Rare Event Simulation for Stochastic Dynamics in Continuous Time

Letizia Angeli[1] · Stefan Grosskinsky[1] · Adam M. Johansen[1] ·
Andrea Pizzoferrato[2]

Abstract

Large deviations for additive path functionals of stochastic dynamics and related numerical approaches have attracted significant recent research interest. We focus on the question of convergence properties for cloning algorithms in continuous time, and establish connections to the literature of particle filters and sequential Monte Carlo methods. This enables us to derive rigorous convergence bounds for cloning algorithms which we report in this paper, with details of proofs given in a further publication. The tilted generator characterizing the large deviation rate function can be associated to non-linear processes which give rise to several representations of the dynamics and additional freedom for associated numerical approximations. We discuss these choices in detail, and combine insights from the filtering literature and cloning algorithms to compare different approaches and improve efficiency.

Keywords Dynamic large deviations · Interacting particle systems · Cloning algorithm · Sequential Monte Carlo

1 Introduction

Large deviation simulation techniques based on classical ideas of evolutionary algorithms [1,2] have been proposed under the name of 'cloning algorithms' in [3] for discrete and in [4] for continuous time processes, in order to study rare events of dynamic observables of interacting lattice gases. This approach has subsequently been applied in a wide variety of contexts (see e.g. [5–8] and references therein), and more recently, the convergence properties of the algorithm have become a subject of interest. Analytical approaches so far are based on a branching process interpretation of the algorithm in discrete time [9], with limited and mostly numerical results in continuous time [10]. Systematic errors arise from the correlation

---

✉ Stefan Grosskinsky
  S.Grosskinsky@googlemail.com

1   University of Warwick, Coventry, UK

2   Imperial College London, London, UK

structure of the cloning ensemble which can be large in practice, and several variants of the approach have been proposed to address those including e.g. a multicanonical feedback control [7], adaptive sampling methods [11] or systematic resampling [12]. A recent survey of these issues and different variants of cloning algorithms in discrete and continuous time can be found in [13, Sect. 3].

In this paper we provide a novel perspective on the underlying structure of the cloning algorithm, which is in fact well established in the statistics and applied probability literature on Feynman–Kac models and particle filters [14–16]. The framework we develop here can be used to generalize rigorous convergence results in [17] to the setting of continuous-time cloning algorithms as introduced in [4]. Full mathematical details of this work are published in [18], and here we focus on describing the underlying approach and report the main convergence results. A second motivation is to use different McKean interpretations of Feynman–Kac semigroups (see Sect. 2.2) to highlight several degrees of freedom in the design of cloning-type algorithms that can be used to improve performance. We illustrate this with the example of current large deviations for the inclusion process (originally introduced in [19]), aspects of which have previously been studied [20]. Current fluctuations in stochastic lattice gases have attracted significant recent research interest (see e.g. [21–23] and references therein), and are one of the main application areas of cloning algorithms which are particularly challenging. In contrast to previous work in the context of cloning algorithms [9,10], our mathematical approach does not require a time discretization and works in a very general setting of a jump Markov process on a compact state space. This covers in particular any finite state Markov chain or stochastic lattice gas on a finite lattice.

The paper is organized as follows. In Sect. 2 we introduce notation, the Feynman–Kac semigroup and several representations of the associated non-linear process. In Sect. 3 we describe different particle approximations including the cloning algorithm, and summarize results published in [18] on convergence properties of estimators based on the latter. In Sect. 4 we describe a modification of the cloning algorithm for a particular class of stochastic lattice gases and apply it to the inclusion process as an example.

2 Mathematical Setting

2.1 Large Deviations and the Tilted Generator

We consider a continuous-time Markov jump process $\big(X(t) : t \geq 0\big)$ on a compact state space E. To fix ideas we can think of a finite state Markov chain, such as a stochastic lattice gas on a finite lattice Λ with a fixed number of particles M. Here E is of the form S^Λ with a finite set S of local states (e.g. $S = \{0, 1\}$ or $\{0, \ldots, M\}$), but continuous settings with compact $E \subset \mathbb{R}^d$ for any $d \geq 1$ are also included. One can in principle also generalize to separable and locally compact state spaces, including countable Markov chains and lattice gases on finite lattices with open boundaries. But this would require more effort and complicate not only the proof but also the presentation of the main results for technical reasons which we want to avoid here (see [18] for a more detailed discussion).

Jump rates are given by the kernel $W(x, dy)$, such that for all $x \in E$ and measurable subsets $A \subset E$

$$\mathbb{P}\big[X(t + \Delta t) \in A \big| X(t) = x\big] = \Delta t \int_A W(x, dy) + o(\Delta t) \quad \text{as } \Delta t \to 0, \qquad (2.1)$$

where $o(\Delta t)/\Delta t \to 0$. We use the standard notation $\mathbb{P}$ and $\mathbb{E}$ for the distribution and the corresponding expectation on the usual path space for jump processes

$$\Omega = \big\{\omega : [0, \infty) \to E \text{ right continuous with left limits}\big\}.$$

If we want to stress a particular initial condition $x \in E$ of the process we write $\mathbb{P}_x$ and $\mathbb{E}_x$. The process can be characterized by the infinitesimal generator

$$\mathcal{L}f(x) = \int_E W(x, dy)[f(y) - f(x)], \quad \forall f \in \mathcal{C}_b(E), \ x \in E, \tag{2.2}$$

acting on all continuous bounded functions $f \in \mathcal{C}_b(E)$ on the state space. The adjoint $\mathcal{L}^\dagger$ of this operator acts on probability distributions μ on E, and determines their time evolution via

$$\frac{d}{dt}\mu_t(dy) = \int_{x \in E} \mu_t(dx)W(x, dy) - \int_{x \in E} \mu_t(dy)W(y, dx), \tag{2.3}$$

where $\mathbb{P}[X_t \in A] = \int_A \mu_t(dy)$ for any regular $A \subset E$ characterizes the distribution of the process at time $t \geq 0$. In case of countable E, (2.3) is simply the usual master equation of the process for $\mu_t(y) = \mathbb{P}[X_t = y]$, but we focus our presentation on the equivalent description via the generator (2.2), which leads to a more compact notation and applies in the general setting. As a technical assumption we require that the total exit rate of the process is uniformly bounded

$$w(x) := \int_E W(x, dy) \leq \bar{w} < \infty \quad \text{for all } x \in E. \tag{2.4}$$

We are interested in the large deviations of additive path space observables $A_T : \Omega \to \mathbb{R}$ of the general form

$$A_T(\omega) := \sum_{\substack{t \leq T \\ \omega(t-) \neq \omega(t)}} g\big(\omega(t-), \omega(t)\big) + \int_0^T h\big(\omega(t)\big)dt, \tag{2.5}$$

where $g \in \mathcal{C}_b(E^2)$ and $h \in \mathcal{C}_b(E)$. Note that A_T is well defined since the bound on $w(x)$ implies that the sum in the first term almost surely contains only finitely many terms for any $T > 0$. The above functional, which recently appeared in this form in [24], assigns a weight via the function g to jumps of the process, as well as to the local time via the function h. Dynamics conditioned on such a functional have been studied in many contexts [5], including driven diffusions on periodic continuous spaces E [25].

As mentioned before, the simplest examples covered by our setting are Markov chains with finite state space E. This includes stochastic particle systems on a finite lattice with periodic or closed boundary conditions such as zero-range or inclusion processes [20,23,26], and also processes with open boundaries and bounded local state space such as the exclusion process [3]. Choosing g appropriately and $h \equiv 0$ the functional A_T can, for example, measure the empirical particle current across a bond of the lattice or within the whole system up to time T.

We assume that A_T admits a large deviation rate function, which is a lower semi-continuous function $I : \mathbb{R} \to [0, \infty]$ such that

$$\lim_{T \to \infty} -\frac{1}{T} \log \mathbb{P}[A_T/T \in U] = \inf_{a \in U} I(a) \tag{2.6}$$

for all regular intervals $U \subset \mathbb{R}$ (see e.g. [27,28] for a more general discussion). Based on the graphical construction of the jump process and the contraction principle, existence and

convexity of I can be established in a very general setting for countable state space (see e.g. [29] and references therein). If I is convex, it is characterized by the scaled cumulant generating function (SCGF)

$$\lambda_k := \lim_{T \to \infty} \frac{1}{T} \log \mathbb{E}\big[e^{kA_T}\big] \tag{2.7}$$

via the Legendre transform

$$I(a) = \sup_{k \in \mathbb{R}} \big(ka - \lambda_k\big). \tag{2.8}$$

It is well known (see e.g. [24,26]) that λ_k can be characterized as the principal eigenvalue of a tilted version of the generator (2.2)

$$\begin{aligned}
\mathcal{L}_k f(x) &:= \int_E W(x, dy)\big[e^{kg(x,y)} f(y) - f(x)\big] + kh(x) f(x) \\
&= \underbrace{\int_E W_k(x, dy)\big[f(y) - f(x)\big]}_{=: \widehat{\mathcal{L}}_k f(x)} + \mathcal{V}_k(x) f(x)
\end{aligned} \tag{2.9}$$

with modified rates for the jump part $\hat{\mathcal{L}}_k$

$$W_k(x, dy) := W(x, dy)e^{kg(x,y)}, \tag{2.10}$$

and potential for the diagonal part

$$\mathcal{V}_k(x) := \int_E W(x, dy)\big[e^{kg(x,y)} - 1\big] + kh(x). \tag{2.11}$$

By (2.4) and the boundedness of g and h, for each $k \in \mathbb{R}$ there exist constants such that we have uniform bounds

$$\int_E W_k(x, dy) \le \bar{w}_k \quad \text{and} \quad \mathcal{V}_k(x) \le \bar{v}_k \quad \text{for all } x \in E. \tag{2.12}$$

Note also that $\mathcal{L}_0 = \mathcal{L}$, but for any $k \ne 0$ the diagonal part of the operator does not vanish and

$$\mathcal{L}_k 1(x) = \mathcal{V}_k(x) \quad \text{for all } x \in E. \tag{2.13}$$

Still, it generates a Feynman–Kac semigroup (see e.g. [17,24] for details), defined as

$$P_t^k f(x) = \big(e^{t\mathcal{L}_k}\big) f(x) := \mathbb{E}_x\big[f e^{kA_t}\big], \tag{2.14}$$

which is the unique solution to the backward equation

$$\frac{d}{dt} P_t^k f = P_t^k (\mathcal{L}_k f) \quad \text{with} \quad P_0^k f = f. \tag{2.15}$$

Due to the diagonal part of $\mathcal{L}_k$ this does not conserve probability, i.e. for the constant function $f \equiv 1$ we get

$$P_t^k 1(x) = \mathbb{E}_x\big[e^{kA_t}\big] \ne 1 \quad \text{for all } k \ne 0, \ t > 0. \tag{2.16}$$

The associated logarithmic growth rate

$$\lambda_k(t) := \frac{1}{t} \log P_t^k 1(x) \tag{2.17}$$

provides a finite-time approximation of the SCGF λ_k, which depends on the initial condition $x \in E$. We require convergence of this approximation as $t \to \infty$ as an asymptotic stability

property of the process, discussed in detail in [18] and references therein. Under exponential mixing assumptions, which are mild in our contexts of interest, it can be shown that for some constant $C > 0$

$$\left| \lambda_k(t) - \lambda_k \right| \leq C/t \quad \text{as } t \to \infty . \tag{2.18}$$

This is for example the case if E is finite and the process necessarily has a spectral gap, as is the case for all finite state lattice gases mentioned earlier. Furthermore, exponential mixing implies that the modified finite-time approximation

$$\lambda_k(at, t) := \frac{1}{(1-a)t} \log \frac{P_t^k 1(x)}{P_{at}^k 1(x)} \quad \text{with } a \in (0, 1), \tag{2.19}$$

with a 'burn-in' time period of length at, significantly improves the convergence in (2.18) to

$$\left| \lambda_k(at, t) - \lambda_k \right| \leq C \rho^{at}/t \quad \text{as } t \to \infty \tag{2.20}$$

for some $\rho \in (0, 1)$. This is of course routinely used in Monte–Carlo sampling where systems are allowed to relax towards stationarity before measuring. These intrinsic properties of the process and the related finite-time errors in the estimation of λ_k are not the main subject of this paper. In the following we simply assume asymptotic stability and (2.18), and focus on the efficient numerical estimation of $\lambda_k(t)$ for any given $t \geq 0$. $\lambda_k(at, t)$ can be treated completely analogously, which is discussed in more detail in [18].

