Bulletin of Mathematical Biology
https://doi.org/10.1007/s11538-019-00576-2

SPECIAL ISSUE: GILLESPIE AND HIS ALGORITHMS

Exact Variance-Reduced Simulation of Lattice Continuous-Time Markov Chains with Applications in Reaction Networks

P. A. Maginnis[1] · **M. West**[1] · **G. E. Dullerud**[1]

Received: 29 January 2018 / Accepted: 4 February 2019

Abstract

We propose an algorithm to reduce the variance of Monte Carlo simulation for the class of countable-state, continuous-time Markov chains, or lattice CTMCs. This broad class of systems includes all processes that can be represented using a random-time-change representation, in particular reaction networks. Numerical studies demonstrate order-of-magnitude reduction in MSE for Monte Carlo mean estimates using our approach for both linear and nonlinear systems. The algorithm works by simulating pairs of negatively correlated, identically distributed sample trajectories of the stochastic process and using them to produce variance-reduced, unbiased Monte Carlo estimates, effectively generalizing the method of antithetic variates into the domain of stochastic processes. We define a method to simulate anticorrelated, unit-rate Poisson process paths. We then show how these antithetic Poisson process pairs can be used as the input for random time-change representations of any lattice CTMC, in order to produce anticorrelated trajectories of the desired process. We present three numerical parameter studies. The first examines the algorithm's performance for the unit-rate Poisson process, and the next two demonstrate the effectiveness of the algorithm in simulating reaction network systems: a gene expression system with affine rate functions and an aerosol particle coagulation system with nonlinear rates. We also prove exact, analytical expressions for the time-resolved and integrated covariance between our antithetic Poisson processes for one technique.

✉ P. A. Maginnis
maginni1@illinois.edu

M. West
mwest@illinois.edu

G. E. Dullerud
dullerud@illinois.edu

[1] Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, 1206 W. Green St, Urbana, IL 61801, USA

Published online: 13 February 2019

Keywords Stochastic simulation · Variance reduction · Random time-change · Antithetic sampling · Monte Carlo · Reaction networks

1 Introduction

The aim of this work is to reduce the computational cost of Monte Carlo simulation of discrete-space, continuous-time Markov chains (lattice CTMCs). Such systems are significant in the stochastic simulation literature, with applications including aerosol modeling (Riemer et al. 2009) and HIV infection (Banks et al. 2012). They are particularly useful when the number of particles of a population is small and are thus poorly approximated by large-concentration ODE limits. While a few ad hoc techniques were devised to simulate such systems, the first generalized algorithm was proposed by Gillespie (1976) with the stochastic simulation algorithm (SSA). In the interim, both as computational resources have grown exponentially and the inherent stochasticity of many systems has become better understood (McAdams and Arkin 1997), such methods have seen increased utility and development. A more comprehensive survey of the major techniques used for the stochastic simulation of such systems was given by Gillespie et al. (2013). As increasingly complex models are developed, the cost of Monte Carlo simulation for their study can become prohibitive. To address this issue, we seek an algorithm to reduce the variance of unbiased Monte Carlo estimates, increasing their accuracy for a fixed or reduced number of sample trajectories.

We achieve this goal by defining an algorithm that simulates stochastic trajectories of such systems that are both exact and identically distributed yet are also mutually negatively correlated. We observe order-of-magnitude reduction in the mean-square error (MSE) of mean estimators constructed using these sample paths in numerical experiments. The key idea of the algorithm is that lattice CTMCs can be expressed in terms of random-time-changes of unit-rate Poisson processes. We have defined an algorithm for simulating negatively correlated pairs of Poisson processes while preserving their marginal distributions. These anticorrelated Poisson process pairs can then be used as random input into random time-change representations of lattice CTMCs to produce negatively correlated sample paths of the lattice CTMC itself. These anticorrelated sample paths can then be used to construct unbiased, reduced-variance Monte Carlo mean estimates of the lattice CTMC distribution. This is effectively a version of the classical method of antithetic variates (Robert and Casella 2004), extended to the realm of continuous-time stochastic processes. In the paper, we define a simple algorithm to simulate anticorrelated Poisson processes, which we refer to as the endpoint method, and use it to motivate a more general algorithm, the binomial midpoint method, that admits the endpoint method as a special case.

The random-time-change (RTC or Kurtz) representation (Ethier and Kurtz 1986) of a lattice CTMC expresses the process as a linear combination of unit-rate Poisson processes, each run at different time-rates determined by the rate functions and current state of the process. Consider a state vector $X(t) \in \mathbb{Z}^D, t \in [0, T]$. If the process has I event channels, each with propensity function $a^i(t, x)$, its RTC representation is given by

$$X(t) = x_0 + \sum_{i=1}^{I} Y^i \left(\int_0^t a^i(s, X(s))\, \mathrm{d}s \right) \nu^i. \tag{1}$$

Here, $\{Y^i\}_{i=1}^I$ are independent, unit-rate Poisson processes and $\nu^i \in \mathbb{Z}^D, i = 1, \ldots, I$, are the state jump vectors. That is, $\nu^i = X(t^+) - X(t^-)$ if the ith event channel experiences a transition at time t. This formulation is increasingly well known in the systems biology literature (Wilkinson 2011). The RTC representation can be extended to represent any discrete-state, continuous-time Markov jump process (Anderson and Kurtz 2015). Note that the only random input to this equation are the iid, unit-rate Poisson processes used to determine the time of system jumps. We will show how to simulate pairs of such unit-rate Poisson processes that are significantly negatively correlated. When these antithetic Poisson processes are input into a random-time-change representation, we demonstrate that the resulting lattice CTMC pair has significant negative correlation, even in nonlinear examples.

In previous work (Maginnis et al. 2016), we produced anticorrelated simulation algorithms to reduce the cost of Monte Carlo simulation for discrete-time Markov chains that can approximate lattice CTMCs, such as those arising from the tau-leaping approximation of Gillespie (2001). Such approximations can reduce computational time, but also introduce bias into the simulated distributions. The algorithm contained here instead produces exact realizations of the lattice CTMC distribution, and Monte Carlo mean estimators constructed using these sample paths are unbiased.

Unlike other simulation approaches for these systems that use exponential random variables to simulate jump times, such as next-reaction methods (Gibson and Bruck 2000; Anderson 2007) or Gillespie's SSA (Gillespie 1975, 1976), our approach simulates candidate jump times using unit-rate Poisson process path segments that first sample the (Poisson distributed) end-state of the process segment, then simulate the conditional (iid, uniform) jump times that occurred during the segment. Our algorithm requires a small amount of additional random variable simulation overhead. The simplest case, for example, simulates one additional random variable per Poisson process segment of length τ_s, as well as additional random variates that may be simulated during the final step but not used in the final trajectory. In most practical situations, the number of random variates simulated will lie somewhere below SSA but above next-reaction. The length τ_s can be increased or decreased for reaction channels that evolve faster or slower, respectively. These tradeoffs are discussed in more detail in Sect. 3.

Our focus is on typical system behavior, and we invoke no measure changes as in well-known variance reduction techniques like importance sampling (Glynn and Iglehart 1989) or restarting (Villen-Altamirano 2012). In this work, we do not address systems that include time delays (Bratsun et al. 2005; Cai 2007; Anderson 2007). Further, we do not directly address issues arising from time-scale separation between reaction channels, which is studied, for example, in (Cao et al. 2005). In the context of sensitivity estimation, variance reduction approaches for stochastic simulation include common reaction path methods (Rathinam et al. 2010) and finite-differencing (Anderson 2012). Multi-level Monte Carlo methods (Giles 2008) have been extended to and studied for this class of processes (Anderson and Higham 2012). We believe that sev-

eral of these techniques are in some sense "orthogonal" to our methods, and would be compatible for use in combination with our techniques, though we leave that work for future exploration.

This paper is organized as follows. In Sect. 2, we define an initial, simple algorithm to simulate antithetic pairs of unit-rate Poisson processes we refer to as the endpoint method. We then present the generalized algorithm, which we call the binomial midpoint method, that further enhances the negative correlation between paths and subsumes the endpoint method as a special case. In Sect. 2.3, we prove several analytical results that characterize the behavior and correlation of such sample paths. Next, in Sect. 2.4, we provide supporting numerical studies that both confirm the analytical results and examine Monte Carlo MSE behavior of estimators constructed using the paths over a range of parameter values. Finally, in Sect. 3, we show how these anticorrelated Poisson processes can be used as inputs to simulation of lattice CTMCs, in order to produce negatively correlated trajectory pairs of any system in this class. We present two example stochastic systems, a simple, linear gene expression model and a nonlinear aerosol coagulation model and numerically demonstrate the error vs cost relationship of the algorithms.

2 Anticorrelated Unit-Rate Poisson Processes

We begin with the problem of simulating antithetic pairs of unit-rate Poisson processes. We will proceed by introducing two algorithms to achieve this goal, with the first motivating the more general second. We provide some theoretical and numerical analysis of their behavior and performance. In Sect. 3, we will apply these algorithms to the problem of simulating lattice CTMCs.

2.1 Endpoint Method for Simulating Antithetic Poisson Processes

The inspiration for the algorithm we have created to simulate antithetic Poisson process paths is the classical technique to sample antithetic pairs of scalar random variables. This generates two Poisson-distributed random samples (X_1, X_2) with

$$X_1 := F_\tau^{-1}(U)$$
$$X_2 := F_\tau^{-1}(1-U),$$

where $U \sim \text{Unif}[0,1]$ is a uniform random variable and F_τ^{-1} is the formal inverse of the Poisson CDF with parameter τ. It is easy to show that $\text{Cov}(X_1, X_2) \leq 0$ for all τ (Maginnis et al. 2016). We note here that the focus of this paper is on reducing the variance of mean estimates. This choice is intimately linked with our choice of one-to-one operator on the unit interval, $g(u) = 1-u$, which is optimal for producing negative correlation between two variates (Whitt 1976). Indeed, for estimators of other statistics, say, for example, $\mathbb{E}[X^2]$, this choice of operator may actually be variance *increasing*. In that case, an alternate sampling operator may be desirable, such as

$$g(u) = \begin{cases} 1/2 - u & 0 \le u \le 1/2 \\ 3/2 - u & 1/2 < u \le 1 \end{cases}.$$

While we will restrict our focus in the sequel to $g(u) = 1-u$ and to variance reduction of mean estimators, we encourage the reader to reference the excellent article by Whitt (1976) for further discussion of this topic. We will denote an antithetic draw from this distribution by $(X_1, X_2) \overset{\text{anti}}{\sim} \text{Pois}(\tau)$. Recall that the distribution of the increment $N(s,t)$ of a unit-rate Poisson process Y over an interval $[s,t]$ is given by $N(s,t) := Y(t) - Y(s) \sim \text{Pois}(t-s)$, and is independent of increments of the same Poisson process over other, disjoint intervals. Further, recall that its arrival times conditioned on its value at the endpoints of any interval are uniformly distributed throughout that interval.

