $(w_{t^\dagger})$. As discussed above for the exponential tilt, it is simpler to consider the distribution of the number of positive jumps by the final time $(N_{t^\dagger})$, which is given by

$$f_{t^\dagger}(n) = \mathsf{P}[N_{t^\dagger} = n].$$

This distribution is related to the one of the terminal states through

$$f_{t^\dagger}(n) = w_{t^\dagger}(2n + n_0 - t^\dagger; n_0).$$

We consider a one-parameter family of uniform distributions,

$$f_{t^\dagger}(n) = \begin{cases} \frac{1}{D+1}, & \text{if } n \in \left[\frac{t^\dagger}{2} - D, \frac{t^\dagger}{2} + D\right], \\ 0, & \text{otherwise.} \end{cases} \qquad (5.16)$$

The support of the distribution Eq. (5.16) can be trivially modified, and so we can easily examine scenarios where the new sampling measure does not adhere to the absolute continuity requirement. In particular, selecting $D < t^\dagger$ violates $\mathsf{P} \ll \mathsf{P}_{n_0}$ and leads to the addition of bias errors (refer to Secs. II D and IV C for details).

Algorithm SM1.5 in SM [43], provides the detailed implementation of backtracking for the general target interval $\mathsf{I}_c(a)$, with any integers $a \geqslant 0$ and c. The results of the Monte Carlo study in Fig. 3(a) show that importance sampling by backtracking is in substantial agreement with the analytical value, for terminal distributions that include the target region, namely, for $D > 10$. But when $D < 10$, the backtracking measure forbids states that are important for the estimation of z. Such values of D lead to bias (systematic) errors. In contrast with this, when $D \geqslant 10$, forbidden areas do not affect the backtracking estimator, even without fulfillment of absolute continuity.

Figure 3(b) shows the influence of the width of the support of the terminal distribution ($2D$) on the relative error. As before, simulations are compared with theoretical values, obtained with the formula

$$z_2 = \mathsf{E}_{\mathsf{P}_{n_0}}\big[I_{\mathsf{I}_0(10)}(X_{t^\dagger})\big[L_{t^\dagger,n_0}(X_{t^\dagger})\big]^2\big]$$
$$= 2D\sum_{n=n_0-t^\dagger}^{n_0+t^\dagger} I_{\mathsf{I}_0(10)}(n)\left[B\left(\frac{t^\dagger+n-n_0}{2}, r, t^\dagger\right)\right]^2.$$

We see that the sampling errors are proportional to D. However, even if for $D < 10$ we observe the smaller statistical errors, discrepancies with the analytical value are bigger due to systematic errors [see Fig. 3(a)].

4. Numerical comparisons II: Homogeneity and extended time target

For the second numerical example, we evaluate the probability that the binomial process hits a target extended in time:

$$z_{t_1,t_2}(\mathsf{I}_0(10)) = \mathsf{P}[\exists s \in [t_1, t_2] : X_s \in \mathsf{I}_0(10)], \tag{5.17}$$

where $0 \leqslant t_1 < t_2 \leqslant t^\dagger$. As in the previous example, the process is biased towards the positive direction with $r = 0.6$. In this case, the hitting probability is not trivial to obtain analytically. Thus, numerical techniques are the preferred option to tackle the problem. Moreover, since the target can be hit at different times, we need to simulate the process at intermediate times.

For *exponential tilting*, we follow Algorithm SM1.2 in SM [43], Sec. SM 1, where we choose the tilting parameter according Eqs. (5.5) and (5.6) such that the tilting parameter sets the mean final state of trajectories equal to the center of the interval $\mathsf{I}_c(a)$, namely, c. Also, we can compute the estimator as

$$\hat{z}_{t_1,t_2,m}(\mathsf{I}_c(a)) = \frac{1}{m}\sum_{i=1}^{m} I\left\{\sum_{s=t_1}^{t_2} I_{\mathsf{I}_c(a)}\big(X_s^{(i)}\big) > 0\right\} L_{t^\dagger}(X_{t^\dagger}, \theta)$$
$$= \frac{1}{m}\sum_{i=1}^{m} I\left\{\sum_{s=t_1}^{t_2} I_{\mathsf{I}_c(a)}\big(X_s^{(i)}\big) > 0\right\}$$
$$\times \left(\frac{r}{q}\right)^{N_{t^\dagger}^{(i)}}\left(\frac{1-r}{1-q}\right)^{t^\dagger - N_{t^\dagger}^{(i)}}, \tag{5.18}$$