2.2 McKean Interpretation of the Feynman–Kac Semigroup

As usual, for a given initial distribution ν_0^k on E the semigroup (2.14) determines a measure ν_t^k at times $t \geq 0$ on E, which can be characterized weakly through integrals of bounded test functions $f \in C_b(E)$ as

$$\nu_t^k(f) := \int_E f(x) \nu_t^k(dx) = \int_E P_t^k f(x) \nu_0^k(dx). \tag{2.21}$$

Here and in the following we use the common short notation $\nu_t^k(f)$ for the integral of the function f under the measure ν_t^k to simplify notation. Note that we can write (2.17) as

$$\lambda_k(t) = \frac{1}{t} \log \nu_t^k(1), \tag{2.22}$$

with a more general initial condition ν_0^k. But since P_t^k does not conserve probability, ν_t^k is not a normalized probability measure and it is consequently impossible to sample from it.

With (2.16) $\nu_t^k(1) > 0$ for all $t \geq 0$, and we can define normalized versions of the measures via

$$\mu_t^k(f) := \nu_t^k(f)/\nu_t^k(1). \tag{2.23}$$

Using (2.9), (2.13) and (2.15) on can derive the evolution equation

$$\begin{aligned}
\frac{d}{dt} \mu_t^k(f) = \frac{d}{dt} \frac{\nu_t^k(f)}{\nu_t^k(1)} &= \frac{1}{\nu_t^k(1)} \nu_t^k(\mathcal{L}_k f) - \frac{\nu_t^k(f)}{\nu_t^k(1)^2} \nu_t^k(\mathcal{L}_k 1) \\
&= \mu_t^k(\mathcal{L}_k f) - \mu_t^k(f) \mu_t^k(\mathcal{L}_k 1) \\
&= \mu_t^k(\widehat{\mathcal{L}}_k f) + \mu_t^k(\mathcal{V}_k f) - \mu_t^k(f) \mu_t^k(\mathcal{V}_k)
\end{aligned} \tag{2.24}$$

with initial condition $\mu_0^k = \nu_0^k$. It can be shown by similar direct computation of $\frac{d}{dt} \log \nu_t^k(1)$, using (2.13) and (2.22) that

$$\lambda_k(t) = \frac{1}{t} \int_0^t \mu_s^k(\mathcal{V}_k) ds. \tag{2.25}$$

So the finite-time approximation of λ_k is given by an ergodic average with respect to the distribution μ_t^k, depending on the initial distribution μ_0^k, with an obvious modification for (2.19). The asymptotic stability of the original process implies that $\mu_t^k \to \mu_\infty^k$ converges to a unique stationary distribution μ_∞^k on E, so that the SCGF (2.7) can be written as the stationary expectation of the potential

$$\lambda_k = \mu_\infty^k(\mathcal{V}_k).$$

Due to the non-linear nature of (2.24), μ_∞^k is characterized by stationarity as the solution of the non-linear equation

$$\mu_\infty^k(\widehat{\mathcal{L}}_k f) = \mu_\infty^k(f) \mu_\infty^k(\mathcal{V}_k) - \mu_\infty^k(\mathcal{V}_k f) \quad \text{for all } f \in C_b(E).$$

Usually μ_∞^k cannot be evaluated explicitly, but from (2.24) it is possible to define a generic processes $\big(X_k(t) : t \geq 0\big)$ with time-marginals μ_t^k, and then use Monte Carlo sampling techniques. The first term of (2.24) already corresponds to a jump process with generator $\widehat{\mathcal{L}}_k$, and we have to rewrite the second non-linear part to be of the form of a generator. There is some freedom at this stage, and we report three common choices from the applied probability literature on particle approximations [17,30], one of which corresponds to the approach in [3,4], and to the best of the authors' knowledge the other two have not been considered in the computational physics literature so far.

For every probability distribution μ on E we can write

$$\mu(\mathcal{V}_k f) - \mu(f) \mu(\mathcal{V}_k) = \mu\big(\mathcal{L}_{k,\mu,c}^- f + \mathcal{L}_{k,\mu,c}^+ f\big), \tag{2.26}$$

where

$$\mathcal{L}_{k,\mu,c}^- f(x) := \big(\mathcal{V}_k(x) - c\big)^- \int_E \big(f(y) - f(x)\big) \mu(dy) \tag{2.27}$$

and

$$\mathcal{L}_{k,\mu,c}^+ f(x) := \int_E \big(\mathcal{V}_k(y) - c\big)^+ \big(f(y) - f(x)\big) \mu(dy), \tag{2.28}$$

using the standard notation $a^+ = \max\{0, a\}$ and $a^- = \max\{0, -a\}$ for positive and negative part of $a \in \mathbb{R}$. We have the freedom to introduce an arbitrary constant $c \in \mathbb{R}$, possibly depending also on the measure μ (but not the state $x \in E$), since the left-hand side of (2.26) is invariant under renormalization of the potential $\mathcal{V}_k(x) \to \mathcal{V}_k(x) - c$. The generators $\mathcal{L}_{k,\mu,c}^-$ and $\mathcal{L}_{k,\mu,c}^+$ describe jump processes on E with rates depending on the probability measure μ. $\mathcal{V}_k(x)$ can be interpreted as a fitness potential for the process, and play exactly that role in the particle approximation of this process based on population dynamics, which is presented in Sect. 3. Generic choices are:

- $c = 0$ is the default and simplest choice, but is usually not optimal as discussed in Sect. 4.
- $c = \mu_t^k(\mathcal{V}_k)$ corresponding to the average potential: If the system in state x is less fit than c it jumps to state y chosen from the distribution $\mu_t^k(dy)$ according to (2.27), and independently, the system jumps to states fitter than c irrespective of its current state according to (2.28).
- $c = \sup_{x \in E} \mathcal{V}_k(x)$ or $\inf_{x \in E} \mathcal{V}_k(x)$, so that $\mathcal{L}_{k,\mu,c}^+(f)(x) \equiv 0$ or $\mathcal{L}_{k,\mu,c}^-(f)(x) \equiv 0$, respectively, and only one of the two processes has to be implemented in a simulation.

Another representation of the non-linear part in (2.24) is (see e.g. [31, Sect. 5.3.1])

$$\mathcal{L}_{k,\mu}^{\mathcal{V}} f(x) := \int_E \big(\mathcal{V}_k(y) - \mathcal{V}_k(x)\big)^+ \big(f(y) - f(x)\big)\mu(dy), \tag{2.29}$$

which is particularly interesting for implementing efficient selection dynamics as discussed in Sect. 4. Here every jump from this part of the generator strictly increases the fitness of the process, which is a stronger version of the previous idea where the process on average increased its fitness above level c. The rate depends on departure state x and target state y, which is in general computationally more expensive to implement than rates in (2.27) and (2.28), but can still be feasible due to simplifications in many concrete examples as demonstrated in Sect. 4. A further improvement of that idea is given by

$$\mathcal{L}_{k,\mu}^{\mathcal{V}} f(x) := \big(\mathcal{V}_k(x) - \mu(\mathcal{V}_k)\big)^-$$
$$\int_E \frac{\big(\mathcal{V}_k(y) - \mu(\mathcal{V}_k)\big)^+}{\mu\big((\mathcal{V}_k - \mu(\mathcal{V}_k))^+\big)}\big(f(y) - f(x)\big)\mu(dy), \tag{2.30}$$

which resembles a continuous-time version of selection processes which are known under the names of stochastic remainder sampling [32] or residual sampling [33] in discrete time. Here selection events change the process from states x of less than average fitness $\mu(\mathcal{V}_k)$ to states y fitter than average, but we will see in Sect. 4 that this variant is harder to implement than (2.29) in our area of interest, and offers only limited extra gain in selection efficiency.

In summary, the evolution equation (2.24) for μ_t^k can be written as

$$\frac{d}{dt}\mu_t^k(f) = \mu_t^k(\widehat{\mathcal{L}}_k f) + \mu_t(\mathcal{L}_{k,\mu_t}^{\mathcal{V}} f) \tag{2.31}$$

where the first choice with (2.27) and (2.28) is included defining $\mathcal{L}_{k,\mu}^{\mathcal{V}} = \mathcal{L}_{k,\mu,c}^- + \mathcal{L}_{k,\mu,c}^+$ in that case. This defines a Markov process $\big(X_k(t) : t \geq 0\big)$ on the state space E with generator

$$\mathcal{L}_{k,\mu_t^k} f(x) := \widehat{\mathcal{L}}_k f(x) + \mathcal{L}_{k,\mu_t^k}^{\mathcal{V}} f(x). \tag{2.32}$$

The process is non-linear since the transition rates in the generator $\mathcal{L}_{k,\mu_t^k}$ depend on the distribution μ_t^k of the process at time t, and in particular the process is also time-inhomogeneous. While the generator is still a linear operator acting on test functions f, the adjoint $\mathcal{L}_{k,\mu_t^k}^{\dagger}$ is a non-linear operator acting on measures μ_t^k, generating their time evolution via

$$\frac{d}{dt}\mu_t^k(dy) = \mathcal{L}_{k,\mu_t^k}^{\dagger}\mu_t^k, \tag{2.33}$$

which is equivalent to (2.31). This microscopic mass transport description consistent with the macroscopic description provided by the Feynman–Kac semigroup P_t^k is also called a McKean representation [16,31]. It is well know that particle approximations of different McKean representations can have very different properties. The first part is similar to the original dynamics with modified rates W_k (2.10), and the second non-linear part depending on the distribution μ_t^k arises from normalizing the measures ν_t^k. Note that μ_t^k and therefore the finite-time approximation $\lambda_k(t)$ in (2.25) are uniquely determined by (2.24), and thus independent of the different McKean representations, as are of course the limiting quantities μ_∞^k and λ_k. Also, these interpretations do not make use of concepts from population dynamics such as branching, which will only come into play when using particle approximations of the measures μ_t^k as explained in the next section.

3 Particle Approximations and the Cloning Algorithm

The rates of the non-linear process $\left(X_k(t) : t \geq 0\right)$ (2.32) depend on the distribution μ_t, which is not known a-priori in the cases in which we are interested. The natural framework to sample such non-linear processes approximately is a particle approximation, see e.g. [30]. Here an ensemble $\left(\underline{X}_k(t) : t \geq 0\right)$ of N processes (also called particles or clones) $X_k^i(t)$, $i = 1, \ldots, N$ is run in parallel on the state space E^N, and μ_t^k is approximated by the empirical distribution $\mu^N(\underline{X}_k(t))$ of the realizations, where for any $\underline{x} \in E^N$ we define

$$\mu^N(\underline{x})(dy) := \frac{1}{N} \sum_{i=1}^N \delta_{x_i}(dy) \quad \text{as a distribution on } E. \tag{3.1}$$

Since $\mu^N(\underline{X}_k(t))$ is fully determined by the state of the ensemble at time t, the particle approximation is a standard (linear) Markov process on E^N. This leads to an estimator for the SCGF using (2.25) given by

$$\Lambda_k^N(t) := \frac{1}{t} \int_0^t \mu^N(\underline{X}_k(s))(\mathcal{V}_k) ds = \frac{1}{t} \int_0^t \frac{1}{N} \sum_{i=1}^N \mathcal{V}_k(X_k^i(s)) ds, \tag{3.2}$$

which is a random object depending on the realization of the particle approximation. The full dynamics can be set up in various different ways such that $\mu^N(\underline{X}_k(t)) \to \mu_t^k$ converges as $N \to \infty$ for any $t \geq 0$. A generic version, directly related to the above McKean representations has been studied in the applied probability literature in great detail [17,30], providing quantitative control on error bounds for convergence. After describing this approach, we present a different approach known in the theoretical physics literature under the name of cloning algorithms [5,13], which provides some computational advantages but lacks general rigorous error control so far [9,10]. We will then set up a framework to identify common aspects of both approaches, which can be used to generalize existing convergence results to obtain rigorous error bounds for cloning algorithms as described in detail in [18], and to compare computational efficiency of both approaches.