The first approach, the endpoint technique, produces an antithetic pair of Poisson process paths (Y^1, Y^2) over an interval $[0, \tau_f]$. One way to specify a pair of continuous-time Poisson process trajectories is to simulate the sequences $\left(\mathcal{A}^1_{\tau_f}, \mathcal{A}^2_{\tau_f}\right)$ of their arrival times. We achieve this as follows. For some step-increment $\tau_s > 0$, sample $\left(Y^1(\tau_s), Y^2(\tau_s)\right) = \left(N^1_E(0,\tau_s), N^2_E(0,\tau_s)\right) \overset{\text{anti}}{\sim} \text{Pois}(\tau_s)$, the state of the pair of Poisson processes at time τ_s as an antithetic pair of Poisson random variables. Here, we denote the increments of process Y^j constructed using the endpoint technique from time s to time t with $N^j_E(s,t)$. Then, we can sample the state of the processes at time $2\tau_s$ by sampling the next process increment as an independent antithetic pair, $\left(N^1_E(\tau_s, 2\tau_s), N^2_E(\tau_s, 2\tau_s)\right) \overset{\text{anti}}{\sim} \text{Pois}(\tau_s)$, and setting $Y^i(2\tau_s) = N^i_E(0,\tau_s) + N^i_E(\tau_s, 2\tau_s)$. We may proceed until we have the sampled values $\{Y^i(\tau_s), Y^i(2\tau_s), \ldots, Y^i((N+1)\tau_s)\}_{i=1}^2$ of the state of the antithetic pair of paths evaluated at multiples of τ_s, where $N := \lfloor \tau_f/\tau_s \rfloor$ is the total number of full steps. We can then finish simulating the paths by sampling the jump times, which are iid uniform random variables. That is, within each interval $(n\tau_s, (n+1)\tau_s]$, we know that $N^i_E(n\tau_s, (n+1)\tau_s)$ jumps occurred, and that each jump is uniformly distributed over the interval. More formally, $t^i_{j,n} \overset{\text{i.i.d.}}{\sim} \text{Unif}(n\tau_s, (n+1)\tau_s)$ for $j = 1, \ldots, N^i_E(n\tau_s, (n+1)\tau_s)$ and for $n = 0, \ldots, N$. Finally, we re-index the jump times in j such that they are sorted in increasing order and we discard any arrival times that lie outside $[0, \tau_f]$. The endpoint technique is summarized in Algorithm 1.

While the anticorrelation is only injected at sample points $\{\tau_s, 2\tau_s, \ldots\}$ and jump times are simulated iid, we will show in 2.3 that negative correlation is felt throughout the time domain of the process, not just at the antithetic sample points.

Though a more complete and precise analysis is provided in Sect. 2.3, we attempt here to provide some intuition regarding the performance and limitations of the endpoint method, in order to motivate the development of the other algorithm we will present, in this work, the binomial midpoint method.

First, we define a useful performance metric. Since all of the mean estimators we construct in this paper are unbiased, we define the scaled mean-square error (MSE) of a pathwise mean estimator δ at time t to be

Algorithm 1 Endpoint Method: Antithetic Poisson process paths via concatenation of step size τ_s

Intialize: $n \leftarrow 0$, $\left(\mathcal{A}^1_{\tau_f}, \mathcal{A}^2_{\tau_f}\right) \leftarrow (\emptyset, \emptyset)$
while $n\tau_s < \tau_f$ **do**
  Sample antithetic Poisson random variables: $\left(N^1_E(0, \tau_s), N^2_E(0, \tau_s)\right) \overset{\text{anti}}{\sim} \text{Pois}(\tau_s)$
  for $j = 1, \ldots, N^i_E(0, \tau_s)$ **do**
    Simulate iid jump times in the interval: $t^i_{j,n} \overset{\text{i.i.d.}}{\sim} \text{Unif}(0, \tau_s),\ i \in \{1, 2\}$
  end for
  Sort and append arrival times: $\mathcal{A}^i_{\tau_f} \leftarrow \mathcal{A}^i_{\tau_f} \cup \text{sort}\left(\left\{n\tau_s + t^i_{j,n}\right\}_{j=1}^{N^i_E}\right),\ i \in \{1, 2\}$
  $n \leftarrow n + 1$
end while
$\mathcal{A}^i_{\tau_f} \leftarrow \mathcal{A}^i_{\tau_f} \cap [0, \tau_f],\ i \in \{1, 2\}$
return $\left(\mathcal{A}^1_{\tau_f}, \mathcal{A}^2_{\tau_f}\right)$

$$\text{MSE}(t) := N\,\text{Var}\left(\delta(t)\right) = N\,\text{Var}\left(\frac{1}{N}\sum_{i=1}^{N} Y^i(t)\right), \tag{2}$$

where N is the number of sample paths used to produce the mean estimate. This quantity is of interest both because it is invariant to the inclusion of additional iid paths (or pairs of antithetic paths, as the case may be) and because it cleanly relates to other quantities of interest. For example, for any mean estimator

$$\delta_{2M}(t) = \frac{1}{2M}\sum_{i=1}^{M}\left[Y^{1,i}(t) + Y^{2,i}(t)\right], \tag{3}$$

where the pairs $(Y^{1,i}, Y^{2,i})$ are iid in i but their elements could be correlated. Then,

$$\begin{aligned}
\text{MSE}_{\delta_{2M}}(t) &= 2M\,\text{Var}\left(\frac{1}{2M}\sum_{i=1}^{M}\left[Y^{1,i}(t) + Y^{2,i}(t)\right]\right) \\
&= \frac{2}{M}\sum_{i=1}^{M}\text{Var}\left(\frac{Y^{1,i}(t) + Y^{2,i}(t)}{2}\right) \\
&= 2\,\text{Var}\left(\delta_2(t)\right) \\
&= \text{MSE}_{\delta_2}(t) \\
&= \text{Var}\left(Y^1(t)\right) + \text{Cov}\left(Y^1(t), Y^2(t)\right) \\
&= t + \text{Cov}\left(Y^1(t), Y^2(t)\right).
\end{aligned}$$

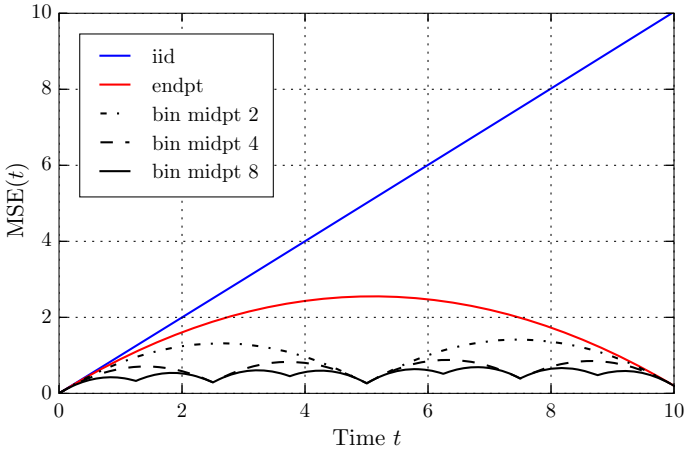

Fig. 1 The scaled MSE of 2-sample mean estimators, each produced from a pair of iid, endpoint, or binomial midpoint antithetic Poisson process paths, all simulated for a single step of length $\tau_s = 10$. For comparison, we show binomial midpoint estimators constructed using 2, 4, and 8 sub-steps. Note that each subsequent mean estimator dominates the previous one, meaning it has lower MSE(t) for all t

In particular, the scaled MSE at time t of a mean estimator constructed from any number of iid sample paths is simply $\mathrm{MSE}_{\widetilde{\delta}_M}(t) = t$. For convenience, we will frequently denote $\mathrm{MSE}_{\delta}(t)$ as simply $\mathrm{MSE}(t)$.

As we will prove in the sequel, a mean estimator $\delta(t)$ constructed using the endpoint method has $\mathrm{MSE}(\tau_s) = \tau_s + \mathrm{Cov}\left(N_E^1(0,\tau_s), N_E^2(0,\tau_s)\right) \leq \tau_s$. That is, its MSE is that of the iid estimator plus the negative covariance between the antithetically sampled Poisson random variables with parameter τ_s. Note that this MSE is also strictly greater than 0, since the Poisson distribution is not symmetric. Due to iid jump times, MSE(t) for this mean estimate is a piecewise concave quadratic function for $t \in (0, \tau_s)$. It is determined by $\mathrm{MSE}(0) = 0$, the value of $\mathrm{MSE}(\tau_s)$ (which is fixed by the covariance between two antithetic Poisson random variables), and the fact that $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{MSE}(t)|_{t=0^+} = 1$ (see Lemma 1). This is illustrated by the red curve in Fig. 1, for $\tau_s = 10$ over an interval $[0, \tau_f = 10]$.