where

$$N_{t^\dagger}^{(j)} = \frac{t^\dagger + X_{t^\dagger}^{(j)}}{2}.$$

For *backtracking*, we use Algorithm SM1.3 in SM [43], Sec. SM 1, with the following uniform terminal distribution:

$$w_{t^\dagger}(n; n_0) = \begin{cases} \frac{1}{D}, & \text{if } n \in \mathsf{I}_c(a), \\ 0, & \text{otherwise,} \end{cases}$$

with $D = 2a$. Also, we can compute the Monte Carlo estimator of of $z_{t_1,t_2}(\mathsf{I}_c(a))$, given in Eq. (5.17), by

$$\hat{z}_{t_1,t_2,m}(\mathsf{I}_c(a)) = \frac{1}{m}\sum_{i=1}^{m} I\left\{\sum_{s=t_1}^{t_2} I_{\mathsf{I}_c(a)}\big(X_s^{(i)}\big) > 0\right\} L_{t^\dagger,n_0}\big(X_{t^\dagger}^{(i)}\big)$$
$$= \frac{2D}{m}\sum_{i=1}^{m} I\left\{\sum_{s=t_1}^{t_2} I_{\mathsf{I}_c(a)}\big(X_s^{(i)}\big) > 0\right\} B\big(N_{t^\dagger}^{(i)}, r, t^\dagger\big),$$

where B is the binomial probability function of Eq. (5.2).

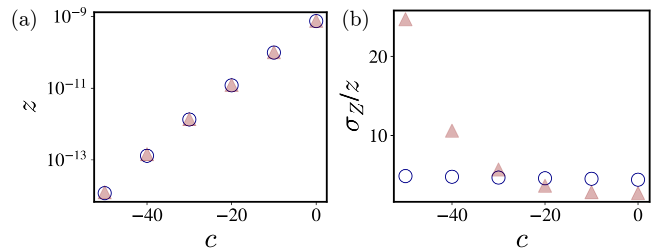

FIG. 4. (a) Results for the estimation of $z = z_{t_1,t_2}(\mathsf{I}_0(10))$ of Eq. (5.17) by backtracking in blue circles and by exponential tilting in red triangles, always with 10^5 generations and with varying values of the center of the target interval c. (b) Monte Carlo relative error obtained by these two methods. Both methods provide good estimations of z, with similar small Monte Carlo errors. Both methods display bounded relative error, which, however, increases with exponential tilting for small values of the center of the target c, namely, as the target becomes more unlikely.

In Fig. 4(a) we show that the sample average for different values of the center c computed with the two methods agree within errors. In Fig. 4(b) we also show that the relative errors in both methods have the same order of magnitude. However, the relative error for the exponential tilt increases as the event becomes rarer, whereas the errors in the backtracking method are more constant. This result is surprising provided that the backtracking process is biased [because our particular choice of $w_{t^\dagger}$ in Eq. (4.9) does not fulfill absolute continuity] whereas the exponential tilted method is unbiased.

B. Process with metastable states

In this section we study a more sophisticated and practical Markovian process where the transition probabilities in Eq. (5.1) do depend on the value of the state n. Precisely, they take the logistic form

$$p_{\bullet n}(j) = \begin{cases} \{1 + e^{j\nu n(n-\ell)(n+\ell)}\}^{-1}, & \text{for } j = -1, 1, \\ 0, & \text{otherwise,} \end{cases} \tag{5.19}$$

where n and $\ell > 0$ are integers and $\nu > 0$ is real. We note that transition probabilities in Eq. (5.19) define a process over a binomial tree (i.e., with stepwise upwards or downwards unit changes), since $p_{\bullet n}(1) + p_{\bullet n}(-1) = 1$. Additionally, this Markov process exhibits two metastable states at positions $n = \pm\ell$. This means that trajectories tend to oscillate around these two states, and it is exceptional to observe a transition from one of these states to the other one. The parameter ν tunes the robustness of the metastable states. If $\nu = 0$, then the process is an unbiased random walk without metastable states. At the opposite, namely, at the limit $\nu \to \infty$, trajectories evolve by forming a straight line (deterministically) towards either ℓ or $-\ell$.

