Scalable Bayesian Inference for Population Markov Jump Processes

Iker Perez and Theodore Kypraios
School of Mathematical Sciences, University of Nottingham

Abstract
Bayesian inference for Markov jump processes (MJPs) where available observations relate to either system states or jumps typically relies on data-augmentation Markov Chain Monte Carlo. State-of-the-art developments involve representing MJP paths with auxiliary candidate jump times that are later thinned. However, these algorithms are i) unfeasible in situations involving large or infinite capacity systems and ii) not amenable for all observation types. In this paper we establish and present a general data-augmentation framework for population MJPs based on uniformized representations of the underlying non-stationary jump processes. This leads to multiple novel MCMC samplers which enable exact (in the Monte Carlo sense) inference tasks for model parameters. We show that proposed samplers outperform existing popular approaches, and offer substantial efficiency gains in applications to partially observed stochastic epidemics, immigration processes and predator-prey dynamical systems.

1 Introduction
Population Markov jump processes (MJPs) are stochastic processes whose dynamics underpin many observable phenomena, in diverse fields such as stochastic epidemic (O’Neill and Roberts, 1999), immigration-death systems (Cappé et al., 2003; Zhang and Rao, 2018), chemical/molecular models (Hobolth and Stone, 2009; Georgoulas et al., 2017) or queueing systems (Sutton and Jordan, 2011; Perez et al., 2018), to name only a few. In this paper, we present a novel and general framework for designing scalable auxiliary-variable data-augmentation algorithms, which allow for exact (in Monte Carlo sense) Bayesian inference for MJPs. In contrast to the state-of-the-art method of Rao and Teh (2013), our framework provides us with a class of Markov chain Monte Carlo (MCMC) algorithms that are amenable to all relevant application fields, where observations may consist of either population counts or jumps. Furthermore, we present efficient algorithmic designs that address augmentation and inferential tasks in systems where the population is large. We demonstrate this by reporting on substantial efficiency and scalability gains, in application to partially-observed birth-death, (stochastic) epidemic and predator-prey models. This general framework is built over uniformized representations of non-stationary jump processes (cf. Jensen, 1953; Van Dijk, 1992), and we show that algorithms presented in Rao and Teh (2012, 2013) are derived as special cases.

1.1 Jump processes
An MJP is a pure-jump right-continuous stochastic process $X = (X_t)_{t \geq 0}$, such that time-indexed variables $X_t$ are defined within some measurable space $(\mathcal{S}, \Sigma_{\mathcal{S}})$. Here, $\mathcal{S}$ is a countable set of states, and $\Sigma_{\mathcal{S}}$ stands for its power set. The process $X$ is governed by a sequence of intensity matrices $Q = \{Q(t) : t \geq 0\}$, indexed over time, so that

$$
P(X_{t+dt} = x'|X_t = x) = \mathbb{1}_{(x=x')} + Q_{x,x'}(t)dt + o(dt)
$$
for all \( x, x' \in \mathcal{S} \) and \( t \geq 0 \); where \( \mathbb{I}_{\{\cdots\}} \) defines a logical indicator function. Hence, elements of \( Q(t) \) describe the rates for jumps between states at time \( t \geq 0 \), and each \( Q_{x,x'}(t) \), \( x, x' \in \mathcal{S} \), is an intensity function over time. Finally, \( 0 \leq Q_{x,x'}(t) < \infty \) for all \( x \neq x' \) and \( Q_x(t) := Q_{x,x}(t) = -\sum_{x' \in \mathcal{S} \setminus \{x\}} Q_{x,x'}(t) \) so that the various rows sum to 0.