3.1 Basic Particle Approximations

The most basic particle approximation is simply to run the McKean dynamics (2.32) in parallel on each of the particles, replacing the dependence on μ_t^k by $\mu^N(\underline{X}_k(t))$ in the jump rates. Mathematically, denoting by $\mathcal{L}_k^N$ the generator of the full N particle process $\left(\underline{X}_k(t) : t \geq 0\right)$ acting on functions $F : E^N \to \mathbb{R}$, this corresponds to

$$\mathcal{L}_k^N F(\underline{x}) := \sum_{i=1}^N \mathcal{L}_{k,\mu^N(\underline{x})}^i F(\underline{x}). \tag{3.3}$$

Here $\mathcal{L}_{k,\mu^N(\underline{x})}^i$ is equivalent to (2.32) acting on particle i only, i.e. on the function $x^i \mapsto F(\underline{x})$ while x^j, $j \neq i$ remain fixed. The linear part $\widehat{\mathcal{L}}_k$ of (2.32) does not depend on μ_t^k and follows the original dynamics for each particle, referred to as 'mutation' events in the standard population dynamics interpretation. In this context, the non-linear parts (2.27) and (2.28) can be interpreted as 'selection' events leading to mean-field interactions between the particles. Using the definition (3.1) of the empirical measures, we can write for the part (2.27)

$$\mathcal{L}^{-,i}_{k,\mu^N(\underline{x}),c} F(\underline{x}) = \left(\mathcal{V}_k(x_i) - c\right)^- \int_E \left(F(\underline{x}^{i,y}) - F(\underline{x})\right) \mu^N(\underline{x})(dy)$$

$$= \left(\mathcal{V}_k(x_i) - c\right)^- \frac{1}{N} \sum_{j=1}^N \left(F(\underline{x}^{i,x_j}) - F(\underline{x})\right) \tag{3.4}$$

with notation $\underline{x}^{i,y} = (x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_N)$. So with a rate depending on the fitness of particle i, it is 'killed' and replaced by a copy of particle j uniformly chosen from all particles. Analogously, we have for (2.28)

$$\mathcal{L}^{+,i}_{k,\mu^N(\underline{x}),c} F(\underline{x}) = \frac{1}{N} \sum_{j=1}^N \left(\mathcal{V}_k(x_j) - c\right)^+ \left(F(\underline{x}^{i,x_j}) - F(\underline{x})\right), \tag{3.5}$$

which leads to the same transition $\underline{x} \to \underline{x}^{i,x_j}$, but with a different interpretation. Each particle j in the system reproduces independently with a rate depending on its fitness (cloning event), and its offspring replaces a uniformly chosen particle, which is equal to i with probability $1/N$. The different nature of killing and cloning events becomes clearer when we write out the full generator (3.3) and switch summation indices for the cloning part (3.5) in the second line,

$$\mathcal{L}^N_k F(\underline{x}) = \sum_{i=1}^N \int_E W_k(x_i, dy)\left(F(\underline{x}^{i,y}) - F(\underline{x})\right)$$

$$+ \sum_{i=1}^N \left(\mathcal{V}_k(x_i) - c\right)^+ \frac{1}{N} \sum_{j=1}^N \left(F(\underline{x}^{j,x_i}) - F(\underline{x})\right)$$

$$+ \sum_{i=1}^N \left(\mathcal{V}_k(x_i) - c\right)^- \frac{1}{N} \sum_{j=1}^N \left(F(\underline{x}^{i,x_j}) - F(\underline{x})\right). \tag{3.6}$$

Analogously, the McKean representations (2.29) and (2.30) lead to basic N-particle systems with generators

$$\mathcal{L}^N_k F(\underline{x}) = \sum_{i=1}^N \int_E W_k(x_i, dy)\left(F(\underline{x}^{i,y}) - F(\underline{x})\right)$$

$$+ \frac{1}{N} \sum_{i,j=1}^N \left(\mathcal{V}_k(x_j) - \mathcal{V}_k(x_i)\right)^+ \left(F(\underline{x}^{i,x_j}) - F(\underline{x})\right) \tag{3.7}$$

and

$$\mathcal{L}^N_k F(\underline{x}) = \sum_{i=1}^N \int_E W_k(x_i, dy)\left(F(\underline{x}^{i,y}) - F(\underline{x})\right)$$

$$+ \frac{1}{N} \sum_{i,j=1}^N \left(\mathcal{V}_k(x_i) - \mu(\mathcal{V}_k)\right)^- \frac{\left(\mathcal{V}_k(x_j) - \mu(\mathcal{V}_k)\right)^+}{\mu\left((\mathcal{V}_k - \mu(\mathcal{V}_k))^+\right)} \left(F(\underline{x}^{i,x_j}) - F(\underline{x})\right). \tag{3.8}$$

Here a particle i is replaced by a particle j with higher fitness, combining killing and cloning into a single event. In the case of (3.8), particle i is furthermore less fit and j is fitter than average. Note that these approximating systems can be seen as particle systems with mean-field or averaged pairwise interaction given by the selection dynamics.

Following established results in [17,30,31], the (random) quantity $\Lambda_k^N(t)$ is an asymptotically unbiased estimator of $\lambda_k(t)$ with a systematic error bounded by

$$\sup_{t \geq 0} \left| \mathbb{E}^N\big[\Lambda_k^N(t)\big] - \lambda_k(t) \right| \leq \frac{C}{N} \quad \text{for all } N \geq 1, \tag{3.9}$$

along with several rigorous convergence results. These include an estimate on the random error in L^p norm for any $p > 1$,

$$\sup_{t \geq 0} \mathbb{E}^N\big[|\Lambda_k^N(t) - \lambda_k(t)|^p\big]^{1/p} \leq \frac{C_p}{\sqrt{N}} \quad \text{for all } N \geq 1, \tag{3.10}$$

as well as other formulations including almost sure convergence. Note that these estimates are uniform in $t \geq 0$, so are not affected by the choice of simulation time. The use of a finite simulation time, t, leads to an additional systematic error to the estimate of the SCGF λ_k, of order $1/t$ as in (2.18) or ρ^{at}/t as in (2.20). The bound (3.10) for $p = 2$ implies for the variance

$$\sup_{t \geq 0} \mathbb{E}^N\left[\left|\Lambda_k^N(t) - \mathbb{E}^N\big[\Lambda_k^N(t)\big]\right|^2\right] \leq \frac{C_2^2}{N} \quad \text{for all } N \geq 1, \tag{3.11}$$

since we have $\mathrm{Var}(Y) = \inf_{a \in \mathbb{R}} \mathbb{E}\big[(Y - a)^2\big]$ for any real-valued random variable Y. Therefore, error bars based on standard deviations are of the usual Monte Carlo order of $1/\sqrt{N}$, and the random error dominates the systematic bias (3.9) for N large enough. Further remarks on possible unbiased estimators can be found at the end of the next subsection.

3.2 Essential Properties of Particle Approximations

Following the standard martingale characterization of Feller-type Markov processes (see e.g. [34], Chap. 3), we know that for every bounded, continuous $F \in C_b(E^N)$

$$\mathcal{M}_F^N(t) := F\big(\underline{X}_k(t)\big) - F\big(\underline{X}_k(0)\big) - \int_0^t \mathcal{L}_k^N F\big(\underline{X}_k(s)\big) ds \tag{3.12}$$

is a martingale on $\mathbb{R}$ with (predictable) quadratic variation

$$\langle \mathcal{M}_F^N \rangle(t) = \int_0^t \Gamma_k^N F\big(\underline{X}_k(s)\big) ds, \tag{3.13}$$

where the associated carré du champ operator Γ_k^N is given by

$$\Gamma_k^N F(\underline{x}) := \mathcal{L}_k^N F^2(\underline{x}) - 2F(\underline{x})\mathcal{L}_k^N F(\underline{x}). \tag{3.14}$$

In analogy to the decomposition of a random variable into its mean and a centred fluctuating part, the martingale (3.12) describes the fluctuations of the process $t \mapsto F\big(\underline{X}_k(t)\big)$. The strength of the noise depends on time and is given by the increasing process (3.13), whose time evolution is generated by the carré du champ operator (3.14). In contrast to the generator $\mathcal{L}_k^N$, this is a quadratic (non-linear) operator and it is the main tool for studying the fluctuations of a process.

Elementary computations for approximations (3.6), (3.7) and (3.8) show that for marginal test functions $F(\underline{x}) = f(x_i)$ depending only on a single particle, we have

$$\mathcal{L}_k^N F(\underline{x}) = \mathcal{L}_{k,\mu^N(\underline{x})} f(x_i) \quad \text{and} \quad \Gamma_k^N F(\underline{x}) = \Gamma_{k,\mu^N(\underline{x})} f(x_i).$$

So generator and carré du champ both coincide with the corresponding operators $\mathcal{L}_{k,\mu^N(\underline{x})}$ and $\Gamma_{k,\mu^N(\underline{x})}$ for the McKean dynamics (2.32). This means that for large enough N and $\mu^N\big(\underline{X}_k(t)\big)$ close to μ_t^k, each marginal process $t \mapsto X_k^i(t)$ has essentially the same distribution as the corresponding McKean process $t \mapsto X_k(t)$. Note that due to selection events these (auxiliary) dynamics do not coincide with the original process conditioned on a large deviation event, and they are also not unique since there are various choices for McKean representations of Feynman–Kac semigroups, as discussed earlier. Trajectories in a particle approximation always correspond to the trajectories of the particular McKean interpretation they are based on, which is usually (2.26) in the context of cloning algorithms. Due to asymptotic stability the particle approximation converges as $t \to \infty$ for fixed N to a unique stationary distribution $\mu_\infty^{N,k}$, and the single-particle marginals of this distribution converge to μ_∞^k as $N \to \infty$. While the marginal processes for a given particle approximation are identically distributed they are not independent, and $\mu_\infty^{N,k}$ exhibits non-trivial correlations between particles resulting from selection events, which we discuss again in Sect. 4.

Now consider averaged observables of the form

$$F(\underline{x}) = \frac{1}{N} \sum_{i=1}^N f(x_i) = \mu^N(\underline{x})(f)$$

as they appear in the eigenvalue estimator (3.2). Since the generator $\mathcal{L}_k^N$ is a linear operator in F, we have the same identity as above for the generator,

$$\mathcal{L}_k^N \mu^N(\underline{x})(f) = \mu^N(\underline{x})\big(\mathcal{L}_{k,\mu^N(\underline{x})}f\big). \tag{3.15}$$

The carré du champ, on the other hand, is non-linear in F and the dependence between particles is captured by this operator. Since for all approximations (3.6), (3.7) and (3.8) in every selection event only a single particle is affected, another elementary, slightly more cumbersome computation shows (see [18] for details)

$$\Gamma_k^N \mu^N(\underline{x})(f) = \frac{1}{N} \mu^N(\underline{x})\big(\Gamma_{k,\mu^N(\underline{x})}f\big). \tag{3.16}$$

The factor $1/N$ results from a self-averaging property of the mean-field interaction through selection dynamics, which is expected from results on other mean-field particle systems (see e.g. [35,36] and references therein), and is fully analogous to the central-limit type scaling of the empirical variance for the sum of N independent random variables. While this scaling remains the same for more general particle approximations with more than one particle being affected by selection events, the simple identity (3.16) does not hold exactly for any $N \geq 1$ as we see in the next subsection.