We may further reduce the MSE over most of this interval by reducing τ_s, and thus, injecting negative correlation more frequently in the interval. Compare the endpoint MSE (red) curve in Fig. 1 ($\tau_s = 10$) to the same curve in Fig. 2 ($\tau_s \approx 2.5$) to observe this process. However, note that the gains here present a tradeoff. Indeed, we can see that MSE(10) is increased by taking four steps instead of one. Using independent increments of the endpoint method at τ_s step intervals:

$$\begin{aligned}\mathrm{MSE}(4\tau_s) &= 4\tau_s + \sum_{n=0}^{3}\mathrm{Cov}\left(N_E^1(n\tau_s,(n+1)\tau_s), N_E^2(n\tau_s,(n+1)\tau_s)\right)\\ &= 4\tau_s + 4\,\mathrm{Cov}\left(N_E^1(0,\tau_s), N_E^2(0,\tau_s)\right)\\ &\geq 4\tau_s + \mathrm{Cov}\left(N_E^1(0,4\tau_s), N_E^2(0,4\tau_s)\right), \end{aligned} \tag{4}$$

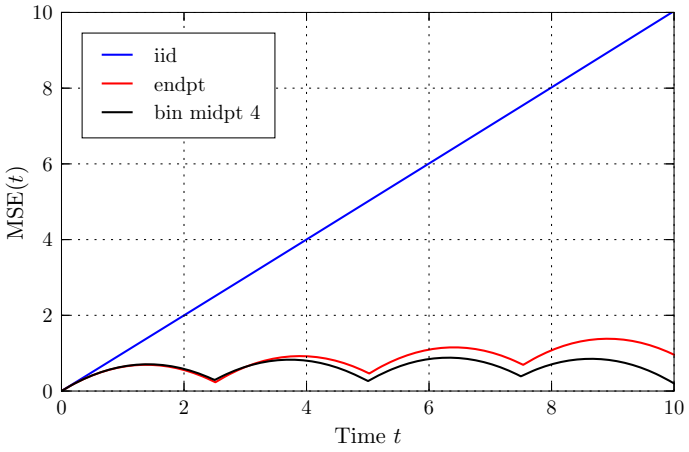

Fig. 2 The scaled MSE of iid, endpoint and binomial midpoint 2-sample mean estimators. Endpoint technique uses $\tau_s \approx 2.5$. Binomial midpoint technique uses $\tau_s = 10.0$ with 4 sub-steps, or sub-step size 2.5. Note that the endpoint estimator achieves similar performance to the binomial midpoint estimator, but accumulates slightly more MSE with each step, as shown in (4). These two antithetic estimators require an almost identical number of random variable draws to simulate on average

since the covariance between antithetic Poisson variables is sub-linear in their parameter Maginnis (2011). So each time we step forward by τ_s, we accumulate MSE from the previous endpoint and this accumulation exceeds the MSE at the endpoint of a single, larger step. This difference is small at first and still reduces the overall MSE in the interval, but as we let τ_s get even smaller, eventually the MSE not only becomes significantly larger at the endpoint of the interval than before, it is also larger over the majority of the interval, as shown in Fig. 3 ($\tau_s \approx 0.0625$). In fact, the expression we will prove in Theorem 2 shows that as $\tau_s \to 0$, $\text{MSE}_\delta(t) \to \text{MSE}_{\widetilde{\delta}}(t)$, the MSE of the iid estimator. So then, is there a way to reduce MSE more evenly for $t \in (0, \tau_s)$? The next algorithm we present does exactly that.

2.2 Binomial Midpoint Method for Increased Variance Reduction

As motivated above, τ_s is the primary parameter that governs the reduction in MSE for antithetic simulation of unit-rate Poisson processes using these techniques. There are limitations to what modifying τ_s alone can do, however. Indeed, as we illustrated, with changing τ_s there is a tradeoff where near-time performance (within a fixed time window, say) competes with long-term performance (after compounding many small steps, say). Instead of reducing τ_s to improve near-time performance, we may instead antithetically sub-sample previous times (using the conditional binomial distribution) so that we improve local performance in much the same way that, say, halving τ_s does, but without sacrificing endpoint performance. We refer to this approach as the binomial midpoint method.

The binomial midpoint method injects more negative correlation into the Poisson process pair by antithetically sampling values in the interior of a step after sampling

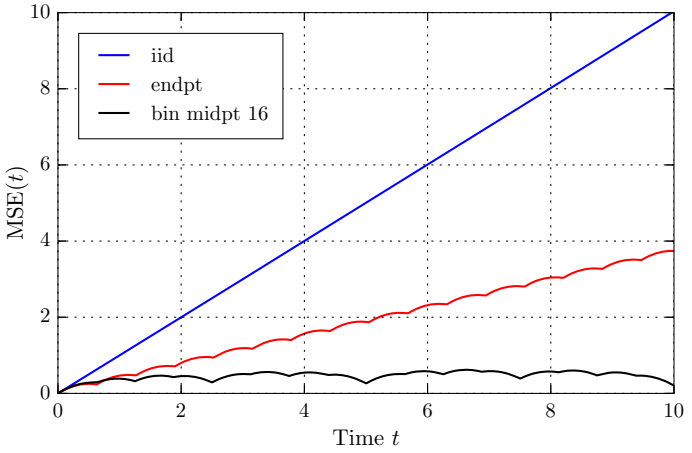

Fig. 3 The scaled MSE of iid, endpoint and binomial midpoint 2-sample mean estimators. Endpoint technique uses $\tau_s \approx 0.63$. Binomial midpoint technique uses $\tau_s = 10.0$ with 16 sub-steps, for sub-step size of $10/16 = 0.625$. When the step size of the endpoint technique becomes sufficiently small, its MSE accumulates rapidly. For the binomial midpoint technique, MSE(10) is not affected by the number of sub-steps it takes in [0, 10.0]. These two antithetic estimators require an almost identical number of random variable draws to simulate on average

its endpoint. Here, we exploit the fact that, conditioned on past and future values, the Poisson process has binomial distribution. First, we may simulate $\left(Y^1(\tau_s), Y^2(\tau_s)\right)$, exactly as in the endpoint method using antithetic Poisson sampling. But then, instead of merely sampling the iid jump times over $[0, \tau_s]$ as in the endpoint method, we first conditionally sample additional antithetic values of the process at interior time points. For example, $\left(Y^1(\tau_s/2), Y^2(\tau_s/2)\right)$, which, conditioned on $\left(Y^1(\tau_s), Y^2(\tau_s)\right)$ are binomially distributed, i.e., $Y^i(\tau_s/2)|Y^i(\tau_s) \sim \text{Bin}\left(Y^i(\tau_s), 1/2\right)$ for $i = 1, 2$, may be sampled antithetically by inverting their respective binomial CDFs. Note that this antithetic pair will no longer have identical conditional CDFs, since, for a particular pair of trajectories $Y^1(\tau_s) \neq Y^2(\tau_s)$. But we can still introduce additional negative correlation at this point since both distribution functions are still non-decreasing. And, importantly, we have not increased $\text{MSE}(\tau_s)$ since it has already been sampled. We find it best in practice to sub-sample at time points that bisect the interval formed by times where the Poisson process has already been sampled, since the binomial parameter $p = 1/2$ symmetrizes the binomial CDF, enhancing the efficacy of antithetic sampling. Other choices of sub-intervals are permissible, however, and it is possible that in some applications, sub-sampling on asymmetric intervals may be desirable. Thus, additional negative correlation can be introduced at subsequent dyadic intervals by conditioning on the nearest previous and future values that have already been sampled. For instance $N^i(3\tau_s/4) := \left(Y^i(3\tau_s/4) - Y^i(\tau_s/2)\right)|Y^i(\tau_s/2), Y^i(\tau_s) \sim \text{Bin}\left(Y^i(\tau_s) - Y^i(\tau_s/2), 1/2\right)$ for $i = 1, 2$. For the sake of compactness, we denote the conditional sub-increment of process Y^i from time s to time $(s + t)/2$ (i.e., from the beginning of the interval $[s, t]$ to its midpoint) as

$$Q^i(s,t) := \left(Y^i((s+t)/2) - Y^i(s)\right) | Y^i(s), Y^i(t), \tag{5}$$

where $Q^i(s,t) \sim \text{Bin}\left(Y^i(t) - Y^i(s), 1/2\right)$. The binomial midpoint technique is summarized in Algorithm 2.

Algorithm 2 Binomial Midpoint Method: Conditional sub-sampling in τ_s-increment on 2^L dyadic points

Intialize: $n \leftarrow 0$, $\left(\mathcal{A}^1_{\tau_f}, \mathcal{A}^2_{\tau_f}\right) \leftarrow (\emptyset, \emptyset)$
while $n\tau_s < \tau_f$ **do**
  $Y^i_B(0) \leftarrow 0$, $\mathcal{A}^i_B \leftarrow \emptyset$, $i \in \{1,2\}$,
  Sample antithetic Poisson random variables: $\left(N^1_B(0,\tau_s), N^2_B(0,\tau_s)\right) \overset{\text{anti}}{\sim} \text{Pois}(\tau_s)$
  Set state value at endpoint τ_s: $Y^i_B(\tau_s) \leftarrow N^i_B(0,\tau_s)$, $i \in \{1,2\}$
  for $\ell = 1$ **to** L **do**
    for $k = 1, 3, 5, \ldots, 2^\ell - 1$ **do**
      Conditionally sample increments at midpoint as antithetic binomial variables (see Eq. (5)):
      $Q^i\left(\frac{k-1}{2^\ell}\tau_s, \frac{k+1}{2^\ell}\tau_s\right) \overset{\text{anti}}{\sim} \text{Bin}\left(Y^i_B\left(\frac{k+1}{2^\ell}\tau_s\right) - Y^i_B\left(\frac{k-1}{2^\ell}\tau_s\right), \frac{1}{2}\right)$, $i \in \{1,2\}$
      Set state value at midpoint: $Y^i_B\left(\frac{k}{2^\ell}\tau_s\right) \leftarrow Y^i_B\left(\frac{k-1}{2^\ell}\tau_s\right) + Q^i\left(\frac{k-1}{2^\ell}\tau_s, \frac{k+1}{2^\ell}\tau_s\right)$, $i \in \{1,2\}$
    end for
  end for
  for $k = 1$ **to** $2^L - 1$ **do**
    for $j = 1$ **to** $Y^i_B\left(\frac{k+1}{2^L}\tau_s\right) - Y^i_B\left(\frac{k}{2^L}\tau_s\right)$ **do**
      Simulate iid jump times: $t^i_{j,k} \overset{\text{i.i.d.}}{\sim} \text{Unif}\left(\frac{k}{2^L}\tau_s, \frac{k+1}{2^L}\tau_s\right)$, $i \in \{1,2\}$
    end for
    Sort and append jump times: $\mathcal{A}^i_B \leftarrow \text{sort}\left(\{n\tau_s + t^i_{j,k}\}_j\right)$, $i \in \{1,2\}$
  end for
  $\mathcal{A}^i_{\tau_f} \leftarrow \mathcal{A}^i_{\tau_f} \cup \mathcal{A}^i_B$, $i \in \{1,2\}$
  $n \leftarrow n + 1$
end while
$\mathcal{A}^i_{\tau_f} \leftarrow \mathcal{A}^i_{\tau_f} \cap [0, \tau_f]$, $i \in \{1,2\}$
return $\left(\mathcal{A}^1_{\tau_f}, \mathcal{A}^2_{\tau_f}\right)$