**Piecewise-constant representation.** An MJP is further characterized by a path or trajectory \((t, x)\), where \( t = \{t_0, \ldots, t_n\} \) denotes a sequence of transition times, s.t. \( t_0 = 0 \), and \( x = \{x_0, \ldots, x_n\} \) are the corresponding states. Over a time interval \([0, T]\), a process \( X \equiv (t, x) \) is a random variable on a measurable space \((\mathcal{X}, \Sigma_X)\) supporting finite \( \mathcal{S}\)-valued trajectories. On a basic level, \( \mathcal{X} = \bigcup_{n=0}^{\infty} ([0, T] \times \mathcal{S})^n \) and the collection \( \Sigma_X \) stands for the corresponding union \( \sigma\)-algebra. This allows the assignment of a dominating base measure \( \mu_X \) w.r.t which define a trajectory density

\[
f_X(t, x | Q) = \pi(x_0) e^{\int_{t_0}^{t} Q_{x,s}(s)ds} \prod_{i=1}^{n} Q_{x_{i-1}, x_i}(t_i) e^{\int_{t_{i-1}}^{t_i} Q_{x_i-1, s}(s)ds},
\]

(1)

where \( \pi(\cdot) \) is the distribution assigned (over \( \mathcal{S} \)) to the starting value. Noticeably, the time to departure or jump from any state \( x \in \mathcal{S} \), regardless of the destination, is driven by a density

\[
f_{t_{i+1}}(t | x_i, t_i) = Q_{x_i}(t) e^{-\int_{t_i}^{t} Q_{x_i}(s)ds}, \quad t > t_i, \ i = 1, \ldots, n-1.
\]

Thus, inter-arrival times in \( t \) are linked to diagonal elements of \( \{Q(t) : t \geq 0\} \), and \( Q_\cdot(s), s \geq 0 \) is often referred to as the hazard function to the origin state \((t, x) \in [0, T] \times \mathcal{S} \). Finally, transitions in \( x \) are proportional to off-diagonal elements, s.t. \( \mathbb{P}(x_{i+1} = x | x_i, t_i, t_{i+1}) = Q_{x_i,x}(t_{i+1})/Q_{x_i}(t_{i+1}) \).

For details, we refer the reader to Daley and Vere-Jones (2007).

**Stationary models.** If \( X \) is assumed to be a time-homogeneous process, ignoring seasonal effects and thus governed by a generator matrix \( Q(t) \equiv Q, t \geq 0 \); then, inter-arrival times in \( t \) are exponentially distributed random variables and \((t, x)\) is a (Markov) renewal process.

**1.2 Population models and Bayesian inferential tasks**

Throughout this paper, a Markov population model is represented by a non-stationary MJP whose support space \( \mathcal{S} \) is countable and possibly infinite. Matrices \( Q(t), t \geq 0 \) are assumed to be sparse and parametrized by some arbitrary vector of independent rates \( \lambda \), which scale along with levels of populations in \( X \). An upper-bound over a sequence of matrices \( Q \equiv Q(\lambda) \) may take extraordinarily large values.

**Bayesian inferential task.** Let \( O = \{O_r\}_{r \geq 1} \) denote some observations at arbitrary (ordered) time points \( t_r \in [0, T], r \geq 1 \), which relate to a population model realization \( X \) with unknown matrices \( Q(\lambda) \). The basis for inference on the (unknown) vector \( \lambda \) is a density or mass \( \mathcal{L}(O | X) \) for the observation model; and posterior rate densities are proportional to an infinite weighted product of MJP path densities \( X \) in (1), i.e.

\[
f_Q(\lambda | O) \propto f_Q(\lambda) \cdot \int_X \mathcal{L}(O | t, x) f_X(t, x | Q(\lambda)) \mu_X(dt, dx),
\]

(2)

where \( f_Q(\lambda) \) defines a prior over the rates. This is an analytically, and often computationally, intractable expression. It is hard to design a generic framework to perform exact Monte Carlo inference, yet remain adaptable to any type of jump process \( X \) and observation model \( \mathcal{L}(O | X) \). Consequently, many solutions either focus on approximate inferential methods, or are limited to homogeneous systems and address constrained biological models where population measurements must be subject to observation noise. Such approaches can lead to computationally efficient methods by relying on simplifying independence assumptions Opper and Sanguinetti (2008), diffusions with continuous support Golightly and Wilkinson (2015) or linear noise approximations Golightly and Sherlock (2018).
1.3 Exact inference and Markov chain Monte Carlo

Exact inference often proceeds by MCMC, and alternates sampling between the latent process \((t, x)\) and rates \(\lambda\). Thus, it is concerned with the joint density \(f_{Q,X}(\lambda, t, x|O)\), and entails data augmentation procedures from a conditional

\[
f_X(t, x|\lambda, O) \propto \mathcal{L}(O|t, x)f_X(t, x|Q(\lambda)),
\]

which may take multiple forms based on observation model dependencies for \(O|t, x\). In every instance, sampling a trajectory from (3) brings about substantial tractability challenges; there can exist infinitely many jumps and we require to explore transitions across large or infinite subsets of \(S\).