Furthermore, the roles of ℓ and ν can be understood intuitively from a physical perspective. The transition probabilities in Eq. (5.19) can be seen as a discretization of the trajectory of a particle that, at position n, is subject to the force

$$-\tanh\frac{\nu n(n-\ell)(n+\ell)}{2},$$

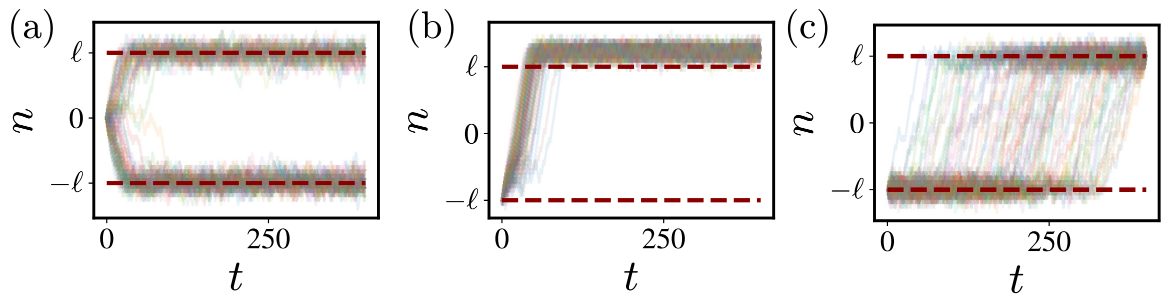

FIG. 5. Panels (a)–(c) show ensembles of 100 sample paths for $t \in \{0, \dots, 400\}$ and parameters $\nu = 10^{-3}$ and $\ell = 15$. Trajectories in (a) are generated with the original process with transition probabilities depending on the states through Eq. (5.19). All paths depart from position $n = 0$ ($\mathsf{P}[X_0 = n] = \delta_{0,n}$). The metastable states $n = \pm\ell$ are indicated by the two horizontal dashed red lines. Due to the symmetry of the transition rates around $n = 0$, on average, one half of the trajectories evolve towards the metastable state $n = \ell$, and the other half evolve towards the state $n = -\ell$. No transitions between the metastable states were observed within the given time horizon. Trajectories in (b) are generated under the tilted measure of Eq. (5.20) with tilting parameter $\theta = 1$. We note that positive values of the tilting parameter favor transitions towards the positive metastable state. We note that the location of the metastable state of the tilted process is slightly above the one of the original process. Trajectories in (c) are obtained by backtracking and transitions over a wide range of intermediate times can be observed. It is shown that the backtracking process does respect the position of the metastable states of the original process.

in the overdamped limit, i.e., in a high-viscosity regime where inertia can be neglected. In Fig. 5(a) we show typical trajectories for this process.

In order to statistically characterize these transitions, we make use of our two change-of-measure strategies.

For *exponential tilting*, by using Eq. (3.13) we obtain the following transition probabilities for any fixed value of the tilting parameter $\theta \in \mathbb{R}$:

$$p_{\bullet n}(j, \theta) = \begin{cases} \{1 + \mathrm{e}^{j[\nu n(n-\ell)(n+\ell)-2\theta]}\}^{-1}, & \text{for } j = -1, 1, \\ 0, & \text{otherwise.} \end{cases} \tag{5.20}$$

The parameter θ breaks the symmetry of the transition rates around $n = 0$ and effectively biases paths either towards the metastable state in the positive position ℓ (for $\theta > 0$) or in the negative position $-\ell$ (for $\theta < 0$). Moreover, Eq. (5.20) tells that the precise positions of the metastable states are not preserved and depend on the value of θ. This fact is illustrated numerically in Fig. 5(b).

In this case the likelihood ratio process of the exponentially tilted measure depends on the whole Markov process, and it is obtained by Eq. (3.14).