Recall that the estimator (3.2) for the principal eigenvalue (2.7) is given by an ergodic integral of the average observable $F(\underline{x}) = \mu^N(\underline{x})(\mathcal{V}_k)$. With (2.12) $\mathcal{V}_k \in C_b(E)$ and rates are bounded, so $\mu^N(\underline{x})\big(\Gamma_{k,\mu^N(\underline{x})}\mathcal{V}_k\big)$ is also bounded and the carré du champ (3.16) vanishes as $N \to \infty$. Therefore the martingale $\mathcal{M}_F^N(t)$ also vanishes[1] for all $t \geq 0$, leading to a convergence of the measures $\mu^N(\underline{X}_k(t)) \to \mu_t^k(t)$ and also of finite time approximations $\Lambda_k^N(t) \to \lambda_k(t)$ as reported in the previous subsection. Due to the time-normalization in (3.2) and the assumed ergodicity, corresponding error bounds hold uniformly in $t \geq 0$. In summary, bounds on the carré du champ are the main ingredient for the proof of convergence results as explained in detail in [18] and references therein. All above properties up to and including (3.15) are generic requirements for any particle approximation. These particle

---

[1] in L^p-sense for any $p > 1$ following with the Burkholder–Davis–Gundy inequality, see e.g. [37, Sect. 11]

approximations can differ in their correlation structures and this freedom can be used to construct numerically more efficient particle approximations as discussed in the next subsection. To optimize sampling, particles should ideally evolve in as uncorrelated a fashion as possible; it is not possible to achieve completely independent evolution due to the non-linearity of the underlying McKean process and resulting selection events and mean-field interactions.

Remarks on Unbiased Estimators. Estimators based on expectations w.r.t. the empirical measures $\mu_t^N = \mu^N(\underline{X}_k(t))$ usually have a bias, i.e. $\mathbb{E}\big[\mu_t^N(f)\big] \neq \mu_t(f)$ for $f \in \mathcal{C}_b(E)$, which vanishes only asymptotically (3.9). This originates from the non-linear time evolution of μ_t^k (2.24) and associated McKean processes. It is straightforward to derive estimators based on the unnormalized measures ν_t^k (2.21) that are unbiased. Based on (2.25) and (3.2), we obtain an estimate of the normalization $\nu_t^k(1)$:

$$\nu_t^N(1) := \exp\Big(\int_0^t \mu_s^N(\mathcal{V}_k)ds\Big), \tag{3.17}$$

and then introduce unnormalized empirical measures on E at time t based on the particle approximation

$$\nu_t^N(f) := \nu_t^N(1)\mu_t^N(f) \quad \text{for all } f \in \mathcal{C}_b(E).$$

The expected time evolution of observables f is then given by

$$\frac{d}{dt}\mathbb{E}\big[\nu_t^N(f)\big] = \mathbb{E}\big[\nu_t^N(f)\mu_t^N(\mathcal{V}_k) + \nu_t^N(1)\mathcal{L}_k^N\mu_t^N(f)\big]. \tag{3.18}$$

Now with (3.15) and the decomposition (2.32) of $\mathcal{L}_{k,\mu_t^N}$ into mutation and selection part, we have

$$\mathcal{L}_k^N\mu_t^N(f) = \mu_t^N\big(\widehat{\mathcal{L}}_k f + \mathcal{L}_{k,\mu_t^N}^{\mathcal{V}} f\big),$$

and with the general construction of McKean representations (2.26)

$$\mu_t^N(\mathcal{L}_{k,\mu_t^N}^{\mathcal{V}} f) = \mu_t^N(\mathcal{V}_k f) - \mu_t^N(f)\mu_t^N(\mathcal{V}_k).$$

Inserting into (3.18), this simplifies to

$$\frac{d}{dt}\mathbb{E}\big[\nu_t^N(f)\big] = \mathbb{E}\big[\nu_t^N(\widehat{\mathcal{L}}_k f) + \nu_t^N(\mathcal{V}_k f)\big].$$

Since with (2.15) $\mathcal{L}_k = \widehat{\mathcal{L}}_k + \mathcal{V}_k$ also generates the time evolution of $\nu_t(f)$, a simple Gronwall argument with $\mathbb{E}\big[\nu_0^N(f)\big] = \nu_0(f)$ gives

$$\mathbb{E}\big[\nu_t^N(f)\big] = \nu_t(f) \quad \text{for all } t \geq 0 \text{ and } N \geq 1. \tag{3.19}$$

Note that choosing $f \equiv 1$ implies that the normalization (3.17) is an unbiased estimator of $e^{t\lambda_k(t)}$, which we will see again in Sect. 3.4 in the context of cloning algorithms. However, in practice the random error dominates the accuracy of estimates of $\lambda_k(t)$, so N has to be chosen large and the bias of the estimator $\Lambda_k^N(t)$ (3.2) is negligible.

3.3 Cloning Algorithms

Cloning algorithms proposed in the theoretical physics literature [3,4] are similar to the particle approximation (3.6), using the same tilted rates W_k for mutations, but combining the cloning and mutation part of the generator. We focus the following exposition around the

algorithm proposed in [4], but other continuous-time versions can be analysed analogously. The idea is simply to sample the cloning process for each particle i together with the mutation process at the same rate

$$w_k(x_i) := \int_E W_k(x_i, dy) = \int_E w(x_i, dy) e^{kg(x_i, y)}.$$

In each combined event, a random number of clones is generated with a distribution $p_{k,x_i}^N(n)$ such that its expectation is

$$m_k^N(x_i) := \sum_{n=0}^N n p_{k,x_i}^N(n) = (V_k(x_i) - c)^+ / w_k(x_i). \tag{3.20}$$

These clones then replace n particles chosen uniformly at random (in the sense that all subsets of size n are equally probable) from the ensemble. In this way, the rate at which a clone of particle i replaces any given particle j is

$$w_k(x_i) \frac{m_k^N(x_i)}{N} = \frac{1}{N} (V_k(x_i) - c)^+,$$

as required for $\mathcal{L}_k^N$ in (3.6). The only additional assumption on $p_{k,x_i}^N(n)$ is that the range of possible values for n has to be bounded by N for the cloning event to be well defined. Since its mean is bounded by $\max_{x \in E} (V_k(x) - c)^+ / w_k(x) < \infty$ independently of N, any distribution with the correct mean and finite range will lead to a valid algorithm for sufficiently large N.

The above cloning process is described by the generator

$$\mathcal{L}_k^{N,mc} F(\underline{x}) := \sum_{i=1}^N \sum_{n=0}^N \frac{1}{\binom{N}{n}} \sum_{\substack{A \subseteq \{1,...,N\} \\ |A|=n}} \int_E W_k(x_i, dy) p_{k,x_i}^N(n)$$

$$\left(F(\underline{x}^{A, x_i; i, y}) - F(\underline{x}) \right). \tag{3.21}$$

Here we have used the notation $\underline{x}^{A, w; i, y}$ for the vector $\underline{z} \in E^N$ with

$$z_j := \begin{cases} x_j & j \notin A \cup \{i\} \\ w & j \in A \setminus \{i\} \\ y & j = i, \end{cases}$$

for $j \in \{1, \ldots, N\}$ and $w, y \in E$. (3.21) combines cloning of x_i into a uniformly chosen subset A of size n, with a subsequent mutation event where the state of particle i changes to y. If we simply write $\mathcal{L}_k^{N,-}$ for the killing part (third line in (3.6)) which remains unchanged, the full generator of this cloning algorithm is given by

$$\mathcal{L}_k^{N,clone} := \mathcal{L}_k^{N,mc} + \mathcal{L}_k^{N,-}. \tag{3.22}$$

It can be shown by direct computation that for marginal test functions of the form $F(\underline{x}) = f(x_i)$ this is equivalent to the generator (3.6)

$$\mathcal{L}_k^N F(\underline{x}) = \mathcal{L}_k^{N,clone} F(\underline{x}), \tag{3.23}$$

and by linearity of generators also for all averaged functions of the form $F(\underline{x}) = \mu^N(\underline{x})(f)$. One can also show that for marginal test functions the carré du champ operators coincide, so the cloning algorithm produces marginal processes or particle trajectories with the same distribution as the simple particle approximation (3.6). For averaged test functions the change

in the correlation structure between particles is picked up by the carré du champ operator. Instead of (3.16) one can derive the following estimate for the mutation and cloning part

$$\Gamma_k^{N,mc} F(\underline{x}) \leq \frac{2}{N} \mu^N(\underline{x}) \Big(\widehat{\Gamma}_k f + \mathbb{1}(\mathcal{V}_k > c) \frac{q_k^N}{m_k^N} \Gamma_{k,\mu^N(\underline{x}),c}^+ f \Big),$$

see [18], the proof of Theorem 3.3.

Here $q_k^N(x_i) := \sum_{n=0}^N n^2 p_{k,x_i}^N(n)$ denotes the second moment of the number of clones for the particle i, and we use the decomposition (2.32) where $\widehat{\Gamma}_k f$ is the carré du champ corresponding to the mutation dynamics $\widehat{\mathcal{L}}_k$, and $\Gamma_{k,\mu^N(\underline{x}),c}^+$ the one corresponding to the cloning part (2.28). This estimate holds, of course, only for N large enough that the cloning event is well defined (see Sect. 5). Note also that $\Gamma_{k,\mu^N(\underline{x}),c}^+ f(x)$ is proportional to $(\mathcal{V}_k(x) - c)^+$ and with (3.20) $(\mathcal{V}_k(x) - c)^+ = 0$ implies $m_k^N(x) = 0$ for the expectation of the distribution $p_{k,x}^N$, leading to the indicator function $\mathbb{1}(\mathcal{V}_k > c) \in \{0, 1\}$.

This is sufficient to carry out the full proof of the convergence results mentioned in Sect. 3 based on results in [17]. This is carried out in [18] in full detail, and here we only report the main result of that work. Recall the bounds (2.12) on $\mathcal{V}_k$ and the total modified exit rate w_k.

Theorem *Denote by $\bar{\Lambda}_k^N(t)$ the eigenvalue estimator (3.2) corresponding to the cloning algorithm (3.22). Then there exist constants α, $\gamma > 0$ and α_p, $\gamma_p > 0$ such that for all N large enough*

$$\sup_{t \geq 0} \Big| \mathbb{E}^N \big[\bar{\Lambda}_k^N(t) \big] - \lambda_k(t) \Big| \leq \frac{\alpha}{N} \bar{v}_k \bar{w}_k \big(\gamma + \|q_k^N\|_\infty \big), \tag{3.24}$$

and for all $p > 1$

$$\sup_{t \geq 0} \mathbb{E}^N \big[|\bar{\Lambda}_k^N(t) - \lambda_k(t)|^p \big]^{1/p} \leq \frac{\alpha_p}{\sqrt{N}} \bar{v}_k \bar{w}_k \big(\gamma_p + \|q_k^N\|_\infty \big)^{1/2}. \tag{3.25}$$

Remarks • Choosing the normalization of the potential $c < \inf_{x \in E} \mathcal{V}_k(x)$ the killing rate in (3.6) vanishes and (3.21) describes the full generator $\mathcal{L}_k^{N,clone}$ for the cloning algorithm. This is computationally cheaper and simpler to implement, since only the mutation process has to be sampled independently for all particles, and cloning events happen simultaneously. However, as is discussed in Sect. 4, this choice in general reduces the accuracy of the estimator.