To allow for a generic adaptable algorithm design, sampling commonly proceeds by means of blocked (Poisson) thinning procedures, in summary

- a set of candidate jump times \(\hat{t}|O, Q(\lambda)\) is first produced, with some conditional intensity process \(\Omega(t), t \geq 0\), and s.t. every \(\Omega(t)\) dominates all diagonal elements of \(Q(t)\),
- an augmented sequence \(\hat{x}|t, O, Q(\lambda)\) is sampled from an appropriate forward-backward algorithm; this must allow for self transitions and thus thin a portion of candidate jump times.

Within time-homogeneous jump systems, such procedures may be supported on matrix exponential representations for transition probabilities (see Fearnhead and Sherlock, 2006) or, ideally, built over uniformization alternatives and the seminal contributions of Hobolth and Stone (2009); Rao and Teh (2013). In broader settings parametrized by hazard functions, dependent thinning alternatives (see Rao and Teh, 2012; Miasojedow et al., 2017) offer the only computationally feasible approach. Overall, data augmentation procedures in all the above instances are rigid, designed with small MJP systems in mind and only accommodate restrictive observation models suitable to few applications. Importantly, they often do not work (or do not scale) for the analysis of population models, where transition rates scale quadratically through interactions of marginal counts, observations are often a consequence of system jumps and unbounded populations are the norm.

Recent developments. Current alternatives sit on top of the aforementioned benchmark algorithms, and are limited to addressing considerations of state-space explosions for stationary systems. In order to preserve asymptotic exactness, without imposing artificial bounds on population levels, sequential particle procedures may be used to target sequences of states in \(\hat{x}\) (Miasojedow and Niemiro, 2015) (subject to particle degeneracy), or arbitrary random truncations imposed over explorable spaces of paths (Georgoulas et al., 2017) (requiring costly Metropolis-Hastings (M-H) acceptance steps to overcome induced bias). Most recent advances towards efficient algorithmic constructions (see Zhang and Rao, 2018) involve updating parameters \(\lambda\) within forward-backward procedures for \((\hat{t}, \hat{x})\), which works reportedly well with small MJP systems.

1.4 Summary of contributions

In this work, we present a novel auxiliary-variable framework leading to data-augmentation techniques for conditional population model trajectories \((t, x)\) in (3). This will yield to computationally tractable joint distributions across both target and auxiliary variables, and readily lead to Gibbs-like procedures satisfying detailed balance (see Higdon, 1998). Hence, we further construct new MCMC samplers adaptable to popular Bayesian inferential tasks; in Figure 1 we summarize efficiency results that compare these to existing benchmark methods, in application to birth-death processes (left), stochastic epidemics (centre) and predator-prey (right) dynamics. The lines represent ratios in effective sample sizes across unknown model parameters, tested at several population capacities specified by the horizontal axis. In each case, ratios are measured against a suitably chosen benchmark (horizontal line at level 1), and include confidence intervals through repetition over several datasets.
Coloured lines correspond to samplers introduced in this paper; dark lines represent existing state of the art alternatives. In all cases, we note significant gains in scalability and efficiency.

![Image](image1.png)

Figure 1: Ratios in effective sample sizes (with confidence intervals) across model parameters, for inferential tasks with birth-death (left), epidemic (centre) and predator-prey (right) systems. The horizontal axes represent population sizes tested. In each case, ratios are measured against a suitably chosen benchmark (horizontal line at level 1). Coloured lines correspond to techniques in this paper; dark lines represent state-of-the-art methods. We notice significant advances in system scalability (existing approaches are often unusable with large populations) and reasonable increments in efficiency in all cases.