For *backtracking*, we cannot use an analytical form for the transition probabilities, and we have to evaluate the backward change-of-measure kernel of Eq. (4.5) numerically. In principle, this evaluation would require to compute the probabilities $\mathsf{P}[X_t = n|X_0 = n_0]$, for all values of n and t, through the iteration of the forward Kolmogorov equation

$$\begin{aligned} \mathsf{P}[X_{t+1} = n|X_0 = n_0] &= p_{\bullet n+1}(-1)\mathsf{P}[X_t = n+1|X_0 = n_0] \\ &\quad + p_{\bullet n-1}(1)\mathsf{P}[X_t = n-1|X_0 = n_0]. \end{aligned} \tag{5.21}$$

In order to ensure numerical tractability of Eq. (5.21), boundary conditions are introduced to the system. We set $\ell_{\max}$ and $-\ell_{\max}$ as boundaries such that

$$p_{\bullet -\ell_{\max}}(-1) = p_{\bullet \ell_{\max}}(1) = 0 \quad \text{and}$$
$$p_{\bullet -\ell_{\max}}(0) = p_{\bullet \ell_{\max}}(0) = \{1 + \mathrm{e}^{-\nu n(n-\ell_{\max})(n+\ell_{\max})}\}^{-1}.$$

These conditions imply that particles are unable to cross the boundaries, but they can avoid jumping, namely, remaining in the same state, precisely at the boundaries. This approximation does not introduce significant systematic errors in the solution as long as the probability for the process to reach the boundaries is negligible (namely, $\mathsf{P}[X_t = \pm\ell_{\max}] \simeq 0$, for $t = 1, \dots, t^\dagger$). Although it is possible to introduce alternative approximations that avoid the use of Eq. (5.21), they are not employed in this work. For more details on these alternative approaches, we refer readers to [16]. In Fig. 5(c) it is shown that the process generated by backtracking can be used to sample paths connecting metastable states.

1. Numerical comparisons III: Inhomogeneity and local time target

As with the homogeneous binomial process in Sec. V A 3, we present another example for which the desired quantity can be evaluated analytically. Let ℓ be a positive integer. We are interested in the probability that a path starting from one of the two metastable states, precisely $-\ell$, hits a one-sided interval that includes the other metastable state, ℓ, at some terminal time $t^\dagger$. This is the probability

$$z(A) = \mathsf{P}[X_{t^\dagger} \in A|X_0 = -\ell],$$

where $A = [a, \infty)$ and the integer a is such that $-\ell \notin A$ and $\ell \in A$, i.e. such that $-\ell < a < \ell$. In fact, this problem can be solved by the direct evaluation of $\mathsf{P}[X_{t^\dagger} = n|X_0 = -\ell]$ at all integers $n \in A$, by means of Eq. (5.21), and then by summing, because

$$z(A) = \sum_{n \in A} \mathsf{P}[X_{t^\dagger} = n|X_0 = -\ell]. \tag{5.22}$$

In this study, we first estimate $z(A)$ through an ensemble of paths generated by backtracking, as described in Algorithm SM1.3 in SM [43], Sec. SM 1.

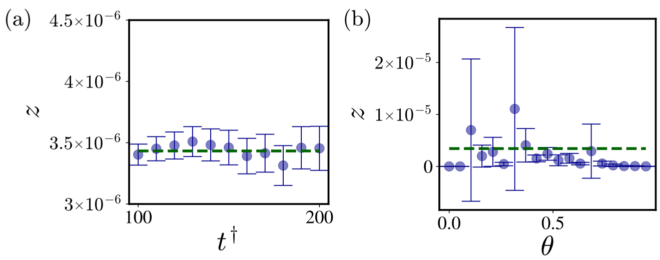

FIG. 6. Estimations of $z(A) = \mathsf{P}[X_{t^\dagger} \in A | X_0 = -\ell]$, for $A = [\ell - 2, \infty]$, $\ell = 15$, $\nu = 10^{-3}$, $t^\dagger = 100$, by backtracking in panel (a) and by exponential tilting in panel (b), always with 10^4 generations. Error bars show the 95% asymptotic normal confidence intervals.

The terminal distribution is the uniform one, given by

$$w_{t^\dagger}(n; -\ell) = \begin{cases} \frac{1}{2D+1}, & \text{if } n \in \mathsf{I}_\ell(D), \\ 0, & \text{otherwise}, \end{cases} \tag{5.23}$$

for some positive integer D.

Our first experiment considers extensions of the temporal horizon of the simulation by backtracking. We thus generate sample paths by backtracking at times $t = 0, \dots, t^*$, this for several values of $t^* \geqslant t^\dagger$. The aim is to determine the effect of the terminal time on the accuracy of the Monte Carlo estimator of $z(A)$. So the terminal distribution in Eq. (5.23) is considered at time t^*, instead of $t^\dagger$.