Note that the endpoint method is merely a special case of the binomial midpoint method. Indeed, the binomial midpoint method with no partitions between steps (i.e., when the order $L = 0$, which can be interpreted as the number of times we halve the sub-interval length) is precisely the endpoint method. Note also that we introduce additional cost (in the form of additional random variable samples) to achieve this variance reduction. If we divide a particular step of length τ_s into 2^L equal-length sub-intervals, we require $2^L - 1$ additional CDF inversions. As we will show in the sequel, this cost-error tradeoff is profitable for a finite order L that depends on the operating parameters of the system and particular simulation. Additionally, while MSE will always be reduced by increasing L, note that this reduction decreases sharply for large L. As a result, it should not be thought of as an asymptotic parameter that drives MSE toward zero for large values and fixed τ_s. Indeed, note that as $\tau_s/2^L$ becomes small, the first parameter of the corresponding binomial distributions will also be small, since

the expected change in the process will be small over this sub-interval. This limits the impact of antithetic sampling. In this case, the corresponding reduction in MSE is small, and it would have been more efficient to partition into 2^{L-1} sub-intervals instead, incurring roughly half the computational cost. So L is best thought of as a finite parameter that has significant benefit for small values and saturates quickly for large values. This saturation point will largely be determined by the quantity $\tau_s/2^L$.

2.3 Analysis of Antithetic Poisson Processes in the Endpoint Case

We now provide some analysis of the antithetic algorithm above. In this section, we present two useful metrics to quantify the expected error from mean estimators constructed using the algorithms: the scaled mean-square error and the integrated scaled mean-square error. Next, we define a special function related to the antithetic simulation of Poisson random *variables* that will help us analyze the behavior of antithetically simulated Poisson random *processes*. Finally, we present several results that explicitly and exactly quantify the scaled MSE and integrated scaled MSE behavior of antithetic endpoint Poisson process simulation, which in particular we then use to obtain asymptotic performance bounds.

Let $\left(\widetilde{Y}^1, \widetilde{Y}^1\right)$ denote a pair of iid, unit-rate Poisson process, so that $\text{Cov}\left(\widetilde{Y}^1(t), \widetilde{Y}^2(t)\right) = 0$ for all $t \geq 0$. Let $\left(Y^1, Y^2\right)$ denote the antithetic, unit-rate Poisson processes constructed using Algorithm 1 above, so that Y^1 and Y^2 are correlated (indeed, we will show that $\text{Cov}\left(Y^1(t), Y^2(t)\right) \leq 0$ for all $t \geq 0$). Let $\widetilde{\delta}(t)$ and $\delta(t)$ denote the 2-sample mean estimators obtained by averaging the iid and endpoint Poisson process pairs, respectively. For brevity, we will refer to mean estimators by the method used to simulate their constituent sample paths (e.g., iid estimator, endpoint estimator, binomial midpoint estimator). Recall the scaled MSE defined in Eq. (2) given by

$$\text{MSE}(t) := N \text{ Var}\left(\delta(t)\right) = N \text{ Var}\left(\frac{1}{N}\sum_{i=1}^{N} Y^i(t)\right),$$

where N is the number of sample paths used to construct the mean estimate.

Note that, as with any Monte Carlo scheme, we may produce more accurate mean estimates by increasing the number of samples used to construct the estimator. Practitioners can simulate a sequence of many antithetic pairs which are iid with respect to each other (each with 2 correlated components, of course) to create mean estimates of sufficient accuracy for their particular application. This decrease in variance will scale in the usual way ($1/N$, or $1/\sqrt{N}$ w.r.t. the standard deviation), so we restrict our analyses to mean estimates constructed from a single antithetic pair of random paths. Further, all comparisons are made to iid mean estimates constructed using two independent sample paths.

Recall that $(X_1, X_2) \overset{\text{anti}}{\sim} \text{Pois}(\tau)$ denotes the anticorrelated scalar Poisson variable pair, i.e.,

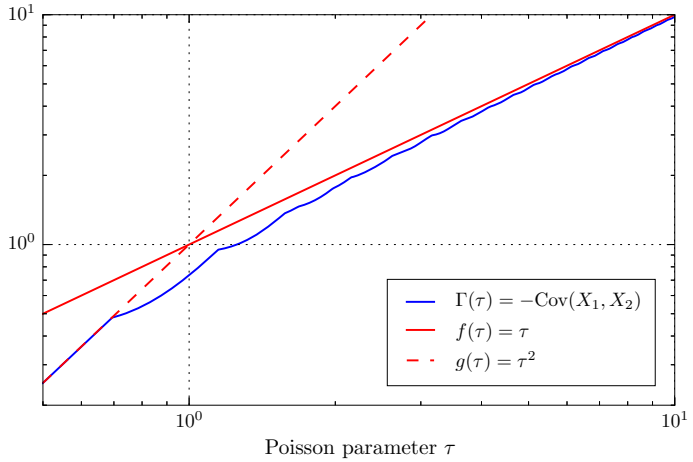

Fig. 4 The special function $\Gamma(\tau) = -\operatorname{Cov}(X_1, X_2)$ (where $(X_1, X_2) \overset{\text{anti}}{\sim} \text{Pois}(\tau)$) plotted versus Poisson parameter τ. For reference, we also show the simple functions $f(\tau) = \tau$ and $g(\tau) = \tau^2$. Note that $0 \leq \Gamma(\tau) < \tau$ for all τ and, for $\tau < \ln 2$, $\Gamma(\tau) = \tau^2$

$$X_1 := F_\tau^{-1}(U)$$
$$X_2 := F_\tau^{-1}(1-U),$$

where $U \sim \text{Unif}[0, 1]$ is a uniform random variable and F_τ^{-1} is the formal inverse of the Poisson CDF with parameter τ. For such an antithetic pair, define the special function

$$\Gamma(\tau) := -\operatorname{Cov}(X_1, X_2) = \tau^2 - \int_{u=0}^{1} F_\tau^{-1}(u) F_\tau^{-1}(1-u)\, \mathrm{d}u \geq 0, \tag{6}$$

the negative covariance of a pair of antithetically sampled Poisson scalar random variables. This function will appear frequently in the analysis of the variance properties of antithetically simulated Poisson process paths. It has several useful properties. In particular, note that $\Gamma(\tau) \leq \tau = \operatorname{Var}(X_1)$ by definition and $\Gamma(\tau) \geq 0$ for all τ (Maginnis et al. 2016). Note that $\Gamma(\tau) = \tau^2$ for all $\tau < \ln 2$ (Maginnis 2011). These relationships, as well as Γ itself, are shown in Fig. 4. The scaled MSE of a mean estimator constructed from (X_1, X_2) is related to Γ by $\text{MSE} = 2\operatorname{Var}\left(\frac{X_1+X_2}{2}\right) = \operatorname{Var}(X_1) - \Gamma(\tau) = \tau - \Gamma(\tau)$. This quantity is plotted in Fig. 5.

We proceed by exactly characterizing the variance properties of a mean estimator constructed from two antithetic sample paths (Y^1, Y^2) of the unit-rate, Poisson process simulated using the endpoint technique defined above. First, we provide an expression for the variance of the estimator at every time, and then, we motivate and provide an expression for a more useful quantity, the integral of the estimator variance over a fixed time window $[0, \tau_\text{f}]$. We begin with a lemma that characterizes how the covariance between two correlated Poisson processes propagates from points of direct anticorrelation to times where direct anticorrelation is *not* applied.

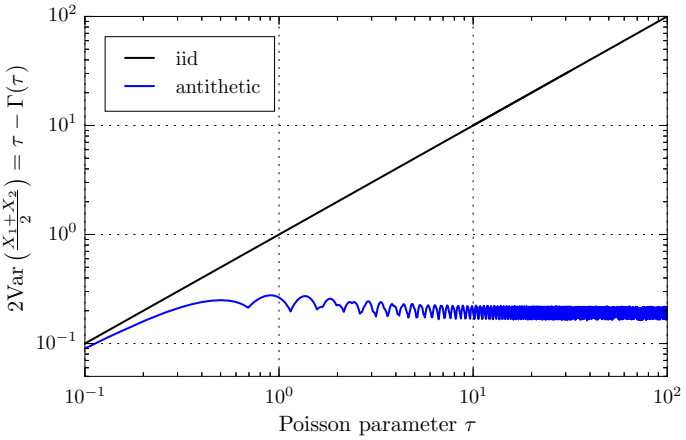

Fig. 5 Scaled MSE of a mean estimate constructed using two Poisson random variables, sampled either using iid or antithetic sampling, plotted versus Poisson parameter τ. Note that this variance remains bounded below by a small positive constant, even for large τ. This suggests that $\Gamma(\tau)$ does not converge to τ as $\tau \to \infty$

Lemma 1 *For antithetic unit-rate Poisson processes Y^1, Y^2, and for $0 \leq T_1 < T_2$, denote by $\mathcal{G}(T_1, T_2) := \sigma\left\{Y^1(T_1), Y^1(T_2), Y^2(T_1), Y^2(T_2)\right\}$, the sigma algebra generated by the 4 random variables obtained by evaluating each process at each endpoint. Then, if $Y^1(t)$ and $Y^2(t)$ are conditionally independent given $\mathcal{G}(T_1, T_2)$ for every $t \in (T_1, T_2)$, then:*

$$\begin{aligned}\mathrm{Cov}\left(Y^1(t), Y^2(t)\right) &= \mathrm{Cov}\left(Y^1(T_1), Y^2(T_1)\right) \\ &\quad + \frac{(t-T_1)^2}{(T_2-T_1)^2}\,\mathrm{Cov}\left(N^1(T_1,T_2), N^2(T_1,T_2)\right), \end{aligned} \tag{7}$$

for every $t \in [T_1, T_2]$, where $N^i(T_1, T_2) := Y^i(T_2) - Y^i(T_1)$ is the increment of the process.