Within the rest of the paper, Section 3 introduces a two-step data-augmentation with random importance weightings; and further describes (i) performance optimization with stationary MJPs and (ii) limiting properties to population systems with infinite capacities. Also, the section draws comparisons and discusses differences with existing uniformization-based methods, and addresses inference with (i) deterministic/random observations of population states, and (ii) observations of population jumps. Section 4 presents auxiliary-variable results to efficiently sample jump trajectories as deviations from deterministic mean-average population dynamics; thus addressing considerations of state-space explo

sions strictly within Gibbs procedures. Finally, Section 5 studies dividing augmentation procedures into smaller computationally tractable counterparts.

2 Uniformization and auxiliary variables

Let \( \bar{t}, \bar{x} \) define an augmented jump trajectory over the finite time interval \([0, T]\), so that \( \bar{x}_i \in S \) for \( i \geq 0 \). Here, inter-arrival times in \( \bar{t} \) are exponentially distributed with a fixed rate

\[
\Omega \geq \max_{x \in S} \sup_{t \in [0, T]} |Q_x(t)|,
\]

and \( \bar{x} \) is a realization from a discrete-time non-homogeneous Markov chain, with initial state \( x_0 \in S \) drawn from \( \pi(\cdot) \), and transition probability matrices \( P(\bar{t}_i) = I + Q(\bar{t}_i)/\Omega, \bar{t}_i \in \bar{t} \). The process \((\bar{t}, \bar{x})\) describes an augmented MJP (allowing for self-transitions) on \( X \), and it is equivalent to \( X = (t, x) \) with intensity matrices \( Q = \{Q(t) : t \geq 0\} \) and density function (1).

Proposition 2.1. The process \((\bar{t}, \bar{x})\) describes an augmented MJP (allowing for self-transitions) on \( X \), and it is equivalent to \( X = (t, x) \) with intensity matrices \( Q = \{Q(t) : t \geq 0\} \) and density function (1).

This is a well known result; it follows since

\[
P(X_{t+s} = x'|X_t = x, Q) = \sum_{k=0}^{\infty} \frac{s^k \Omega^k}{k!} e^{-\Omega t} \int_{H^k} [P(u_1) \times \cdots \times P(u_k)]_{x,x'} \, dH(u_1, \ldots, u_k)
\]

for all \( x, x' \in S \) and \( 0 < t < t + s < T \) offers a randomized or uniformized representation of transition probabilities across times and states in the original MJP. There, \( dH(u_1, \ldots, u_k) = k! / s^k \cdot du_1 \ldots du_k \)
denotes the density of a \( k \)-dimensional vector of order statistics on \( H^k = \{(u_1, \ldots, u_k) \in [t, t + s]^k : u_1 < u_2 < \cdots < u_k \} \), and a proof of equivalence may be found on e.g. Van Dijk (1992); Van Dijk et al. (2018). Commonly used for simulation in homogeneous systems, the computational procedure that constructs these augmented sets of times \( \hat{t} = \{t_0, \ldots, t_m\} \) and states \( \hat{x} = \{\hat{x}_0, \ldots, \hat{x}_m\} \) offers an efficient alternative to Gillespie’s algorithm (Gillespie, 1977), and is commonly referred to as uniformization (cf. Jensen, 1953). Whenever \( \hat{x}_i = \hat{x}_{i-1} \), we refer to a transition \( i \) as a virtual jump. For example, in Figure 2 (left) we observe an augmented birth-death trajectory; there, we spot virtual jumps at times \( t_2, t_6, t_7 \) and \( t_9 \), which are represented by white circles on the time axis. On the right hand side, we observe the equivalent trajectory after virtual times and states have been removed.

Along with the uniformized trajectory \( \hat{t}, \hat{x} \), let \( u = \{u_i\}_{i=1,\ldots,m} \) define an auxiliary family of random variables, s.t.

\[
\mathbb{P}(u_i^{-1}(A)|\hat{x}) = \int_A g(a|u_{i-1}, \hat{x}_{i-1}, \hat{x}_i, \hat{t}_i) \mu_J(da)
\]

for all \( i = 2, \ldots, m \) and \( A \in \Sigma_J \), with \( \mathbb{P}(u_1^{-1}(A)|\hat{x}) = \int_A g(a|\hat{x}_0, \hat{x}_1, \hat{t}_1) \mu_J(da) \). Here, \( (J, \Sigma_J) \) denotes an arbitrary support space and \( \mu_J \) is its corresponding base measure. We impose that a density/mass \( g(\cdot) \) in (4) must be defined s.t. for any sequence \( u \), along with corresponding holding times in \( \hat{t} \), there must exist multiple probabilistically compatible choices of \( \hat{x} \).