In Fig. 6(a) we see that the Monte Carlo estimations of $z(A)$ are close to the value computed analytically by Eq. (5.22). We also see that the Monte Carlo errors increase as t^* increases away from $t^\dagger$.

In Fig. 6(b) we show the computation of $z(A)$ over an ensemble of paths generated by Algorithm SM1.2 in SM [43], Sec. SM 1, of exponential tilting, for different values of the tilting parameter θ. In contrast with backtracking, the exponential tilting estimator does not seem to converge to $z(A)$. Furthermore, certain sample paths possess large fluctuations, visible in the error bars of Fig. 6(b). Thus, we observe that the distribution of the exponential tilting estimator exhibits heavy tails. Consequently, a reliable estimation for the first moment necessitates a very large number of simulations. As this number increases, we observe a notable improvement of the exponential tilting estimation in Fig. 7, in which exponential tilting is based on a larger number of simulations than in Fig. 6(b). Figure 7 shows the optimal tilting parameter θ is located around 0.3. Large discrepancies between the numerical estimator and the analytical result are observed for values of θ close to one, pointing out the importance of using the optimal tilting parameter for this type of application.

The example presented in this last section offers insights into the suitability of backtracking for addressing problems involving transition paths between metastable states. Remarkably, backtracking, although biased due to the violation of absolute continuity of measures, proves to be more effective in such scenarios than exponential tilting, which is unbiased.

VI. FINAL REMARKS

Problems that require sampling rare trajectories within a feasible amount of time appear in various scientific disciplines

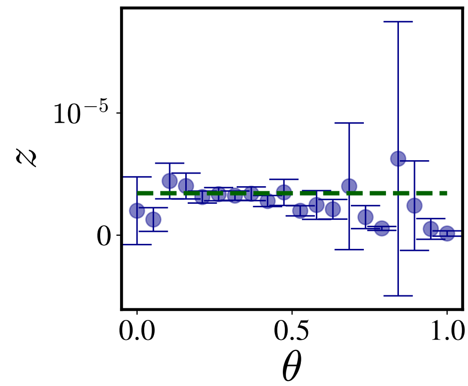

FIG. 7. Estimations of $z(A) = \mathsf{P}[X_{t^\dagger} \in A | X_0 = -\ell]$, for $A = [\ell - 2, \infty]$, $\ell = 15$, $\nu = 10^{-3}$, $t^\dagger = 100$ and by exponential tilting with 10^6 generations. Error bars show the 95% asymptotic normal confidence intervals.

(such as physics, mathematical biology, and finance). In this article, we analyze two paradigms of methods for sampling rare trajectories of Markov processes: exponential tilting and stochastic bridge. Our main contribution is to show that these two methods can be re-expressed within the general theory of change of measure and importance sampling. Through practical numerical examples, we illustrate the applicability of both Monte Carlo methods to the computation of probabilities of rare trajectories. Moreover, we provide various Monte Carlo algorithms in detailed form, so to make them directly accessible applied scientists.

We have demonstrated, both theoretically and through simulations, that techniques relying on sampling stochastic bridges may lead to systematic errors in the estimation of averages (even when the bridge generator itself is unbiased) if absolute continuity is not satisfied. To the best of the authors' knowledge, this represents a unique finding to consider when employing such methods. It highlights that calculations with a small Monte Carlo error could prove significantly inaccurate in the absence of absolute continuity.

For the case of the binomial process, we provide numerical evidence that the relative errors of these two sampling strategies have small and comparable order. In problems involving transition paths between metastable states, stochastic bridges appear to assign consistent values to the estimators for different choices of terminal distributions. In contrast, with exponential tilting the distribution of the estimator appears heavy-tailed. Consequently, we expect that a large number of Monte Carlo sampling is required in order to obtain an accurate estimation of the first moment. We nevertheless expect that exponential tilting could be particularly valuable when dealing with multidimensional processes, where the application of backtracking becomes challenging. We also keep in mind that exponential tilting is more restrictive, in the sense that it requires the existence of the cumulant generating function of the transition probabilities of the Markov process (as mentioned at the end of Sec. III C). In contrast, there is no similar restriction with backtracking.