• A common choice in the physics literature for the distribution p_{k,x_i}^N of the clone size event (see e.g. [3,13]) is

$$p_{k,x_i}^N(n) = \begin{cases} m_k^N(x_i) - \lfloor m_k^N(x_i) \rfloor, & \text{for } n = \lfloor m_k^N(x_i) \rfloor + 1 \\ \lfloor m_k^N(x_i) \rfloor + 1 - m_k^N(x_i), & \text{for } n = \lfloor m_k^N(x_i) \rfloor \\ 0, & \text{otherwise} \end{cases}, \tag{3.26}$$

so the two adjacent integers to the mean are chosen with appropriate probabilities, which minimizes the variance of the distribution for a given mean. This choice therefore minimizes the contribution of the second moment q_k^N to the bound for the errors in (3.24) and (3.25), and is also simple to implement in practice.

• Due to (3.23), trajectories of individual particles follow the same law as the simple particle approximation (3.6) and therefore the same McKean process as explained in Sect. 2.2 The cloning approach can introduce additional correlations between particles due to large cloning events, which is quantified by the second moment q_k^N entering the error bounds in (3.24) and (3.25).

3.4 The Cloning Factor

In the physics literature an alternative estimator to Λ_t^N (3.2) is often used, based on a concept called the 'cloning factor' (see e.g. [3,5,13]). This is essentially a continuous-time jump process $(C_k^N(t) : t \geq 0)$ on $\mathbb{R}^+$ with $C_k^N(0) = 1$, where at each selection event of size $n \geq -1$ at a given time t the value is updated as

$$C_k^N(t) = C_k^N(t-)(1 + n/N). \tag{3.27}$$

Here $n = -1$ indicates a killing event, and $n \geq 0$ a cloning event according to the two parts of the generator (3.22). This idea comes from a branching process interpretation of the cloning ensemble related to the unnormalized measure v_t^k, since with (2.17) we have that

$$v_t^k(1) \approx e^{\lambda_k t} \quad \text{as } t \to \infty \,,$$

so λ_k corresponds to the volume expansion factor of the clone ensemble due to selection dynamics.

In our setting, the dynamics of $C_k^N(t)$ can be defined jointly with the cloning process via an extension of the generator (3.22)

$$\bar{\mathcal{L}}_k^{N,clone} F(\underline{x}, \zeta) = \bar{\mathcal{L}}_k^{N,mc} F(\underline{x}, \zeta) + \bar{\mathcal{L}}_k^{N,-} F(\underline{x}, \zeta) \,,$$

acting on functions that depend on the state $\underline{x} \in E^N$ and the cloning factor $\zeta \in \mathbb{R}^+$. With (3.21) we have for cloning events

$$\bar{\mathcal{L}}_k^{N,mc} F(\underline{x}, \zeta) := \sum_{i=1}^N \sum_{n=0}^N \frac{1}{\binom{N}{n}} \sum_{\substack{A \subseteq \{1,\ldots,N\} \\ |A|=n}} \int_E W_k(x_i, dy) p_{k,x_i}^N(n)$$

$$\left(F\left(\underline{x}^{A,x_i;i,y}, \zeta(1 + n/N)\right) - F(\underline{x}, \zeta)\right)$$

and with the third line of (3.6) for killing events

$$\bar{\mathcal{L}}_k^{N,-} F(\underline{x}, \zeta) = \sum_{i=1}^N \left(\mathcal{V}_k(x_i) - c\right)^- \frac{1}{N} \sum_{j=1}^N \left(F(\underline{x}^{i,x_j}, \zeta(1 - 1/N)) - F(\underline{x}, \zeta)\right).$$

So the joint process $\left((\underline{X}_k^N(t), C_k^N(t)) : t \geq 0\right)$ is Markov, and observing only the cloning factor with the simple test function $G(\underline{x}, \zeta) = \zeta$ we get

$$\bar{\mathcal{L}}_k^{N,mc} G(\underline{x}, \zeta) = \sum_{i=1}^N \sum_{n=0}^N \frac{1}{\binom{N}{n}} \sum_{\substack{A \subseteq \{1,\ldots,N\} \\ |A|=n}} \int_E W_k(x_i, dy) p_{k,x_i}^N(n) \zeta \cdot \frac{n}{N}$$

$$= \frac{\zeta}{N} \sum_{i=1}^N \left(\mathcal{V}_k(x_i) - c\right)^+. \tag{3.28}$$

In the last line we have used (3.20), and in a similar fashion we get for killing events

$$\bar{\mathcal{L}}_k^{N,-} G(\underline{x}, \zeta) = -\frac{\zeta}{N} \sum_{i=1}^N \left(\mathcal{V}_k(x_i) - c\right)^-. \tag{3.29}$$

Therefore

$$\bar{\mathcal{L}}_k^{N,clone} G(\underline{x}, \zeta) = \frac{\zeta}{N} \sum_{i=1}^{N} \left(\mathcal{V}_k(x_i) - c \right) = \zeta m^N(\underline{x})(\mathcal{V}_k) - \zeta c,$$

and analogously to (3.18), the expected time evolution of $C_k^N(t)$ is then given by

$$\frac{d}{dt} \mathbb{E}[C_k^N(t)] = \mathbb{E}[C_k^N(t) \cdot \mu_t^N(\mathcal{V}_k - c)].$$

This is also the evolution of $\nu_t^N(e^{-tc}) = e^{-tc} \nu_t^N(1)$, since

$$\frac{d}{dt} \mathbb{E}[\nu_t^N(e^{-tc})] = \mathbb{E}[\mu_t^N(\mathcal{V}_k) e^{-tc} \nu_t^N(1) - c\, e^{-tc} \nu_t^N(1)]$$

$$= \mathbb{E}[\nu_t^N(e^{-tc}) \cdot \mu_t^N(\mathcal{V}_k - c)].$$

With initial conditions $C_k^N(0) = 1 = \nu_t^N(1)$, a Gronwall argument analogous to (3.19) gives

$$\mathbb{E}[e^{tc} C_k^N(t)] = \mathbb{E}[\nu_t^N(1)] = \nu_t(1) \quad \text{for all } t \geq 0 \text{ and } N \geq 1.$$

So $e^{tc} C_k^N(t)$ is an unbiased estimator for $\nu_t(1)$, which leads also to an alternative estimator for $\lambda_k(t)$ (2.17) given by

$$\overline{\Lambda}_k^N(t) := \frac{1}{t} \log C_k^N(t) + c. \tag{3.30}$$

Note that this is not itself unbiased as a consequence of the nonlinear transformation involving the logarithm.

Since $C_k^N(t)$ is defined as a product (3.27), we can use another simple test function $G(\underline{x}, \zeta) = \log \zeta$ to analyze the convergence behaviour of $\overline{\Lambda}_k^N(t)$. Analogously to (3.28) and (3.29) we get

$$\bar{\mathcal{L}}_k^{N,clone} G(\underline{x}, \zeta) = \frac{1}{N} \sum_{i=1}^{N} \left(\mathcal{V}_k(x_i) - c \right) = m^N(\underline{x})(\mathcal{V}_k) - c + O\left(\frac{1}{N}\right),$$

where we have also used (3.20) and assumed that the support of p_{k,x_i}^N is bounded independently of N (which is the case for common choices in the literature such as (3.26)). This allows us to approximate $\log(1 + n/N) = n/N + O(1/N^2)$ as $N \to \infty$, leading to error terms of order $1/N$. Then, analogously to (3.12) we get with $\log C_k^N(0) = 0$ that

$$\mathcal{M}_C^N(t) := \log C_k^N(t) - \int_0^t \bar{\mathcal{L}}_k^{N,clone} G(\underline{X}_k(s), C_k^N(s))\, ds$$

$$= \log C_k^N(t) - t(\Lambda_k^N(t) - c) + t\, O\left(\frac{1}{N}\right)$$

is a martingale. For the carré du champ we obtain from a straightforward computation that

$$\Gamma_k^{N,clone} G(\underline{x}, \zeta) = \frac{1}{N^2} \sum_{i=1}^{N} |\mathcal{V}_k(x_i) - c| + O\left(\frac{1}{N^2}\right),$$

and since the potential $\mathcal{V}_k$ is bounded (2.12), the quadratic variation of the martingale is bounded by

$$\langle \mathcal{M}_C^N \rangle(t) \leq \frac{t}{N}(\bar{v}_k + |c|) + O\left(\frac{1}{N^2}\right).$$

Therefore the estimator (3.30) based on the cloning factor

$$\bar{\Lambda}_k^N(t) = \frac{1}{t} \log C_k^N(t) + c = \Lambda_k^N(t) + O\left(\frac{1}{N}\right) + \frac{1}{t} \mathcal{M}_C^N(t)$$

is asymptotically equal to the basic estimator $\Lambda_k^N(t)$ (3.2), with corrections that vanish as $1/N$ in the L^p-norm as $N \to \infty$ uniformly in $t \geq 0$, analogously to the discussion in Sect. 3.2. Therefore, the same convergence results as stated in the Theorem apply for $\bar{\Lambda}_k^N(t)$. Similar convergence results can be shown to hold for $e^{tc} C_k^N(t)$ as an estimator of $\nu_t(1)$ for fixed $t > 0$, but naturally cannot hold uniformly in time. Since the object of interest is usually the long-time limit λ_k (2.17), the practical relevance of this is limited, in addition to the general point that random errors dominate the convergence as mentioned in Sect. 3.2. In practice, the basic ergodic average $\Lambda_k^N(t)$ (3.2) is more useful than the cloning factor in the application areas we have in mind. In particular, for alternative particle approximations such as (3.7) or (3.8) where cloning and killing events are effectively combined, it is not clear how to define a cloning factor, whereas $\Lambda_k^N(t)$ is always easily accessible.

4 Efficiency and Application of Particle Approximations

4.1 Efficiency of Algorithms

Selection events (cloning or killing) in a particle approximation increase the correlations among the particles in the ensemble, and thereby decrease the resolution in the empirical distribution $\mu_t^N = \mu^N(\underline{X}_k(t))$, and ultimately the quality of the sample average in the estimator (3.2). Therefore it is desirable to minimize the total rate $S_k(\underline{x})$ of selection events for a particle approximation. For algorithm (3.6) this is given by

$$S_k^1(\underline{x}) = \sum_{i=1}^N |\mathcal{V}_k(x_i) - c|, \tag{4.1}$$

and the same holds for the cloning algorithm (3.22), since the change in cloning rate is compensated exactly by the average number of clones created to obtain the same overall rate. It is easy to see that for a given state $\underline{x}$ of the clone ensemble, there is an optimal choice of c to minimize this expression, given by the median of the fitness distributions $\mathcal{V}_k(\underline{x}) := \{\mathcal{V}_k(x_i) : i = 1, \ldots, N\}$. If the distribution of $\underline{X}_k(t)$ is unimodal with light enough tails, the median can be well approximated by the mean $\mu_t^N(\mathcal{V}_k)$. Since both quantities can be computed with similar computational effort (or well approximated at reduced cost using only a subset of the ensemble), choosing

$$c = c(t) = \text{median}\left(\mathcal{V}_k(\underline{X}_k(t))\right)$$

should be computationally optimal. In particular, the simplest choice $c = 0$ in the cloning algorithm is in general far from optimal, so is choosing $c = \inf_{x \in E} \mathcal{V}_k(x)$ to get rid of the killing part of the dynamics (see first remark in Sect. 3.3).