**Definition 2.2.** Let \( u = \{u_i\}_{i=1,\ldots,m} \) with \( u_i \in J \) be a sequence of auxiliary observations at times \( \hat{t} \) with \( 0 \leq \hat{t}_1, \ldots, \hat{t}_m \) \( \leq T \). We refer to a uniformized sequence \( \hat{x} \) as ‘compatible’ with \( u \) given \( \hat{t} \) whenever \( |\hat{x}| = |\hat{t}| \), \( g(u_1|\hat{x}_0, \hat{x}_1, \hat{t}_1) > 0 \) and \( g(u_i|u_{i-1}, \hat{x}_{i-1}, \hat{x}_i, \hat{t}_i) > 0 \) for all \( i = 2, \ldots, m \).

Trivially, a pair \( (\hat{t}, \hat{x}) \) is compatible with \( u \) if a strictly positive mass is assigned by means of \( g(\cdot|\hat{t}, \hat{x}) \) to the auxiliary realization. Conditioned on \( u \), we may restrict or assign importance weights across uniformized trajectories in \( \mathcal{X} \) within resampling procedures. Throughout the paper, the reader will be presented with multiple designs of densities \( g \) in (4), targeted both at general-form population models or specific jump systems in common application domains.

**Augmenting a trajectory through uniformization.** Assume the existence of a fixed, parametrized sequence of matrices \( Q = Q(\lambda) \), a dominating rate \( \Omega \) and an MJP trajectory \( (t, x) \in \mathcal{X} \), s.t. \( t = \{t_0, \ldots, t_n\} \) and \( x = \{x_0, \ldots, x_n\} \). Then, we may sample an augmented pair \( \hat{(t, x)} \) from within the family of uniformized representations equivalent to \( (t, x) \). Marginalised over \( u \), a conditional density for times \( \hat{t} \) is, up to proportionality, given by

\[
f_\hat{t}(\hat{t}_0, \ldots, \hat{t}_m|t, x, \Omega, Q) \propto \prod_{i=1}^n P_{x_{i-1}, x_i}(t_i) \cdot \prod_{i=0}^m \prod_{j=0}^m P_{x_j, x_i}(\hat{t}_j)^{I[\hat{t}_j \in (t_i, t_{i+1})]} \cdot \Omega^m e^{-T \Omega}
\]

\[
\propto \prod_{i=0}^m \prod_{j=0}^m (\Omega + Q_{x_j, x_i}(\hat{t}_j))^{I[\hat{t}_j \in (t_i, t_{i+1})]} \tag{5}
\]
with \( P(t) = I + Q(t)/\Omega, \, \tau_{n+1} = T \) and whenever \( m \geq n \) and \( t \in \{\hat{t}_0, \ldots, \hat{t}_m\} \). This corresponds to adding \textit{virtual} (self-transition) times to the sequence \( t \), by using successive Poisson processes with rates \( \Omega + Q_{\lambda}(t), \, t \in (t_i, t_{i+1}) \), for every \( i = 0, \ldots, n \) (cf. Rao and Teh, 2013).

Next, a \textit{uniformized} sequence of states \( \tilde{x} \) can be deterministically assigned given knowledge of \( \hat{t}, \hat{t}, x, \) and an auxiliary sequence \( u(\tilde{x}) \) sampled from a mass/density \( g(.) \) in (4). These steps correspond to the top left/right diagrams in the birth-death example within Figure 3. There, a trajectory \( (\hat{t}, \hat{t}, x) \) is complemented with virtual jumps (white circles on horizontal axis), states (white circles within trajectory) and auxiliary evidence (blue rectangles on some virtual epochs). In this example, \( u \) represents randomly \textit{locked or clamped} jumps, and will become clearer to the reader soon.