There is an evident need for future research in this topic. In our numerical examples, we showed that there are optimal choices of the tilting parameter and of the terminal distribution, that minimize the errors of exponential tilting and backtracking, respectively. Analytical optimality results and practical formulas for the tilting parameter and for the terminal distribution are important questions that remain mainly open. Another subject of investigation could be the extension of the exponential tilting likelihood ratio in Eq. (3.14) from one to several tilting parameters: at each transition of the process, a specific tilting parameter could be used, thus depending on the current state. Finally, we could consider other practical settings for comparing backtracking and exponential tilting, such as the one of the insurer ruin with recuperation (cf., e.g., [66] in relation with spectrally negative Lévy processes).

The Python codes used for the creation of the figures of this article are freely available at [67], with appropriate credit to the authors.

ACKNOWLEDGMENTS

The authors thank the anonymous referees for constructive suggestions. They also thank Raúl Toral, Pablo I. Hurtado, Carlos Pérez-Espigares, Rubén Hurtado, Raphaël Chetrite, and David Ginsbourger for useful discussions. Research was partially supported by the University of Bern, UniBE ID Grant 2021, and the Conselleria d'Educació, Universitat i Recerca of the Balearic Islands (Grant No. FPI/006/2020). Further financial support has been received from the Agencia Estatal de Investigación (AEI, MCI, Spain) MCIN/AEI/10.13039/501100011033, Fondo Europeo de Desarrollo Regional (FEDER, UE) under Project APASOS (PID2021-122256NB-C21/C22), the María de Maeztu Program for units of Excellence in R&D, Grant No. CEX2021-001164-M, and the contract ForInDoc (GOIB).

[1] S. Asmussen and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis* (Springer, New York, 2010), Vol. 57.
[2] R. Toral and P. Colet, *Stochastic Numerical Methods: An Introduction for Students and Scientists* (John Wiley & Sons, 2014).
[3] N. Masuda and C. L. Vestergaard, *Gillespie Algorithms for Stochastic Multiagent Dynamics in Populations and Networks*, Elements in Structure and Dynamics of Complex Networks (Cambridge University Press, Cambridge, 2023).
[4] J. Aguilar, J. J. Ramasco, and R. Toral, Numerical methods for stochastic simulations: application to contagion processes, arXiv:2305.02902.
[5] R. Hurtado-Gutiérrez, F. Carollo, C. Pérez-Espigares, and P. Hurtado, Building continuous time crystals from rare events, Phys. Rev. Lett. **125**, 160601 (2020).
[6] D. Ciccarese, G. Micali, B. Borer, C. Ruan, D. Or, and D. R. Johnson, Rare and localized events stabilize microbial community composition and patterns of spatial self-organization in a fluctuating environment, ISME J. **16**, 1453 (2022).
[7] M. Mobilia and M. Assaf, Fixation in evolutionary games under non-vanishing selection, Europhys. Lett. **91**, 10002 (2010).
[8] D. A. Kessler and N. M. Shnerb, Extinction rates for fluctuation-induced metastabilities: A real-space WKB approach, J. Stat. Phys. **127**, 861 (2007).
[9] C. Frei and C. Schär, Detection probability of trends in rare events: Theory and application to heavy precipitation in the alpine region, J. Clim. **14**, 1568 (2001).
[10] A. Gabrielov, I. Zaliapin, W. I. Newman, and V. I. Keilis-Borok, Colliding cascades model for earthquake prediction, Geophys. J. Int. **143**, 427 (2000).
[11] A. Ramoso, J. Magalang, D. Sánchez-Taltavull, J. Esguerra, and É. Roldán, Stochastic resetting antiviral therapies prevent drug resistance development, Europhys. Lett. **132**, 50003 (2020).
[12] J. M. Harrison and S. I. Resnick, The stationary distribution and first exit probabilities of a storage process with general release rule, Math. Oper. Res. **1**, 347 (1976).
[13] S. Haberman, An introduction to mathematical risk theory, J. Inst. Actuaries **108**, 109 (1981).
[14] J. F. Ziegler and W. A. Lanford, The effect of sea level cosmic rays on electronic devices, J. Appl. Phys. **52**, 4305 (1981).
[15] D. Siegmund, Importance sampling in the Monte Carlo study of sequential tests, Ann. Stat. **4**, 673 (1976).
[16] J. Aguilar, J. W. Baron, T. Galla, and R. Toral, Sampling rare trajectories using stochastic bridges, Phys. Rev. E **105**, 064138 (2022).
[17] S. N. Majumdar and H. Orland, Effective langevin equations for constrained stochastic processes, J. Stat. Mech.: Theory Exp. (2015) P06039.
[18] S. Wang, A. Venkatesh, D. Ramkrishna, and V. Narsimhan, Brownian bridges for stochastic chemical processes—An approximation method based on the asymptotic behavior of the backward Fokker–Planck equation, J. Chem. Phys. **156**, 184108 (2022).
[19] S. Wang, D. Ramkrishna, and V. Narsimhan, Exact sampling of polymer conformations using Brownian bridges, J. Chem. Phys. **153**, 034901 (2020).
[20] B. De Bruyne, S. N. Majumdar, and G. Schehr, Optimal resetting Brownian bridges via enhanced fluctuations, Phys. Rev. Lett. **128**, 200603 (2022).
[21] H. Orland, Generating transition paths by Langevin bridges, J. Chem. Phys. **134**, 174114 (2011).
[22] P. Koehl and H. Orland, Sampling constrained stochastic trajectories using Brownian bridges, J. Chem. Phys. **157**, 054105 (2022).
[23] V. Lecomte and J. Tailleur, A numerical approach to large deviations in continuous time, J. Stat. Mech.: Theory Exp. (2007) P03004.
[24] C. Pérez-Espigares and P. I. Hurtado, Sampling rare events across dynamical phase transitions, Chaos **29**, 083106 (2019).
[25] T. Brewer, S. R. Clark, R. Bradford, and R. L. Jack, Efficient characterisation of large deviations using population dynamics, J. Stat. Mech.: Theory Exp. (2018) 053204.
[26] P. Grassberger, Go with the winners: A general Monte Carlo strategy, Comput. Phys. Commun. **147**, 64 (2002).
[27] G. Claussen, A. K. Hartmann, and S. N. Majumdar, Convex hulls of random walks: Large-deviation properties, Phys. Rev. E **91**, 052104 (2015).
[28] A. K. Hartmann, S. N. Majumdar, H. Schawe, and G. Schehr, The convex hull of the run-and-tumble particle in a plane, J. Stat. Mech.: Theory Exp. (2020) 053401.