Proof We proceed using the law of total expectation and the conditional independence hypothesis as follows:

$$\begin{aligned}\mathbb{E}\left[Y^1(t)Y^2(t)\right] &= \mathbb{E}\left[\mathbb{E}\left[Y^1(t)Y^2(t)\middle|\mathcal{G}\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[Y^1(t)\middle|\mathcal{G}\right]\mathbb{E}\left[Y^2(t)\middle|\mathcal{G}\right]\right].\end{aligned}$$

We may then apply the fact that, conditioned on its endpoints (information of which is contained in the filtration $\mathcal{G}$), a unit-rate Poisson process increment at time t is binomially distributed, with number of trials $N^i(T_1, T_2) = Y^i(T_2) - Y^i(T_1)$ and probability of success $(t - T_1)/(T_2 - T_1)$. This fact, combined with the independent increments property gives us

$$\begin{aligned}\mathbb{E}\left[Y^1(t)Y^2(t)\right] &= \mathbb{E}\Bigg[\left(Y^1(T_1) + \frac{t-T_1}{T_2-T_1}N^1(T_1,T_2)\right)\\ &\quad\cdot\left(Y^2(T_1) + \frac{t-T_1}{T_2-T_1}N^2(T_1,T_2)\right)\Bigg]\\ &= \mathbb{E}\left[Y^1(T_1)Y^2(T_1)\right]\\ &\quad + 2T_1(t-T_1) + \frac{(t-T_1)^2}{(T_2-T_1)^2}\mathbb{E}\left[N^1(T_1,T_2)N^2(T_1,T_2)\right].\end{aligned}$$

Note also that

$$t^2 = T_1^2 + 2T_1(t-T_1) + \frac{(t-T_1)^2}{(T_2-T_1)^2}(T_2-T_1)^2, \tag{8}$$

so that

$$\begin{aligned}\mathrm{Cov}(Y^1(t), Y^2(t)) &= \mathbb{E}\left[Y^1(t)Y^2(t)\right] - t^2\\ &= \mathbb{E}\left[Y^1(T_1)Y^2(T_1)\right] - T_1^2\\ &\quad + \frac{(t-T_1)^2}{(T_2-T_1)^2}\left(\mathbb{E}\left[N^1(T_1,T_2)N^2(T_1,T_2)\right] - (T_2-T_1)^2\right)\\ &= \mathrm{Cov}(Y^1(T_1), Y^2(T_1))\\ &\quad + \frac{(t-T_1)^2}{(T_2-T_1)^2}\mathrm{Cov}(N^1(T_1,T_2), N^2(T_1,T_2)),\end{aligned}$$

and the claim holds. □

Lemma 1 allows us to express the scaled MSE of an endpoint estimator with step size τ_s at any time t in terms of the special function Γ evaluated at τ_s. This explicit expression is derived in Theorem 2.

Theorem 2 *For anticorrelated unit-rate Poisson processes (Y^1, Y^2), sampled using the antithetic endpoint technique with step size τ_s, the scaled MSE of the corresponding mean estimator is piecewise quadratic and is given exactly by:*

$$\mathrm{MSE}(t) = 2\,\mathrm{Var}\left(\frac{Y^1(t)+Y^2(t)}{2}\right) = t - n\Gamma(\tau_s) - \frac{(t-n\tau_s)^2}{{\tau_s}^2}\Gamma(\tau_s), \tag{9}$$

for every $t \in [n\tau_s, (n+1)\tau_s]$.

Proof We proceed by first noting that the covariance between the Poisson increments $\mathrm{Cov}\left(N^1(T_1,T_2), N^2(T_1,T_2)\right)$ present in the last term of (7) is exactly equal to $-\Gamma(T_2-T_1)$, since $\left(N^1(T_1,T_2), N^2(T_1,T_2)\right)$ are just antithetically sampled Poisson random variables with parameter T_2-T_1. We proceed by induction on n. First, note that the conditions of Lemma 1 are satisfied for $T_1 = 0 < T_2 = \tau_s$ since, conditioned

on the σ-algebra $\mathcal{G}(0, \tau_s) = \sigma\left\{Y^1(0) = 0, Y^2(0) = 0, Y^1(\tau_s), Y^2(\tau_s)\right\}$, the random variables $Y^1(t)$, $Y^2(t)$ are independent for all $t \in (0, \tau_s)$. Thus

$$\begin{aligned}\mathrm{Cov}\left(Y^1(t), Y^2(t)\right) &= \mathrm{Cov}\left(Y^1(0), Y^2(0)\right) + \frac{t^2}{{\tau_s}^2}\,\mathrm{Cov}\left(N^1(0, \tau_s), N^2(0, \tau_s)\right)\\ &= -\frac{t^2}{{\tau_s}^2}\Gamma(\tau_s),\end{aligned}$$

and

$$\begin{aligned}\mathrm{MSE}(t) &= \mathrm{Var}\left(Y^1(t)\right) + \mathrm{Cov}\left(Y^1(t), Y^2(t)\right)\\ &= t - \frac{t^2}{{\tau_s}^2}\Gamma(\tau_s)\end{aligned}$$

for all $t \in [0, \tau_s]$ (i.e., $n = 0$).

Now, suppose that the claim holds for $n - 1$, namely that, for $t \in [(n-1)\tau_s, n\tau_s]$,

$$\mathrm{MSE}(t) = 2\,\mathrm{Var}\left(\frac{Y^1(t) + Y^2(t)}{2}\right) = t - (n-1)\Gamma(\tau_s) - \frac{(t - (n-1)\tau_s)^2}{{\tau_s}^2}\Gamma(\tau_s),$$

and in particular that

$$\begin{aligned}\mathrm{Var}\left(Y^1(n\tau_s)\right) + \mathrm{Cov}\left(Y^1(n\tau_s), Y^2(n\tau_s)\right) &= n\tau_s - (n-1)\Gamma(\tau_s)\\ &\quad - \frac{(n\tau_s - (n-1)\tau_s)^2}{{\tau_s}^2}\Gamma(\tau_s)\\ \implies \mathrm{Cov}\left(Y^1(n\tau_s), Y^2(n\tau_s)\right) &= -(n-1)\Gamma(\tau_s) - \frac{(\tau_s)^2}{{\tau_s}^2}\Gamma(\tau_s)\\ &= -n\Gamma(\tau_s).\end{aligned}$$

By construction, for $t \in (n\tau_s, (n+1)\tau_s)$, $Y^1(t)$ and $Y^2(t)$ are independent conditioned on $\mathcal{G}(n\tau_s, (n+1)\tau_s)$, since all random sampling inside the interval is iid uniform, given the endpoints. So the conditions of Lemma 1 again hold, and

$$\begin{aligned}\mathrm{Cov}\left(Y^1(t), Y^2(t)\right) &= \mathrm{Cov}\left(Y^1(n\tau_s), Y^2(n\tau_s)\right) + \frac{(t - n\tau_s)^2}{((n+1)\tau_s - n\tau_s)^2}\\ &\quad \mathrm{Cov}\left(N^1(n\tau_s, (n+1)\tau_s), N^2(n\tau_s, (n+1)\tau_s)\right)\\ &= -n\Gamma(\tau_s) - \frac{(t - n\tau_s)^2}{{\tau_s}^2}\Gamma(\tau_s),\end{aligned}$$

for all $t \in [n\tau_s, (n+1)\tau_s]$. So then, for all $t \in [n\tau_s, (n+1)\tau_s]$,

$$\begin{aligned}\mathrm{MSE}(t) &= \mathrm{Var}\left(Y^1(t)\right) + \mathrm{Cov}\left(Y^1(t), Y^2(t)\right) \\ &= t - n\Gamma(\tau_s) - \frac{(t - n\tau_s)^2}{{\tau_s}^2}\Gamma(\tau_s)\end{aligned}$$

and the claim holds for n. □

The expression (9) proven in Theorem 2 combined with our intuition about the function Γ suggests that the larger step size τ_s we take, the greater variance reduction we will observe over a long period of time. In practice, however, stochastic simulation will often be performed over a relatively fixed finite time window, dictated by the system parameters or problem of interest. Reductions in estimator variance beyond that window of interest, which we will denote by $[0, \tau_f]$, are of little benefit since they will never be observed, and in particular they may adversely affect the performance of mean estimates in the window of simulation. Thus, a better metric for comparison between techniques is the total MSE over a fixed finite time interval $[0, \tau_f]$. For simplicity, we proceed using the $\mathcal{L}_1$-integral of MSE as our metric of choice.