![Figure 3: Sketch of a birth-death sampling iteration with an auxiliary-variable naive procedure. Top left, a reference path \( (\hat{t}, \hat{t}, x) \); top right, augmentation with virtual jumps and states (represented by white circles), and auxiliary evidence \( u \) (blue rectangles). Bottom left, fixed Poisson holding times \( \hat{t} \) for a new trajectory; instances with rectangles are removed in this example. Bottom right, new trajectory sampled from within a compatible space in \( \mathcal{X} \), with a forward-backward procedure conditioned on \( u \); virtual epochs have been removed.](image)

**Resampling a new trajectory according to compatibility rules.** A new trajectory within a restricted space of \( \mathcal{X} \) may be obtained, by sampling a fresh augmented sequence \( \hat{x} | \hat{t}, u, Q(\lambda) \) and removing all virtual entries. To this end, we ought to target the discrete-time representation

\[
f_{\tilde{x}}(\hat{x}_0, \ldots, \hat{x}_m | \hat{t}, u, \lambda) \propto \pi(\hat{x}_0) \cdot g(\hat{x}_0, \hat{x}_1, \hat{t}_1) \cdot \prod_{i=2}^{m} g(u_i | u_{i-1}, \hat{x}_{i-1}, \hat{x}_i, \hat{t}_i) \prod_{i=1}^{m} P_{\hat{x}_{i-1}, \hat{x}_i}(\hat{t}_i), \tag{6}\]

which readily simplifies to forward-backward steps with initial distribution \( \pi(x) \), no importance updates and (non-stochastic) transition weight matrices \( \bar{P}(\hat{t}; u) \), defined s.t.

\[
\bar{P}_{\hat{x}_{i-1}, x}(\hat{t}_i; u) = \mathbb{P}(\hat{x}_i = x | u_i, \hat{x}_{i-1}, u_{i-1}, \hat{t}_i) = g(u_i | u_{i-1}, \hat{x}_{i-1}, \hat{x}_i, \hat{t}_i) \cdot P_{\hat{x}_{i-1}, x}(\hat{t}_i) \tag{7}\]

for all states \( x \in \mathcal{S} \) and epochs \( i = 1, \ldots, m \). This step corresponds to the bottom left/right diagrams in Figure 3. On the left, we see an \textit{empty} frame of Poisson holding times \( \hat{t} \), which defines a random time-discretization of the time interval \([0, T] \). In this example \( u \) is defined so that, whenever a blue rectangle is shown, \( P_{\hat{x}, \hat{x}'}(\hat{t}; u) = 0 \) for all \( x \neq x' \) within \( \mathcal{S} \); hence, these epochs correspond with self-transitions in any newly sampled sequence of states. On the right, we find a new trajectory after a forward-backward pass and discarding all virtual transitions.
3 Efficient augmentation over restricted sets of candidate times

Next, we present novel designs of auxiliary variables associated with reference uniformization-based data-augmentation algorithms, and further highlight the shortcomings of traditional methods for inference with population models (see e.g. Hobolth and Stone, 2009; Rao and Teh, 2013, and references therein). In later sections, we build over these results in order to scale sampling procedures, leading to efficiency results reported in Subsection 1.4.

3.1 Two-step data augmentation

To begin with, let \( u = \{u_i\}_{i=1,...,m} \) in (4) be defined on some set \( J = \{\phi, \bar{\phi}\} \), where \( \phi \) denotes an arbitrary undefined element, and \( \bar{\phi} \) is a complementary locked element. Throughout this section, elements of \( u \) are assumed mutually independent given \( (t, \hat{x}) \). We define a probability mass function for the conditional distribution \( \phi|\hat{x}_{i-1}, \hat{x}, \hat{t}_i \) with respect to a suitable count measure, for all \( (\hat{t}_i, \hat{x}_{i-1}, \hat{x}_i) \in [0, T] \times S^2 \) as follows:

\[
g(\phi|\hat{x}_{i-1}, \hat{x}_i, \hat{t}_i) = \frac{\psi(\hat{t}_i, \hat{x}_i)}{\Omega + Q_{\hat{x}_i}(\hat{t}_i)} \quad \text{and} \quad g(\bar{\phi}|\hat{x}_{i-1}, \hat{x}_i, \hat{t}_i) = 1 - g(\phi|\hat{x}_{i-1}, \hat{x}_i, \hat{t}_i), \quad \text{if} \quad \hat{x}_{i-1} = \hat{x}_i,
\]

with \( g(\phi|\hat{x}_{i-1}, \hat{x}, \hat{t}_i) = 1 \) otherwise. Here, \( \psi : [0, T] \times S \rightarrow \mathbb{R}_+ \) is any operator that assigns real-valued intensities across time to the various states in \( S \), and must satisfy \( \psi(t, x) \in (0, \Omega + Q_x(t)) \) for all \( (t, x) \in [0, T] \times S \).