[29] P. Heidelberger, *Fast Simulation of Rare Events in Queueing and Reliability Models* (Association for Computing Machinery, New York, USA, 1995), Vol. 5, pp. 43–85.

[30] J. Blanchet and H. Lam, State-dependent importance sampling for rare-event simulation: An overview and recent advances, Surv. Oper. Res. Manage. Sci. **17**, 38 (2012).

[31] A. Guyader and H. Touchette, Efficient large deviation estimation based on importance sampling, J. Stat. Phys. **181**, 551 (2020).

[32] C. Hartmann, L. Richter, C. Schütte, and W. Zhang, Variational characterization of free energy: Theory and algorithms, Entropy **19**, 626 (2017).

[33] T. H. Oakes, A. Moss, and J. P. Garrahan, A deep learning functional estimator of optimal dynamics for sampling large deviations, Mach. Learn.: Sci. Technol. **1**, 035004 (2020).

[34] R. Gatto, Stochastic simulation of rare events, in *Wiley StatsRef: Statistics Reference Online* (John Wiley & Sons, 2015), pp. 1–11.

[35] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (Springer Science & Business Media, 1998), Vol. 38.

[36] R. J. Allen, C. Valeriani, and P. R. Ten Wolde, Forward flux sampling for rare event simulations, J. Phys.: Condens. Matter **21**, 463102 (2009).

[37] S. W. Hall, G. Díaz Leines, S. Sarupria, and J. Rogal, Practical guide to replica exchange transition interface sampling and forward flux sampling, J. Chem. Phys. **156**, 200901 (2022).

[38] T. S. Van Erp, Dynamical rare event simulation techniques for equilibrium and nonequilibrium systems, Adv. Chem. Phys. **151**, 27 (2012).

[39] P. G. Bolhuis and C. Dellago, Trajectory-based rare event simulations, Rev. Comput. Chem. **27**, 111 (2010).