Intuitively, algorithms (3.7) and (3.8) should lead to even lower total selection rates since every selection event increases the fitness potential, while in algorithms based on (3.6) it increases only on average and may also decrease as the result of selection events. Indeed for (3.7) we have

$$S_k^2(\underline{x}) = \frac{1}{N} \sum_{i,j=1}^{N} \left(\mathcal{V}_k(x_i) - \mathcal{V}_k(x_j)\right)^+ = \frac{1}{2N} \sum_{i,j=1}^{N} \left|\mathcal{V}_k(x_i) - \mathcal{V}_k(x_j)\right|$$

$$\leq \frac{1}{2} \sum_{i=1}^{N} \left(\left|\mathcal{V}_k(x_i) - c\right| + \left|c - \mathcal{V}_k(x_i)\right|\right) = S_k^1(\underline{x}), \tag{4.2}$$

by symmetry of summations and the triangle inequality. The inequality is strict except for degenerate cases, e.g. if $\mathcal{V}_k(x_i)$ takes only two values, and c lies in between the two. In practice, in the scenarios which we have investigated, it turns out that unless the distribution of $\underline{X}_k(t)$ is seriously skewed, S_k^2 is strictly smaller than S_k^1 by a sizeable amount, as is illustrated later in Fig. 2 for the inclusion process. Algorithm (3.8) provides further improvement with

$$S_k^3(\underline{x}) = \frac{1}{N} \sum_{i,j=1}^{N} \left(\mathcal{V}_k(x_i) - \mu^N(\underline{x})(\mathcal{V}_k)\right)^- \frac{\left(\mathcal{V}_k(x_j) - \mu^N(\underline{x})(\mathcal{V}_k)\right)^+}{\mu^N(\underline{x})\left((\mathcal{V}_k - \mu^N(\underline{x})(\mathcal{V}_k))^+\right)}$$

$$= \frac{1}{2} \sum_{i=1}^{N} \left|\mathcal{V}_k(x_i) - \mu^N(\underline{x})(\mathcal{V}_k)\right| \leq S_k^2(\underline{x}) . \tag{4.3}$$

Here we have used $\sum_{i=1}^{N} \left(\mathcal{V}_k(x_i) - \mu^N(\underline{x})(\mathcal{V}_k)\right) = 0$ and Jensen's inequality to compare with $S_k^2(\underline{x})$, since $v \mapsto |a - v|$ is convex for all $a \in \mathbb{R}$. Note that the rate of change of the mean fitness $\mu_t^N(\mathcal{V}_k)$ is given by the same expression in all the above particle approximations,

$$\mu_t^N(\widehat{\mathcal{L}}_k \mathcal{V}_k) + \mu_t^N(\mathcal{V}_k^2) - \mu_t^N(\mathcal{V}_k)^2. \tag{4.4}$$

The first term due to mutation dynamics $\widehat{\mathcal{L}}_k$ can have either sign and is identical in all algorithms, while the second due to selection is positive and given by the empirical variance of $\mathcal{V}_k$. This follows from direct computations using the averaged test function $F(\underline{x}) = \mu^N(\underline{x})(\mathcal{V}_k)$ in (3.6), (3.7), (3.8) and (3.22), and is consistent with the evolution equation (2.24). So the mean fitness evolves until a mutation selection balance is reached and the rate of change (4.4) vanishes, characterizing the stationary state of the particle approximation process. Note that this basic mechanism is identical in all particle approximations discussed here, so we expect the mean fitness to show a very similar behaviour. While finite size effects can lead to deviations also in the mean, the main difference between the algorithms is found on the level of variances and time correlations, which can be significantly reduced using (3.7) or (3.8) as illustrated in the next subsections. Since our main observable of interest $\Lambda_k^N(t)$ is an ergodic time average of $\mu_t^N(\mathcal{V}_k)$, this can lead to significant improvements in the accuracy of the estimator (3.2).

The correlations introduced by selection are counteracted by mutation dynamics, which occur independently for each particle and decorrelate the ensemble. The dynamics of correlation structures in cloning algorithms has been discussed in some detail recently in [7,8,13,38], and can be understood in terms of ancestry in the generic population dynamics interpretation. Those results also discuss important non-ergodicity effects in the measurement of path properties and the interpretation of particle trajectories, which were already pointed out in [3] and are also a subject of recent research [39]. This poses interesting questions for rigorous mathematical investigations which are left to future work. Here we simply conclude with a numerical test in the next subsections, which supports the intuition that approximation (3.7) with minimal selection rates leads to variance reduction in the relevant estimators compared to the cloning algorithm. Since the selection rate in (3.7) depends on potential differences between pairs, implementation is in general more involved than for algorithms based on

(3.6). While the scaling $t N \log N$ of computational complexity with the size N of the clone ensemble is the same, the prefactor and computational cost in practice may be higher and this has to be traded off against gains in accuracy on a case by case basis. For the examples studied below we find a computationally efficient implementation of (3.7) providing a clear improvement over the standard cloning algorithm, which is the main contribution of this paper in this context. Algorithm (3.8), on the other hand, provides only marginal improvement over (3.7), but cannot be implemented as efficiently in our area of interest.

4.2 Current Large Deviations for Lattice Gases

In the following we consider one-dimensional stochastic lattice gases with periodic boundary conditions on the discrete torus $\mathbb{T}_L$ with L sites and a fixed number of particles M. Within our general framework, they are simply Markov chains on the finite state space E of all particle configurations, which have been of recent research interest in the context of current fluctuations. We denote configurations by $\eta = (\eta_x : x \in \mathbb{T}_L)$ where $\eta_x \in \mathbb{N}_0$ is interpreted as the mass (or number of monomers) at site x, and the process is denoted as $(\eta(t) : t \geq 0)$. In order to use standard notation for lattice gases, in this and the following subsection we change notation, and in particular the use of $x, y \in \mathbb{T}_L$ is different to the use of those symbols in previous sections where they denoted states in E. Monomers jump to nearest neighbour sites with rates $u(\eta_x, \eta_y) \geq 0$ for $y = x \pm 1$ depending on the occupation numbers of departure and target site, multiplied with a spatial bias $p = 1 - q \in [0, 1]$. The generator is of the form

$$
\mathcal{L} f(\eta) = \sum_{x \in \mathbb{T}_L} \Big[p \, u(\eta_x, \eta_{x+1}) \big(f(\sigma_{x,x+1}\eta) - f(\eta) \big)
$$
$$
+ q \, u(\eta_x, \eta_{x-1}) \big(f(\sigma_{x,x-1}\eta) - f(\eta) \big) \Big],
\tag{4.5}
$$

where $\sigma_{x,y}\eta$ results from the configuration η after moving one particle from x to y. The number of particles $M = \sum_{x \in \mathbb{T}_L} \eta_x$ is a conserved quantity, but otherwise we assume the process to be irreducible for any fixed M, which is ensured for example by positivity of the rates, i.e. for all $k, l \geq 0$

$$
u(k, l) = 0 \quad \Leftrightarrow \quad k = 0.
$$

This class includes various models that have been studied in the literature, for example the inclusion process introduced in [19], where

$$
u(k, l) = k(d + l) \quad \text{for all } k, l \geq 0,
\tag{4.6}
$$

with a positive parameter $d > 0$. Particles perform independent jumps with rate d and in addition are attracted by each particle on the target site with rate 1, giving rise to the 'inclusion' interaction. This model has attracted recent attention due the presence of condensation phenomena [40,41] and in the context of large deviations of the particle current [20], and we will use this as an example in Sect. 4.3. Other well-studied models covered by our set-up are the exclusion process with state space $E \subset \{0, 1\}^{\mathbb{T}_L}$ and $u(\eta_x, \eta_y) = \eta_x(1 - \eta_y)$, or zero-range processes with $E \subset \mathbb{N}_0^{\mathbb{T}_L}$ and rates $u(\eta_x, \eta_y) = u(\eta_x)$ depending only on the occupation number on the departure site.

In terms of previous notation, the jump rates for a lattice gas of type (4.5) between any two configurations η and ζ are given as

$$
W(\eta, \zeta) = \sum_{x \in \mathbb{T}_L} \Big(p \, u(\eta_x, \eta_{x+1}) \delta_{\zeta, \sigma_{x,x+1}\eta} + q \, u(\eta_x, \eta_{x-1}) \delta_{\zeta, \sigma_{x,x-1}\eta} \Big).
\tag{4.7}
$$

In the following we focus on lattice gases where $\sum_x u(\eta_x, \eta_{x+1}) = \sum_x u(\eta_x, \eta_{x-1})$ for all configurations η. While this is not true in general for models of type (4.5), it holds for many examples including inclusion, exclusion and zero-range processes mentioned above. With $p + q = 1$, the total exit rate out of configuration η is then simply given by

$$w(\eta) = \sum_{x \in \mathbb{T}_L} \left(p\, u(\eta_x, \eta_{x+1}) + q\, u(\eta_x, \eta_{x-1}) \right) = \sum_{x \in \mathbb{T}_L} u(\eta_x, \eta_{x+1}). \tag{4.8}$$

We are interested in an observable A_T measuring the total particle current up to time T, which is achieved by choosing $h(\eta) \equiv 0$ in (2.5) and

$$g(\eta, \zeta) = \pm 1 \quad \text{if } \zeta = \sigma_{x,x\pm1}\eta \quad \text{and} \quad g(\eta, \zeta) = 0 \quad \text{otherwise} .$$

Using (4.8) we see by direct computation that the potential (2.11) takes the simple form

$$\mathcal{V}_k(\eta) = (Q_k - 1)w(\eta) \quad \text{where} \quad Q_k := pe^k + qe^{-k}. \tag{4.9}$$

Modified mutation rates $W_k(\eta, \zeta)$ are given by (4.7) replacing (p, q) by (pe^k, qe^{-k}), leading to modified total exit rates

$$w_k(\eta) = Q_k \sum_{x \in \mathbb{T}_L} u(\eta_x, \eta_{x+1}) = Q_k w(\eta). \tag{4.10}$$

The similarity of $\mathcal{V}_k$ and w_k for lattice gases (4.5) that obey (4.8) provides a direct relation between mutation and selection rates, and allows us to set up an efficient rejection based implementation of a particle approximation $(\underline{\eta}_k(t) : t \geq 0)$ based on the efficient algorithm (3.7). In the following we omit the subscript k for configurations and write $\underline{\eta}(t) = (\eta^i(t), i = 1, \ldots, N)$ to simplify notation. For given parameters $p, q = 1 - p$ and fixed $k \in \mathbb{R}$ we distinguish two cases.

$\mathbf{Q_k < 1}$. We sample the ensemble of N clones at a total rate of $\mathcal{W}(\underline{\eta}) := \sum_{i=1}^{N} w(\eta^i)$, and pick a clone i with probability $w(\eta^i)/\mathcal{W}(\underline{\eta})$ for the next event. With probability $Q_k \in (0, 1)$ this is a simple mutation within clone i, and then we replace η^i by ζ^i with probability $W_k(\eta^i, \zeta^i)/w_k(\eta^i)$. Otherwise, with probability $1 - Q_k$ we perform a selection event following the second line in (3.7): Pick a clone j uniformly at random (including i). If

$$\mathcal{V}_k(\eta^j) > \mathcal{V}_k(\eta^i) \quad \text{or equivalently} \quad w(\eta^j) < w(\eta^i)$$

(with (4.9) and since $Q_k < 1$), replace η^i by η^j with probability $\left(w(\eta^i) - w(\eta^j) \right)/w(\eta^i)$. This procedure ensures that mutation and selection events are sampled with the correct rates as required in (3.7).

$\mathbf{Q_k > 1}$. We sample the ensemble of N clones at a total rate of $Q_k\mathcal{W}(\underline{\eta})$, and pick a clone i with probability $w(\eta^i)/\mathcal{W}(\underline{\eta})$ and a clone j uniformly at random. If $w(\eta^j) < w(\eta^i)$ we replace η^j by η^i with probability $\frac{Q_k-1}{Q_k} \frac{w(\eta^i)-w(\eta^j)}{w(\eta^i)}$. Then we mutate clone i as above, combining the mutation and selection event as in the cloning algorithm.