**Proposition 3.1.** Let \( X = (t, x) \) be an MJP realization with intensity \( Q = \{Q(t) : t \geq 0\} \), s.t. \( t = \{t_0, \ldots, t_n\} \) and \( x = \{x_0, \ldots, x_n\} \). Consider an augmentation procedure for \( X \), to a triplet \((\hat{t}, \hat{x}, u)\), where

- A sequence \( \hat{t} \) augments \( t \) by adding virtual times from two jointly independent Poisson processes,
  - a ‘controlled’ process with rate \( \psi(t, x) > 0 \), \( t \in (t_i, t_{i+1}) \) within intervals of \( t \), and
  - a ‘compensating’ process with rate \( \Omega + Q_{x_i}(t) - \psi(t, x) \), \( t \in (t_i, t_{i+1}), i \geq 0 \).
- An augmented sequence of states \( \hat{x} \) is deterministically assigned given knowledge of \( \hat{t}, t, x \).
- Auxiliary variables \( u \) are deterministically assigned so that
  - \( u_i = \phi \) for all \( i \geq 1 \) where time \( \hat{t}_i \) in \( \hat{t} \) was sampled from the ‘compensating’ Poisson process,
  - \( u_i = \varphi \) otherwise; i.e. either \( \hat{t}_i \) was sampled from the ‘controlled’ process, or \( \hat{t}_i \in t \).

Then, this construction yields an statistically equivalent triplet \((\hat{t}, \hat{x}, u)\), when compared to sampling \((\hat{t}, \hat{x})\) from (5) followed by auxiliary variables \( u \) from (8).

Due to Markovian properties, we only require to test equivalence in density representations for realizations \((\hat{t}, \hat{x}, u)\) restricted to intervals \([t_i, t_{i+1}], i \geq 0 \). This is however straightforward, by noting that \( \phi \) can only be sampled from the ‘compensating’ Poisson process in the newly described data augmentation procedure; thus, the proof is omitted. At a basic level, we note that locked elements \( \phi \) in \( u \) may only be a consequence of virtual transitions within any uniformized trajectory representation \((\hat{t}, \hat{x}) \in \mathcal{X})\; in Figure 3, these correspond to the blue rectangles on the top-right diagram.