Lemma 3 *For the endpoint technique implemented with step size τ_s over time interval $[0, \tau_f]$, let $N := \lfloor \tau_f/\tau_s \rfloor$ be the number of incremental Poisson samples (i.e., full steps) taken in the interval. Then,*

$$\begin{aligned}\int_0^{\tau_f} \mathrm{MSE}(t)\,\mathrm{d}t &= \int_0^{\tau_f} 2\,\mathrm{Var}\left(\frac{Y^1(t) + Y^2(t)}{2}\right)\mathrm{d}t \\ &= \frac{N(N-1)}{2}\tau_s\left(\tau_s - \Gamma(\tau_s)\right) + \frac{N}{6}\tau_s\left(3\tau_s - 2\Gamma(\tau_s)\right) \\ &\quad + N\left(\tau_s - \Gamma(\tau_s)\right)\left(\tau_f - N\tau_s\right) + \frac{(\tau_f - N\tau_s)^2}{2} - \frac{(\tau_f - N\tau_s)^3}{3\tau_s^2}\Gamma(\tau_s).\end{aligned} \tag{10}$$

The proof is lengthy but straightforward and is omitted here for brevity. While this expression is exact, it can be difficult to parse in the general case. For comparison, note that the integrated variance of the iid mean estimator is given by

$$\int_0^{\tau_f} \mathrm{MSE}_{\widetilde{\delta}}(t)\,\mathrm{d}t = \int_0^{\tau_f} 2\,\mathrm{Var}\left(\frac{\widetilde{Y}^1(t) + \widetilde{Y}^2(t)}{2}\right)\mathrm{d}t = \int_0^{\tau_f} t\,\mathrm{d}t = \frac{1}{2}\tau_f^2. \tag{11}$$

As will be shown in Fig. 7, we can see that the integrated MSE of the endpoint estimator is always less than the same quantity for the iid mean estimator. Two simple extremal cases are also illustrative. Consider the case when $\tau_s > \tau_f$, i.e., when less than one step is used for simulation. In this case, $N = 0$ and (10) reduces to

$$\int_0^{\tau_f} \mathrm{MSE}(t)\,\mathrm{d}t = \frac{1}{2}\tau_f^2 - \frac{\Gamma(\tau_s)}{3\tau_s^2}\tau_f^3. \tag{12}$$

For fixed τ_f, as $\tau_s \to \infty$, the performance of the antithetic estimator will degrade back to the iid estimator, since $\Gamma(\tau_s) \leq \tau_s$. Alternatively, for fixed τ_f, when $\tau_s \to 0$, i.e., many Poisson steps are being taken during the course of a simulation, $0 \leq \tau_f - N\tau_s \leq \tau_s \to 0$ and $\Gamma(\tau_s) = \tau_s^2$. Thus (10) becomes

$$\int_0^{\tau_f} \mathrm{MSE}(t)\,\mathrm{d}t \approx \frac{1}{2}\tau_f^2\,(1-\tau_s) - \frac{1}{2}\tau_f\tau_s + \mathcal{O}(\tau_s^2), \tag{13}$$

and performance again degrades to the iid case. This suggests that the best performing τ_s is one that is neither too large nor too small relative to the window of interest, a claim that is further supported by the numerical results in Sect. 2.4.

2.4 Numerical Results for Antithetic Poisson Processes

We now support the analytical results of the previous section with numerical experiments. In particular, we are interested in examining the relationship between the choice of Poisson simulation step time τ_s and the integrated scaled MSE over the time window $[0, \tau_f]$. Note here that even though we find it helpful to parameterize the algorithms by Poisson step size τ_s, this is not the finest increment in the processes produced. Those are the inter-arrival times of the Poisson processes. However, these are stochastic and not suitable as parameters for our approach. As a result, we will consider cases where the Poisson step size τ_s is larger than the time window endpoint τ_f for the sake of completeness.

For each of the Poisson process simulation algorithms, iid and binomial midpoint, a pair of unit-rate Poisson processes is simulated from $t = 0$ to τ_f using a step size of τ_s, and averaged to produce a single, 2-sample mean estimate of the Poisson process. For illustrative purposes, we will continue to identify the endpoint case separately from the binomial midpoint algorithm when $L > 0$, but recall that it is simply a binomial midpoint algorithm with $L = 0$. This sampling is repeated to form an ensemble of such mean estimators, and the ensemble is then used to construct an estimate of the integrated MSE for each algorithm. This process is repeated for a wide range of τ_s, and the results are plotted as follows. First, we examine the endpoint mean estimator in order to verify both the exact analytical expression proven in (10) and the asymptotic bounds given by (13) and (12). The results are collected in Fig. 6. Next, we compare the performance of each of the proposed algorithms with each other and examine how they vary with τ_s for fixed τ_f. These results are collected in Fig. 7.

It is important to note here that the operating points of each of these methods and values of τ_s correspond to different computational costs, which we will define in this work as the expected number of random variable draws required to simulate a path. For example, suppose we are simulating unit-rate Poisson processes over the interval $[0, \tau_f]$ using the endpoint method ($L = 0$) with step size τ_s. On average, we will draw roughly τ_f/τ_s antithetic pairs of Poisson random variables for each step we take in the interval. Then, we will sample approximately τ_f random uniform jump times to simulate a path. The details of these costs are sensitive to the many optimizations that are possible for a particular implementation of continuous-time antithetic stochastic

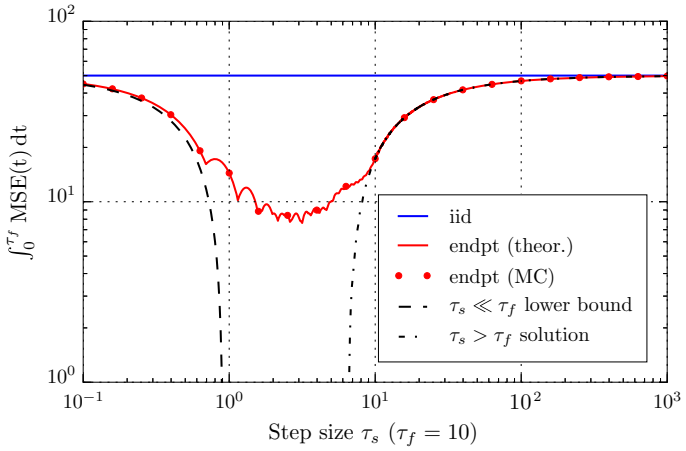

Fig. 6 Exact analytical solution for the integrated scaled MSE of the endpoint estimator (10) versus step size τ_s, compared to empirical observation of the same. The right lower bound is defined in (12), and the left lower bound is given by (13). Empirical results obtained using Monte Carlo simulation using ensemble sizes of 1 440 000 samples or more. Error bars are very small and are thus omitted

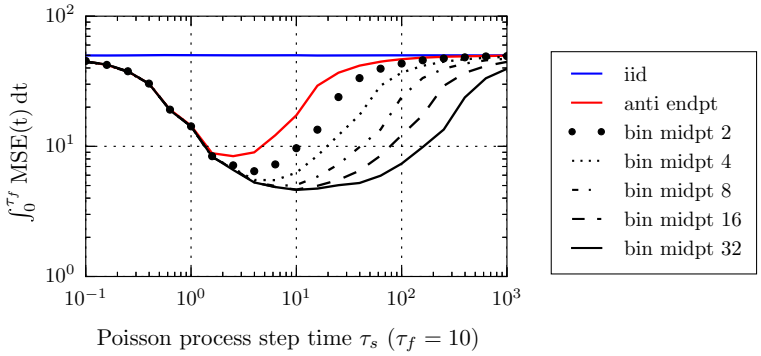

Fig. 7 Integrated scaled variance versus step size τ_s for mean estimators produced using iid, endpoint and binomial midpoint simulation. Results obtained using Monte Carlo simulation using ensemble sizes of 360,000 samples or more. Error bars are very small and are thus omitted

simulation. For the purpose of this work, we focus on a crude implementation: steps of length τ_s are taken until the final time τ_f is strictly exceeded. For each step taken, every corresponding uniform jump time is simulated, including those lying outside $[0, \tau_f]$. Thus, we will tend to incur significant overhead relative to MSE reduction when $\tau_s \gg \tau_f$ or when $\tau_s \ll \tau_f$ compared to existing methods such as SSA or next-reaction.

For this particular implementation, we thus estimate the number of random draws necessary for an antithetic method with order L using the expression $2^L\,(\lceil \tau_f/\tau_s \rceil \vee 1) + \lceil \tau_f/\tau_s \rceil \tau_s \vee \tau_s$, where $\vee$ denotes the maximum operator. The lower limit for this cost is τ_f, achieved, for example, using various next-reaction methods (Gibson and Bruck 2000), so we will use this as our baseline for comparison. We will restrict our attention to τ_s values that lie in the Pareto front, the region of values for which error and cost

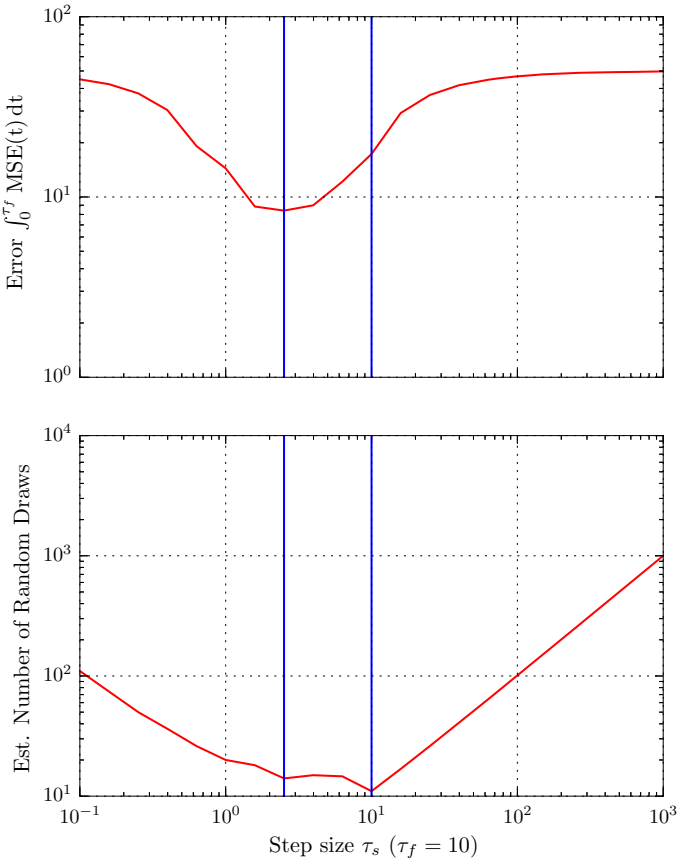

Fig. 8 Integrated scaled variance (error) and estimated number of random draws (cost), each plotted versus step size τ_s for a mean estimator produced using endpoint simulation. The τ_s values shown in blue are the Pareto front: the set of values for which error (integrated MSE) and cost (expected number of random draws) cannot be simultaneously improved. The Pareto region was estimated numerically by sampling τ_s values and selecting those values which were Pareto optimal with respect to the sampled set. In this case, the Pareto front is composed of two points. Results obtained using Monte Carlo simulation using ensemble sizes of 360,000 samples or more. Error bars are very small and are thus omitted

cannot be simultaneously improved. The Pareto region was estimated numerically by sampling τ_s values and selecting those values which were Pareto optimal with respect to the sampled set. To illustrate this region, consider Fig. 8, where we plot the MSE results for the antithetic endpoint method that appear in Fig. 7 along with the estimated cost of simulation versus τ_s.