Remarks Note that $Q_k = 1$ is equivalent to $k = 0$, which corresponds to the original process with $\lambda_0 = 0$ and does not require any estimation. For $Q_k > 1$ we perform mutation and selection events simultaneously, in analogy to the cloning procedure explained in Sect. 3.3, but can use the efficient algorithm (3.7). For $Q_k < 1$ no mutation or selection event occurs with probability $(1 - Q_k)\frac{w(\eta^j)}{w(\eta^i)}\mathbb{1}(w(\eta^j) < w(\eta^i))$, and a high rate of such rejections is not desirable for computational efficiency. But even for very small values of Q_k the second factor is usually significantly smaller than 1 (or simply 0), since clone i was picked with probability proportional to $w(\eta^i)$ and j uniformly at random.
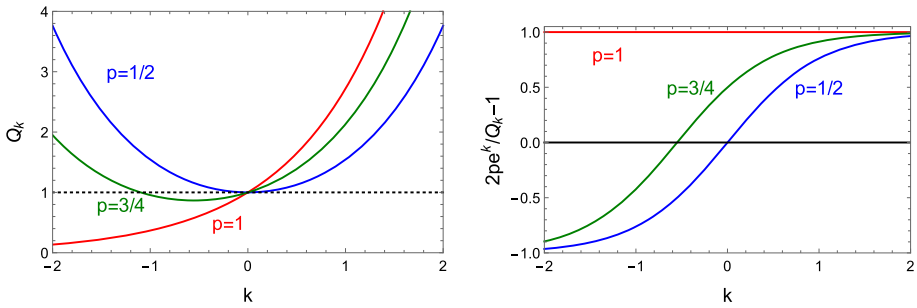
Fig. 1 Illustration of Q_k (left) as given in (4.9) and the drift $2pe^k/Q_k - 1$ for the modified dynamics (right) as a function of k for different values of the asymmetry $p = 1 - q$. The minimum of Q_k is $2\sqrt{pq}$, attained at $k = \frac{1}{2}\log\frac{q}{p} \in [-\infty, \infty]$, which is also where the modified drift vanishes

Note also that if the cloning algorithm (3.22) is implemented with the common choice $c = 0$ for a lattice gas of the type discussed here, due to (4.9) and (4.10) the average number of clones per event (3.20) is

$$m_k^N(\eta^i) = \left(\mathcal{V}_k(\eta^i)\right)^+/w_k(\eta^i) = \frac{Q_k - 1}{Q_k} \in (0, 1) \quad \text{if } Q_k > 1,$$

and 0 for $Q_k < 1$, where only killing occurs. In particular, this is independent of the state $\underline{\eta}$ of the clone ensemble, and the standard distribution of the form (3.26) is a simple Bernoulli random variable.

While with (4.10) the total mutation rate is $Q_k\mathcal{W}(\underline{\eta})$, selection rates (4.1), (4.2) and (4.3) can be written as

$$S_k^1(\underline{\eta}) = \sum_{i=1}^{N} \left|(Q_k - 1)w(\eta^i) - c\right| \stackrel{c=0}{=} |Q_k - 1|\mathcal{W}(\underline{\eta})$$

$$S_k^2(\underline{\eta}) = |Q_k - 1|\frac{1}{2N}\sum_{i,j=1}^{N}\left|w(\eta^i) - w(\eta^j)\right|$$

$$S_k^3(\underline{\eta}) = |Q_k - 1|\frac{1}{2}\sum_{i=1}^{N}\left|w(\eta^i) - \mu^N(\underline{\eta})(w)\right|. \tag{4.11}$$

So for very small values of Q_k close to 0 the mutation rate can become very small in comparison to selection, which means that significant computation time is devoted to re-weighting by selection, rather than advancing the dynamics via mutation events. This effect is typically much stronger for the standard cloning algorithm with $c = 0$, and occurs for example for totally asymmetric lattice gases with $p = 1$ and negative k conditioning on low currents. In Fig. 1 we include a sketch of Q_k for different values of asymmetry, including also the drift of the modified dynamics, which can be reversed in partially asymmetric systems.

In Fig. 2 we compare the cloning algorithm to algorithms (3.7) and (3.8) for an inclusion process with $d = 1$, $L = 64$, $M = 128$ and asymmetry $p = 0.7$. It is known [20] that the SCGF λ_k scales linearly with the system size L, and outside the convergent regime $k \in [-\ln(\frac{1-p}{p}), 0] \approx [-0.85, 0]$ the rescaled SCGF λ_k/L diverges as $L \to \infty$ (divergent regime). We compare estimates $\Lambda_k^N(t)$ for the cloning algorithm (3.22) with $c = 0$ and algorithm (3.7) in the convergent regime. We use initial conditions where M particles are
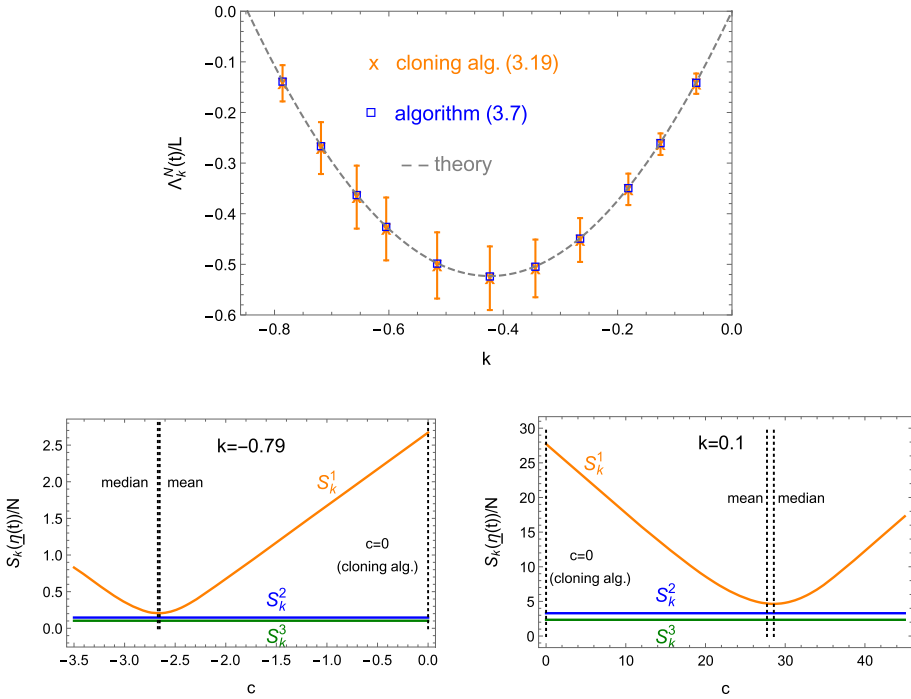
Fig. 2 Inclusion process (4.6) with $d = 1$, system size $L = 64$, $M = 128$ particles, asymmetry $p = 0.7$ and $N = 2^{11}$ clones at time $t = 42000$. (Top) The rescaled estimator $\Lambda_k^N(t)/L$ as a function of k in the convergent regime, comparing the cloning algorithm (3.22) with $c = 0$ (orange) and algorithm (3.7) (blue). Error bars indicate 5 standard deviations, which are bounded by the size of the symbols for (3.7). (Bottom) Illustration of the relationship between S_k^1 depending on c, and S_k^2 and S_k^3 (4.11) for $k = -0.79$ (left) and $k = 0.1$ (right) based on the state $\underline{\eta}(t)$ of the clone ensemble

distributed on L lattice sites uniformly at random, and a burn-in time of $10 \cdot L = 640$ as discussed in (2.19) and (2.20). This leads to an obvious adaption of the integration interval in the estimator $\Lambda_k^N(t)$ (3.2), but we do not alter the notation here to keep it simple. Both algorithms perform very well and agree with a simple theoretical estimate based on bias reversal, which is not the main concern in this paper and we refer the reader to [20]. Enlarged error bars indicating 5 standard deviations reveal that (3.7) is significantly more accurate than (3.22). This is due to lower total selection rates S_k illustrated at the bottom in the converging and diverging regime. While S_k^2 for (3.7) is much lower than S_k^1 with $c = 0$, S_k^3 does not offer significant further improvement. Since the efficient rejection based implementation of (3.7) explained above does not work for (3.8), we focus on (3.7) in our context. The much higher selection rate for the cloning algorithm with $c = 0$ leads to a significantly higher time variation of the average potential in the convergent regime compared to algorithm (3.7), as is illustrated in Fig. 3. So in comparison to standard cloning, algorithm (3.7) leads to reduced finite size effects and/or a significant variance reduction in this example, and a significant improvement of convergence of the estimator (3.2). We have checked that this also holds for zero-range processes with bounded rates. These promising first numerical results pose interesting questions for a systematic study of practical properties of the algorithms and associated time correlations for future work, also in comparison with various recent results on improvements of cloning algorithms [7,11,12].
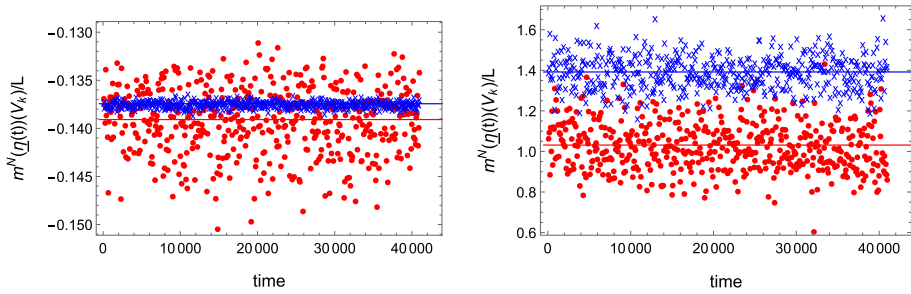
Fig. 3 Inclusion process (4.6) with $d = 1$, system size $L = 64$, $M = 128$ particles, asymmetry $p = 0.7$ and $N = 2^{11}$ clones. Time series of the mean fitness $m^N(\underline{\eta}(t))(\mathcal{V}_k)/L$ for the cloning algorithm (red dots) and algorithm (3.7) (blue crosses), with time averages indicated by full lines. (Left) In the convergent regime for $k = -0.79$ we see a clear variance reduction using (3.7) with similar time average. (Right) In the divergent regime for $k = 0.1$ we have similar variance but (3.7) improves on the time average

4.3 Details for the Inclusion Process

We summarize the procedure outlined in the previous subsection for the inclusion process with rates (4.6) on the torus $\mathbb{T}_L$ with M particles in pseudo-code given below. Besides fixing the model parameter $d > 0$ and the tilt $k \in \mathbb{R}$, the specific parameters for the estimator are the ensemble size N and the total simulation time t, which lead to the estimator $\Lambda_k^N(t)$ as given in (3.2). For simplicity we do not include any burn-in times in this description, which would obviously be used in practice. In this implementation we make a further simplification which is very common for continuous-time jump dynamics of large systems: We replace exponentially distributed random time increments by their expectation, given by $\Delta t = 1/\mathcal{W}(\underline{\eta})$ for $Q_k < 1$ and $\Delta t = 1/(Q_k \mathcal{W}(\underline{\eta}))$ for $Q_k > 1$. Since with (4.9)

$$\frac{1}{N}\sum_{i=1}^{N}\mathcal{V}_k(\eta^i(s)) = \frac{Q_k - 1}{N}\mathcal{W}(\underline{\eta}) ,$$

we get for increments in the evaluation of the ergodic time integral in (3.2)

$$\Delta t \frac{1}{N}\sum_{i=1}^{N}\mathcal{V}_k(\eta^i(s)) = \begin{cases} \frac{Q_k - 1}{N} , & Q_k < 1 \\ \frac{Q_k - 1}{Q_k N} , & Q_k > 1 \end{cases} .$$

These are independent of the actual state $\underline{\eta}$ of the clones, so evaluation of $\Lambda_k^N(t)$ in (3.2) can be achieved by a simple integer counter $\hat{\Lambda}_k^N$ as explained in the pseudocode Algorithm 1 and 2. While this counter may appear similar to the cloning factor explained in Sect. 3.4 at first glance, we want to stress that here finer increments of $+1$ are added after every event (not only selections).