[40] C. Dellago and P. G. Bolhuis, Transition path sampling and other advanced simulation techniques for rare events, *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, edited by C. Holm (Springer, Berlin, Heidelberg, 2009), pp. 167–233.

[41] M. Heymann and E. Vanden-Eijnden, Pathways of maximum likelihood for rare events in nonequilibrium systems: Application to nucleation in the presence of shear, Phys. Rev. Lett. **100**, 140601 (2008).

[42] T. Grafke and E. Vanden-Eijnden, Numerical computation of rare events via large deviation theory, Chaos **29**, 063118 (2019).

[43] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.109.034113 for details in the calculation of conditioned moments of the binomial process.

[44] J. A. Bucklew and J. Bucklew, *Introduction to Rare Event Simulation* (Springer, 2004), Vol. 5.

[45] S. Asmussen, *Applied Probability and Queues* (Springer, 2003), Vol. 2.

[46] W. Paul and J. Baschnagel, *Stochastic Processes From Physics to Finance* (Springer, Berlin, Heidelberg, 2010).

[47] C. Van den Broeck *et al.*, Stochastic thermodynamics: A brief introduction, Phys. Complex Colloids **184**, 155 (2013).

[48] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Rep. Prog. Phys. **75**, 126001 (2012).

[49] J. Wang, K. Zhang, L. Xu, and E. Wang, Quantifying the Waddington landscape and biological paths for development and differentiation, Proc. Natl. Acad. Sci. USA **108**, 8257 (2011).

[50] F. Esscher, On the probability function in the collective theory of risk, Scandinavian Actuarial J. **1932**, 175 (1932).

[51] F. Esscher, On approximate computation of distribution functions when the corresponding characteristic functions are known, Scand. Actuar. J. **1963**, 78 (1963).

[52] H. E. Daniels, Saddlepoint approximations in statistics, Ann. Math. Stat. **25**, 631 (1954).

[53] E. T. Copson, *Asymptotic Expansions*, Cambridge Tracts in Mathematics (Cambridge University Press, Cambridge, 1965).

[54] R. Gatto, Saddlepoint approximations, in *Wiley StatsRef: Statistics Reference Online* (John Wiley & Sons, Ltd, 2015), pp. 1–7.

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

[56] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, First-order dynamical phase transition in models of glasses: An approach based on ensembles of histories, J. Phys. A: Math. Theor. **42**, 075007 (2009).

[57] R. L. Jack and P. Sollich, Large deviations and ensembles of trajectories in stochastic models, Prog. Theor. Phys. Suppl. **184**, 304 (2010).

[58] J. Tailleur and V. Lecomte, *AIP Conference Proceedings* (American Institute of Physics, 2009), Vol. 1091, pp. 212–219.

[59] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, Dynamical first-order phase transition in kinetically constrained models of glasses, Phys. Rev. Lett. **98**, 195702 (2007).

[60] R. Chetrite and H. Touchette, Nonequilibrium microcanonical and canonical ensembles and their equivalence, Phys. Rev. Lett. **111**, 120601 (2013).

[61] R. Chetrite and H. Touchette, *Annales Henri Poincaré* (Springer, 2015), Vol. 16, pp. 2005–2057.

[62] F. Carollo, J. P. Garrahan, I. Lesanovsky, and C. Pérez-Espigares, Making rare events typical in Markovian open quantum systems, Phys. Rev. A **98**, 010103(R) (2018).

[63] M. Delorme and K. J. Wiese, Extreme-value statistics of fractional Brownian motion bridges, Phys. Rev. E **94**, 052105 (2016).

[64] F. Mori, S. N. Majumdar, and G. Schehr, Time between the maximum and the minimum of a stochastic process, Phys. Rev. Lett. **123**, 200201 (2019).

[65] O. Bénichou, P. Krapivsky, C. Mejía-Monasterio, and G. Oshanin, Joint distributions of partial and global maxima of a Brownian bridge, J. Phys. A: Math. Theor. **49**, 335002 (2016).

[66] R. Gatto, A logarithmic efficient estimator of the probability of ruin with recuperation for spectrally negative Lévy risk processes, Stat. Probab. Lett. **99**, 177 (2015).

[67] https://github.com/jvrglr/A-Unified-Perspective-on-Exponential-Tilt-and-Bridge-Algorithms-for-Rare-Trajectories/tree/main (2024).