Restricting our attention only to Pareto values of τ_s, we may plot the error vs cost as shown in Fig. 9. As discussed above, the antithetic endpoint ($L = 0$) or binomial midpoint ($L = 1, 2, 3$) methods can offer significant performance improvement at relatively modest cost increases. These gains saturate relative to cost for larger L values (e.g., $L = 5$). Thus, we may conclude that, in practice, τ_s should be tuned to the native speed of the process (relative to the time window τ_f that we are studying) and

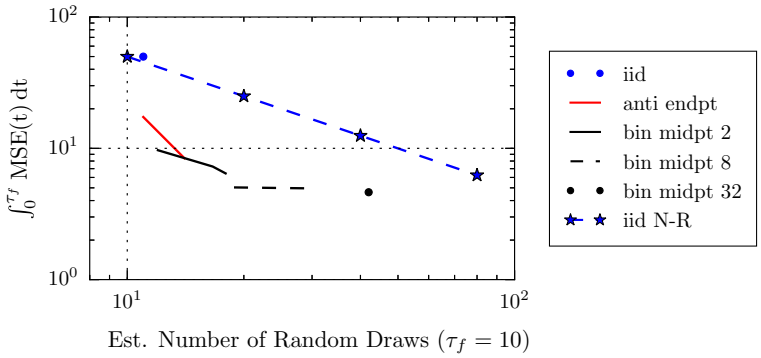

Fig. 9 Integrated scaled variance (error) versus estimated number of random draws (cost), for various estimators constructed using iid, endpoint and binomial midpoint Poisson process simulation. Only points corresponding to τ_s values in the Pareto region are shown. The Pareto region was estimated numerically by sampling τ_s values and selecting those values which were Pareto optimal with respect to the sampled set. The cost baseline for iid simulation is a next-reaction algorithm that simulates trajectories using no excess random variable draws. For cost comparison, we plot the unscaled, integrated MSE for the next-reaction estimator as the number of iid sample paths used in the N-R estimate is repeatedly doubled. Results obtained using Monte Carlo simulation using ensemble sizes of 360,000 samples or more. Error bars are very small and are thus omitted

that significant performance gains can be achieved using binomial midpoint sample with moderate order L.

3 Antithetic Simulation of Lattice CTMCs

We can employ negatively correlated pairs of unit-rate Poisson processes (as simulated using Algorithms 1 and 2 shown in Sect. 2) to simulate negatively correlated pairs of lattice continuous-time Markov chains (CTMCs). We define this anticorrelated pair of stochastic processes $(X^{(1)}, X^{(2)})$ as follows:

$$X^{(j)}(t) = x_0 + \sum_{i=1}^{I} Y^{i,j}\left(\int_0^t a^i(s, X^{(j)}(s))\,\mathrm{d}s\right)\nu^i, \tag{14}$$

for $j \in \{1, 2\}$. In other words, to simulate a pair of trajectories of a lattice CTMC system with I reaction channels, we simulate I antithetic pairs of unit-rate Poisson processes and assign one element of each pair to a reaction channel in each path $X^{(j)}$.

Note that, to simulate these CTMCs, we use the fact that each trajectory is piecewise constant while waiting for the next jump to occur. So for each reaction channel, we can use the value of the reaction rate to compute the time until the next transition occurs for that Poisson process. The smallest of these times is the one that will occur first, so we may move each process forward until this event occurs, update the state of the system and repeat. Thus, we can simulate a process trajectory using only the ordered jump times $\mathcal{A}$ of I unit-rate Poisson processes.

By construction, each stochastic process path is simulated using I iid, unit-rate Poisson processes, and the exact marginal distribution of the system is preserved. The only difference is that the pair of lattice CTMC paths are now negatively correlated, and will produce variance-reduced mean estimates $\delta = \frac{X^{(1)}+X^{(2)}}{2}$. To quantify this reduction in variance, we define the scaled stochastic process MSE to be

$$\mathrm{MSE}(t) = 2\,\mathrm{tr}\,\mathrm{Var}\left(\delta(t)\right). \tag{15}$$

As discussed above, the scaled MSE is insensitive to the inclusion of additional anti-correlated pairs of paths, so we need only consider estimators constructed from a single pair of stochastic processes for the subsequent studies. As above, we use the integrated scaled MSE to quantify estimator performance. Note that in this case, the time over which we will integrate is the physical time t, and no longer the "internal times" of the Poisson processes. We use the iid mean estimator as a baseline, as its MSE does not depend on the choice of τ_s. In both numerical studies, we construct an ensemble of 2-sample mean estimators each built from a single pair of system trajectories simulated using either iid, endpoint or binomial midpoint simulation for a given value of τ_s. We then use this ensemble to estimate the integrated MSE of each of these estimators, and let τ_s vary over a large range of values to examine the dependence of estimator MSE. In both systems, we will see that a similar relationship between MSE and τ_s holds as in the Poisson process case, save that the artifact τ_f determining the time window of interest is now replaced by the interaction between T, the final time of simulation, and the reaction rates and particular trajectory of the system.

We now introduce two example systems to illustrate the performance of the anti-correlated Monte Carlo for stochastic process paths using the RTC (1), driven by the above algorithm for generating antithetic Poisson process pairs. The first is a gene expression system with rates that are an affine function of the system state, and the second is an aerosol coagulation system driven by rates that are a nonlinear function of the system state. In both cases, the endpoint and binomial midpoint algorithms are used to generate the unit-rate Poisson processes $\{(Y^{i,1}, Y^{i,2})\}_{i=1}^{I}$ that are the sole source of random input to the models.

3.1 Gene Expression

First, we examine a linear gene expression system. The system has two components: mRNA that is produced and decays, and a protein it produces which also decays. This particular model appears in Briat and Khammash (2012), and its reactions are given by

$$\begin{aligned} \emptyset &\xrightarrow{k_r} \mathrm{mRNA} \\ \mathrm{mRNA} &\xrightarrow{\gamma_r} \emptyset \\ \mathrm{mRNA} &\xrightarrow{k_p} \mathrm{protein} + \mathrm{mRNA} \\ \mathrm{protein} &\xrightarrow{\gamma_p} \emptyset. \end{aligned}$$

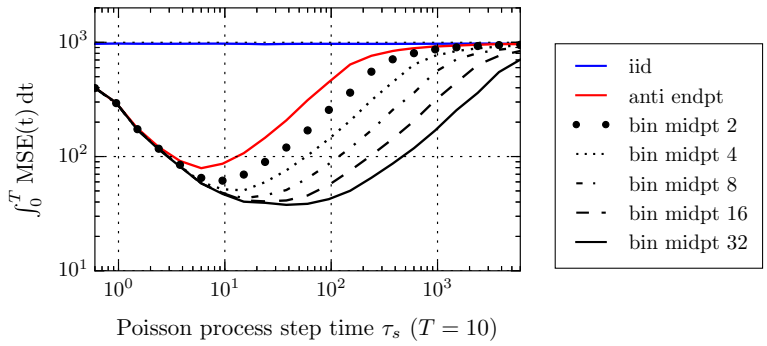

Fig. 10 Integrated scaled MSE versus step size τ_s for mean estimators of the gene expression system produced using iid, endpoint and binomial midpoint simulation of unit-rate Poisson processes. The system is simulated with volume parameter $V = 100$, rate parameters $(k_r, \gamma_r, k_p, \gamma_p) = (0.01, 0.03, 0.06, 0.0066)$, and initial condition $X_0 = [V\ V/2]^\top$. MSE estimates are obtained using Monte Carlo simulation using ensemble sizes of 360,000 samples or more. Error bars are very small and are thus omitted

The system state is a vector $X \in \mathbb{Z}^2$ whose components represent the number of mRNA and protein particles. We set the initial condition $x_0 = V \cdot [1.0\ 0.5]^\top$ (where V is a system volume scaling parameter, fixed here at $V = 100$) with $I = 4$ reaction channels, given by:

$$\begin{aligned} \nu^1 &= [1\ 0]^\top & a^1(X(t)) &= k_r V \\ \nu^2 &= [-1\ 0]^\top & a^2(X(t)) &= \gamma_r X_1(t) \\ \nu^3 &= [0\ 1]^\top & a^3(X(t)) &= k_p X_1(t) \\ \nu^4 &= [0\ -1]^\top & a^4(X(t)) &= \gamma_p X_2(t), \end{aligned}$$

for $k_r, \gamma_r, k_p, \gamma_p > 0$, and where $X_d(t)$ denotes the dth component of the state vector at time t. The system is simulated using the random-time-change representation (1) run from time $t = 0$ to time $t = T = 10$. Our primary interest is the dependence of the integrated scaled MSE for the gene expression estimator on the choice of step size τ_s for the Poisson process trajectories. The results of this study are shown in Fig. 10.

Again, it is instructive to compare the different implementations on the basis of cost (as measured by estimated number of random variable simulations). Restricting our attention to only Pareto-optimal points, we obtain the error vs cost relationship shown in Fig. 11.