5 Discussion

We have presented an analytical approach to cloning algorithms based on McKean interpretations of Feynman–Kac semigroups that have been introduced in the applied probability literature. This allows us to establish rigorous error bounds for the cloning algorithm in

---

Algorithm 1: IP (4.5) with rates (4.6) and $Q_k < 1$ (4.9)

Parameters N number of clones; t simulation time;
Initialize configurations η^i, $i = 1, \ldots, N$ with mass M each;
$w(\eta^i) = d\,M + \sum_x \eta_x^i \eta_{x+1}^i$; $\mathcal{W}(\underline{\eta}) = \sum_j w(\eta^j)$; $s = \frac{1}{\mathcal{W}(\underline{\eta})}$; $\hat{\Lambda}_k^N = 1$;

while $s < t$ **do**

     pick clone i with probability $w(\eta^i)/\mathcal{W}(\underline{\eta})$;

     if $R \sim U(0, 1) < Q_k$ **then**

         $\eta^i \leftarrow \zeta$, ζ chosen with probability $\frac{W_k(\eta^i, \zeta)}{Q_k w(\eta^i)}$ (mutation);

     else

         pick clone j uniformly at random;

         if $w(\eta^j) < w(\eta^i)$ **then**

             $\eta^i \leftarrow \eta^j$ with probability $\frac{w(\eta^i) - w(\eta^j)}{w(\eta^i)}$ (selection);

         end

     end

     $s \leftarrow s + \frac{1}{\mathcal{W}(\underline{\eta})}$; $\hat{\Lambda}_k^N \leftarrow \hat{\Lambda}_k^N + 1$;

end

Output $\Lambda_k^N(t) = \frac{Q_k - 1}{Nt} \hat{\Lambda}_k^N$;

Algorithm 2: IP (4.5) with rates (4.6) and $Q_k > 1$ (4.9)

Parameters N number of clones; t simulation time;
Initialize configurations η^i, $i = 1, \ldots, N$ with mass M each;
$w(\eta^i) = d\,M + \sum_x \eta_x^i \eta_{x+1}^i$; $\mathcal{W}(\underline{\eta}) = \sum_j w(\eta^j)$; $s = \frac{1}{Q_k \mathcal{W}}$; $\hat{\Lambda}_k^N = 1$;

while $s < t$ **do**

     pick clone i with probability $w(\eta^i)/\mathcal{W}(\underline{\eta})$;

     pick clone j uniformly at random;

     if $w(\eta^j) < w(\eta^i)$ **then**

         $\eta^j \leftarrow \eta^i$ with probability $\frac{Q_k - 1}{Q_k} \frac{w(\eta^i) - w(\eta^j)}{w(\eta^i)}$ (selection);

     end

     $\eta^i \leftarrow \zeta$, ζ chosen with probability $\frac{W_k(\eta^i, \zeta)}{Q_k w(\eta^i)}$ (mutation);

     $s \leftarrow s + \frac{1}{Q_k \mathcal{W}(\underline{\eta})}$; $\hat{\Lambda}_k^N \leftarrow \hat{\Lambda}_k^N + 1$;

end

Output $\Lambda_k^N(t) = \frac{Q_k - 1}{Q_k Nt} \hat{\Lambda}_k^N$;

---

continuous time, and to suggest a more efficient variant of the algorithm which can be implemented effectively for current large deviations in stochastic lattice gases. The latter is based on minimizing the selection rate in a standard population dynamics interpretation of particle approximations of non-linear processes. We include a first application of this idea in the context of inclusion processes, but its full potential will be explored in future more systematic studies of optimization of cloning-type algorithms. The rigorous results fully reported in [18] apply under very general conditions, demanding bounded jump rates and existence of a spectral gap for the underlying jump process. These impose no restriction for lattice gases with a fixed number of particles, which are essentially finite state Markov chains. We anticipate that these techniques can also be applied for more general processes including

diffusive, piecewise deterministic, or possibly non-Markovian dynamics (see [42] for first heuristic results in this direction). Another interesting direction would be a rigorous analysis of the detailed ergodic properties of trajectories in the clone ensemble based on recent results in [7,8,39].

References

1. Anderson, J.B.: A random-walk simulation of the Schrödinger equation: H_3^+. J. Chem. Phys. **63**, 1499 (1975)
2. Grassberger, P.: Go with the winners: a general Monte Carlo strategy. Comput. Phys. Commun. **147**(1), 64–70 (2002)
3. Giardina, C., Kurchan, J., Peliti, L.: Direct evaluation of large-deviation functions. Phys. Rev. Lett. **96**(12), 120603 (2006)
4. Lecomte, V., Tailleur, J.: A numerical approach to large deviations in continuous time. J. Stat. Mech.: Theory Exp. **2007**(03), P03004 (2007)
5. Giardina, C., Kurchan, J., Lecomte, V., Tailleur, J.: Simulating rare events in dynamical processes. J. Stat. Phys. **145**(4), 787–811 (2011)
6. Jack, R.L., Sollich, P.: Large deviations and ensembles of trajectories in stochastic models. Prog. Theor. Phys. Suppl. **184**, 304–317 (2010)
7. Nemoto, T., Bouchet, F., Jack, R.L., Lecomte, V.: Population-dynamics method with a multicanonical feedback control. Phys. Rev. E **93**, 062123 (2016)
8. Hidalgo, E. G.: Cloning algorithms: from large deviations to population dynamics. Ph.D. thesis, Université Sorbonne Paris Cité — Université Paris Diderot 7, (2018)
9. Nemoto, T., Guevara Hidalgo, E., Lecomte, V.: Finite-time and finite-size scalings in the evaluation of large-deviation functions: analytical study using a birth-death process. Phys. Rev. E **95**, 012102 (2017)
10. Guevara Hidalgo, E., Nemoto, T., Lecomte, V.: Finite-time and finite-size scalings in the evaluation of large-deviation functions: Numerical approach in continuous time. Phys. Rev. E **95**, 062134 (2017)
11. Ferré, G., Touchette, H.: Adaptive sampling of large deviations. J. Stat. Phys. **172**(6), 1525–1544 (2018)
12. Brewer, T., Clark, S.R., Bradford, R., Jack, R.L.: Efficient characterisation of large deviations using population dynamics. J. Stat. Mech. **2018**(5), 053204 (2018)
13. Pérez-Espigares, C., Hurtado, P. I.: Sampling rare events across dynamical phase transitions. arXiv:1902.01276
14. Del Moral, P., Miclo, L.: A Moran particle system approximation of Feynman-Kac formulae. Stoch. Process. Appl. **86**(2), 193–216 (2000)
15. Del Moral, P., Pierre, M.: Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In: Azema, J., Emery, M., Ledoux, M., Yor, M. (eds.) Seminaire de probabilites, XXXIV, pp. 1–145. Springer, Berlin (2000)
16. Del Moral, P.: Feynman-Kac Formulae. Springer, New York (2004)
17. Rousset, M.: On the control of an interacting particle estimation of Schrödinger ground states. SIAM J. Math. Anal. **38**(3), 824–844 (2006)
18. Angeli, L., Grosskinsky, S., Johansen, A. M.: Limit theorems for cloning algorithms. under review (arxiv:1902.00509)
19. Giardinà, C., Kurchan, J., Redig, F., Vafayi, K.: Duality and hidden symmetries in interacting particle systems. J. Stat. Phys. **135**(1), 25–55 (2009)
20. Chleboun, P., Grosskinsky, S., Pizzoferrato, A.: Current large deviations for partially asymmetric particle systems on a ring. J. Phys. A **51**(40), 405001 (2018)

21. Lazarescu, A.: The physicist's companion to current fluctuations: one-dimensional bulk-driven lattice gases. J. Phys. A **48**(50), 503001 (2015)
22. Hurtado, P.I., Espigares, C.P., del Pozo, J.J., Garrido, P.L.: Thermodynamics of currents in nonequilibrium diffusive systems: theory and simulation. J. Stat. Phys. **154**(1), 214–264 (2014)
23. Chleboun, P., Grosskinsky, S., Pizzoferrato, A.: Lower current large deviations for zero-range processes on a ring. J. Stat. Phys. **167**(1), 64–89 (2017)
24. Chetrite, Rl, Touchette, H.: Nonequilibrium Markov processes conditioned on large deviations. Annales de L'Institut Henri Poincaré **16**, 2005–2057 (2015)
25. Nyawo, P.T., Touchette, H.: Large deviations of the current for driven periodic diffusions. Phys. Rev. E **94**, 032101 (2016)
26. Harris, R.J., Schütz, G.M.: Fluctuation theorems for stochastic dynamics. J. Stat. Mech. **2007**(07), P07020 (2007)
27. Den Hollander, F.: Large deviations, volume 14 of Graduate Texts in Mathematics. American Mathematical Society, Providence (2008)
28. Dembo, A., Zeitouni, O.: Large deviations techniques and applications, vol. 38. Springer, New York (2009)
29. Bertini, L., Faggionato, A., Gabrielli, D.: Large deviations of the empirical flow for continuous time Markov chains. Annales de L'Institut Henri Poincaré **51**(3), 867–900 (2015)
30. Del Moral, P., Miclo, L.: Particle approximations of Lyapunov exponents connected to Schrödinger operators and Feynman-Kac semigroups. ESAIM: Prob Stat **7**, 171–208 (2003)
31. Del Moral, P.: Mean field simulation for Monte Carlo integration. CRC Press, Boca Raton (2013)
32. Baker, J. E.: Adaptive selection methods for genetic algorithms. In: Proceedings of an International Conference on Genetic Algorithms and their applications, pp. 101–111 (1985)
33. Kong, A., Liu, J.S., Wong, W.H.: Sequential imputations and Bayesian missing data problems. J. Am. Stat. Assoc. **89**(425), 278–288 (1994)
34. Liggett, T.M.: Continuous time Markov processes: an introduction, volume 113 of Graduate Texts in Mathematics. American Mathematical Society, Providence (2010)
35. Grosskinsky, S., Jatuviriyapornchai, W.: Derivation of mean-field equations for stochastic particle systems. Stoch. Process. Appl. **129**(4), 1455–1475 (2019)
36. Pra, P. D.: Stochastic mean-field dynamics and applications to life sciences (2017). http://www.cirm-math.fr/ProgWeebly/Renc1555/CoursDaiPra.pdf
37. Chow, Y.S., Teicher, H.: Probability Theory—Independence, Interchangeability, Martingales, 3rd edn. Springer, New Yok (1998)
38. Garrahan, J.P., Jack, R.L., Lecomte, V., Pitard, E., van Duijvendijk, Kristina, van Wijland, F.: First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories. J. Phys. A **42**(7), 075007 (2009)
39. Ray, U., Chan, G.K.-L., Limmer, D.T.: Importance sampling large deviations in nonequilibrium steady states. i. J Chem Phys **148**(12), 124120 (2018)
40. Grosskinsky, S., Redig, F., Vafayi, K.: Dynamics of condensation in the symmetric inclusion process. Electron. J. Prob. **18**(66), 1–23 (2013)
41. Bianchi, A., Dommers, S., Giardinà, C.: Metastability in the reversible inclusion process. Electron. J. Prob. **22**(70), 1–34 (2017)
42. Cavallaro, M., Harris, R.J.: A framework for the direct evaluation of large deviations in non-markovian processes. J. Phys. A **49**(47), 47LT02 (2016)