3.2 Nonlinear Aerosol Coagulation Due to Gravitational Settling

Finally, we examine the MSE of a nonlinear lattice CTMC when we apply antithetic simulation to its driving Poisson processes. We consider a water aerosol system subject to gravitational settling that undergoes coagulation events as it falls. This system can be found in Seinfeld and Pandis (2006), Chapter 13, and the underlying assumptions and construction of the model used here are discussed in some detail in Maginnis

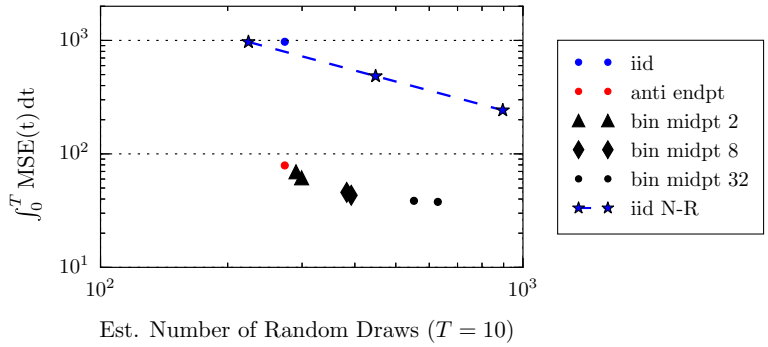

Fig. 11 Integrated scaled variance (error) versus estimated number of random draws (cost), for various estimators of the gene expression system constructed using iid, endpoint and binomial midpoint Poisson process simulation. The cost baseline for iid simulation is a next-reaction algorithm that simulates trajectories using no excess random variable draws. For cost comparison, we plot the unscaled, integrated MSE for the next-reaction estimator as the number of iid sample paths used in the N-R estimate is repeatedly doubled. The system is simulated with volume parameter $V = 100$, rate parameters $(k_r, \gamma_r, k_p, \gamma_p) = (0.01, 0.03, 0.06, 0.0066)$, and initial condition $X_0 = [V\ V/2]^\top$. Only points corresponding to τ_s values in the Pareto region are shown. Results obtained using Monte Carlo simulation using ensemble sizes of 360,000 samples or more. Error bars are very small and are thus omitted

et al. (2016). For the sake of brevity, we omit those details here. The system is composed of large and small water particles falling in a control volume. These classes of particles have different terminal velocities and thus may experience collisions as they fall leading to coagulation events. We fix the velocity of the control volume to be the same velocity as the large particles, so that small particles may enter the system and also coagulate with large particles. The state of the system can be expressed as $X = (N_s, M_l) \in \mathbb{R}^2$, where N_s denotes the number of small particles and M_l is the total mass of the large particles. For convenience, we set the mass of the small particles to be $m = 1$, and the reaction channels and rates of the system are given by:

$$\begin{aligned} \nu^1 &= [1\ 0]^\top & a^1(X(t)) &= V \\ \nu^2 &= [-1\ 1]^\top & a^2(X(t)) &= \alpha K_{sl}^{\mathrm{GS}} V X_1(t) \end{aligned}$$

where $\alpha = 5 \cdot 10^{-4}$ is a proportionality constant and

$$K_{sl}^{\mathrm{GS}} = \frac{1}{V}\left(\sqrt[3]{X_2(t)/V} + \sqrt[3]{m}\right)^3 \left(\sqrt[3]{X_2(t)/V} - \sqrt[3]{m}\right).$$

The state is initialized from $X_0 = V \cdot [100, 10]$ and is simulated for 10s from $t = 0$ to $T = 10$. The integrated scaled MSE of the coagulation system plotted versus the Poisson process step size τ_s is shown in Fig. 12.

As above, the cost-error tradeoff is visualized by restricting attention to Pareto-optimal points and is shown in Fig. 13.

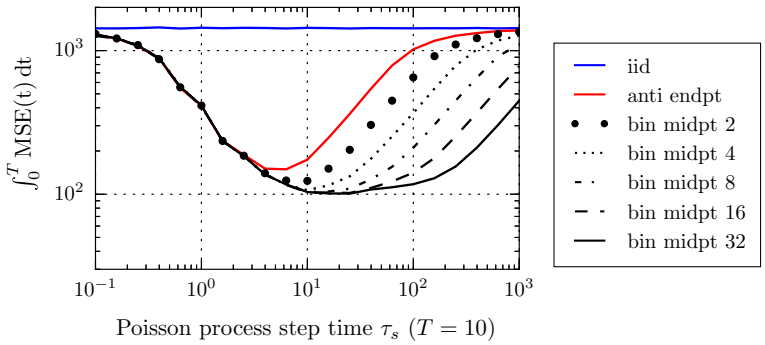

Fig. 12 Integrated scaled MSE versus step size τ_s for mean estimators of the nonlinear aerosol coagulation system produced using iid, endpoint and binomial midpoint simulation of unit-rate Poisson processes. We take volume parameter $V = 100$, proportionality constant $\alpha = 5 \cdot 10^{-4}$, and small particle mass $m = 1$. Results obtained via Monte Carlo simulation using ensemble sizes of 360 000 samples or more. Error bars are small and are thus omitted

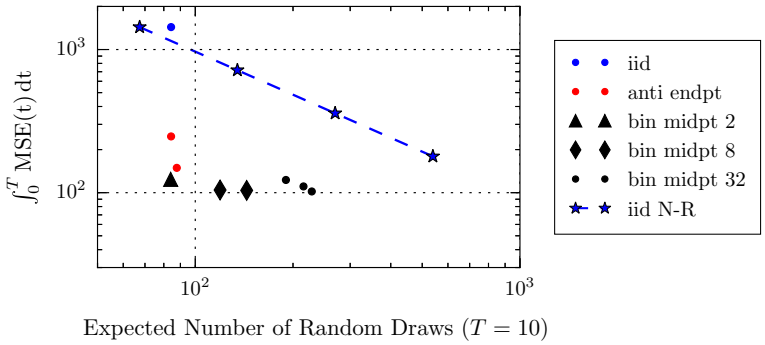

Fig. 13 Integrated scaled variance (error) versus estimated number of random draws (cost), for various estimators of the nonlinear coagulation system constructed using iid, endpoint and binomial midpoint Poisson process simulation. The cost baseline for iid simulation is a next-reaction algorithm that simulates trajectories using no excess random variable draws. For cost comparison, we plot the unscaled, integrated MSE for the next-reaction estimator as the number of iid sample paths used in the N-R estimate is repeatedly doubled. We take volume parameter $V = 100$, proportionality constant $\alpha = 5 \cdot 10^{-4}$, and small particle mass $m = 1$. Only points corresponding to τ_s values in the Pareto region are shown. Results obtained using Monte Carlo simulation using ensemble sizes of 360 000 samples or more. Error bars are small and are thus omitted

4 Conclusions

In this paper, we introduced a new algorithm to simulate lattice CTMCs that provides order-of-magnitude speedup versus traditional iid Monte Carlo methods, while simultaneously introducing no bias error and remaining cost-competitive. We achieved these improvements by producing negatively correlated, identically distributed sample trajectories to produce unbiased mean estimators for both linear and nonlinear stochastic systems. We produced these paths by first showing how antithetic, unbiased unit-rate Poisson process paths could be simulated and used as the random input for time-change representations of lattice CTMCs. Further, we provided both theoretical

and numerical analysis of this Poisson process algorithm that exactly characterize its performance and provide simple asymptotic bounds for large and small leap lengths τ_s.

Acknowledgements Funding was provided by Air Force Office of Scientific Research (Grant No. FA9550-15-1-0059).

References

Anderson DF (2007) A modified next reaction method for simulating chemical systems with time dependent propensities and delays. J Chem Phys 127(21):214107

Anderson DF (2012) An efficient finite difference method for parameter sensitivities of continuous time Markov chains. SIAM J Numer Anal 50(5):2237–2258

Anderson DF, Higham DH (2012) Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics. Multiscale Model Simul 10(1):146–179

Anderson DF, Kurtz TG (2015) Stochastic analysis of biochemical systems. Springer, Berlin

Banks HT, Hu S, Joyner M, Broido A, Canter B, Gayvert K, Link K (2012) A comparison of computational efficiencies of stochastic algorithms in terms of two infection models. Math Biosci Eng 9(3):487–526

Bratsun D, Volfson D, Tsimring LS, Hasty J (2005) Delay-induced stochastic oscillations in gene regulation. PNAS 102(41):14593–14598. https://doi.org/10.1073/pnas.0503858102

Briat C, Khammash M (2012) Computer control of gene expression: robust setpoint tracking of protein mean and variance using integral feedback. In: Proceedings of the IEEE conference on decision and control

Cai X (2007) Exact stochastic simulation of coupled chemical reactions with delays. J Chem Phys 126(12):124108. https://doi.org/10.1063/1.2710253

Cao Y, Gillespie DT, Petzold LR (2005) The slow-scale stochastic simulation algorithm. J Chem Phys 122(1):014116. https://doi.org/10.1063/1.1824902

Ethier SN, Kurtz TG (1986) Markov processes: characterization and convergence. Wiley, Hoboken

Gibson M, Bruck J (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels. J Phys Chem A 105:1876

Giles MB (2008) Multilevel Monte Carlo path simulation. Oper Res 56:607–617

Gillespie DT (1975) An exact method for numerically simulating the stochastic coalescence process in a cloud. J Atmos Sci 32:1977–1989

Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J Comput Phys 22:403–434

Gillespie DT (2001) Approximate accelerated stochastic simulation of chemically reacting systems. J Chem Phys 115(4):1716–1733

Gillespie DT, Hellander A, Petzold LR (2013) Perspective: Stochastic algorithms for chemical kinetics. J Chem Phys 138(17):170901

Glynn PW, Iglehart DL (1989) Importance sampling for stochastic simulation. Manag Sci 35:1367–1392. https://doi.org/10.1287/mnsc.35.11.1367

Maginnis PA (2011) Variance reduction for Poisson and Markov jump processes. Master's thesis, University of Illinois at Urbana-Champaign

Maginnis PA, West M, Dullerud GE (2016) Variance-reduced simulation of lattice discrete-time Markov chains with applications in reaction networks. J Comput Phys 322:400–414. https://doi.org/10.1016/j.jcp.2016.06.019

McAdams HH, Arkin A (1997) Stochastic mechanisms in gene expression. Proc Natl Acad Sci 94(3):814–819

Rathinam M, Sheppard PW, Khammash M (2010) Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks. J Chem Phys 132(3):034103

Riemer N, West M, Zaveri RA, Easter RC (2009) Simulating the evolution of soot mixing state with a particle-resolved aerosol model. J Geophys Res 114:D09202. https://doi.org/10.1029/2008JD011073

Robert CP, Casella G (2004) Monte Carlo statistical methods, 2nd edn. Springer, Berlin

Seinfeld JH, Pandis SN (2006) Atmospheric chemistry and physics. Wiley, Hoboken

Villen-Altamirano M (2012) Rare event simulation: The RESTART methods, In: Proceedings of the international conference on high performance computing and simulation, pp 32–41
Whitt W (1976) Bivariate distributions with given marginals. Ann Stat 4(6):1280–1289
Wilkinson DJ (2011) Stochastic modelling for systems biology. CRC Press, Boca Raton