In view of (8), note that forward filtering steps for a sequence \( \hat{x} \) in (6), conditioned on \( u \), reduce
to
\[
\mathbb{P}(\hat{x}_i = x|u_1, \ldots, u_{i-1}, u_i = \bar{\phi}; \hat{t}) = \sum_{x' \in \mathcal{S}} \mathbb{P}(\hat{x}_i = x, \hat{x}_{i-1} = x'|u_1, \ldots, u_{i-1}, u_i = \bar{\phi}; \hat{t})
\]
\[
\propto g(\bar{\phi}|x, x, \hat{t}_i) \cdot P_{x,x}(\hat{t}_i) \cdot \mathbb{P}(\hat{x}_{i-1} = x|u_1, \ldots, u_{i-1}; \hat{t})
\]
\[
\propto \left(1 + \frac{Q_x(\hat{t}_i) - \psi(\hat{t}_i, x)}{\Omega}\right) \cdot \mathbb{P}(\hat{x}_{i-1} = x|u_1, \ldots, u_{i-1}; \hat{t}),
\]
for all \(x \in \mathcal{S}\) and whenever \(u_i = \bar{\phi}, i \geq 1\); i.e. there exists a direct probabilistic correspondence across states over \textit{locked} time epochs. Moreover, if \(u_j = \bar{\phi}\) for subsequent \(j = i + 1, \ldots, i + k\), it follows
\[
\mathbb{P}(\hat{x}_{i+k} = x|u_1, \ldots, u_i, u_{i+1} = \bar{\phi}, \ldots, u_{i+k} = \bar{\phi}; \hat{t}) \propto \prod_{j=1}^{k} \left(1 + \frac{Q_x(\hat{t}_{i+j}) - \psi(\hat{t}_{i+j}, x)}{\Omega}\right) \cdot \mathbb{P}(\hat{x}_i = x|u_1, \ldots, u_i; \hat{t}),
\]
for all \(x \in \mathcal{S}\). Similarly, probabilities for \textit{backward sampling} steps, conditioned on \(u\), are given by
\[
\mathbb{P}(\hat{x}_i = x|\hat{x}_{i+1}, u_{i+1} = \bar{\phi}; \hat{t}) = \mathbb{I}(x = \hat{x}_{i+1}), \quad \text{for all } x \in \mathcal{S}, \text{ and } i \geq 0 \text{ s.t. } u_{i+1} = \bar{\phi},
\]
and are thus deterministic at times with auxiliary \textit{locked} instances. Trivially, the output of a \textit{uniformized} augmented sequence \(\hat{\mathbf{x}}\) must satisfy \(\hat{x}_i = \hat{x}_{i+1} = \cdots = \hat{x}_{i+k-1} = \hat{x}_{i+k}\) whenever \(u_j = \bar{\phi}\) for all \(j = i + 1, \ldots, i + k\). Hence, jumps in \(\hat{\mathbf{x}}\) are restricted to times with auxiliary \textit{open} instances \(\phi\). Additionally, since these correspond to \(\mathbf{t}\) along with random draws from the \textit{controlled} process in Proposition 3.1, a \textit{(non-stochastic) weight} matrix for transitions in \(\hat{\mathbf{x}}\) follows from (7)-(8), i.e.
\[
\tilde{P}_{\hat{x}_{i-1},x}(\hat{t}_i; u_i = \bar{\phi}) = g(\bar{\phi}|\hat{x}_{i-1}, \hat{x}_i = x, \hat{t}_i) \cdot P_{\hat{x}_{i-1},x}(\hat{t}_i) = \frac{Q_{\hat{x}_{i-1},x}(\hat{t}_i)}{\Omega}
\]
whenever \(x \neq \hat{x}_{i-1}\), and
\[
\tilde{P}_{\hat{x}_{i-1},x}(\hat{t}_i; u_i = \bar{\phi}) = g(\bar{\phi}|\hat{x}_{i-1}, \hat{x}_i = \hat{x}_{i-1}, \hat{t}_i) \cdot P_{x,x}(\hat{t}_i) = \frac{\psi(\hat{t}_i, \hat{x}_{i-1})}{\Omega}
\]
otherwise. In conclusion, owing to (9)-(11) and Proposition 3.1, a sampler of sequentially correlated MJP trajectories \((\mathbf{t}, \mathbf{x})\mid Q(\lambda), \mathbf{O}\), where candidate times \(\mathbf{t}\) are governed by an arbitrary intensity \(\psi(\cdot)\), is formalized in Algorithm 1. There, note that the dominating rate \(\Omega\) within transition matrices \(\tilde{P}(\cdot)\) in (12) is dropped; this corresponds to equations (10)-(11) and fades up to proportionality.

We refer to Algorithm 1 as a \textit{naïve} approach; its purpose is to serve as a starting point. Noticeably, the procedure requires forward-backward steps. It is thus inefficient to sample plain MJP trajectories \(\mathbf{X} \equiv (\mathbf{t}, \mathbf{x}) \in \mathcal{X}\) subject to no observations, in comparison to a generative approach such as Gillespie’s algorithm for stationary systems (Gillespie, 1977). However, Algorithm 1 is readily amendable for conditioning on observations \(\mathbf{O} = \{O_r\}_{r \geq 1}\) commonly encountered in applications. This is because, by assumption, observation models are independent of auxiliary jump events within augmented representations \((\mathbf{t}, \hat{\mathbf{x}})\), and further independent of auxiliary variables \(\mathbf{u}\).

\textbf{Conditioning on system state observations.} In the traditional set-up, \(\mathbf{O} = \{O_r\}_{r \geq 1}\) is a sequence of population level observations at (ordered) time points \(t_r \in [0,T], r \geq 1\), s.t. \(\mathcal{L}(\mathbf{O}|\mathbf{X}) = \prod_{r \geq 1} f(O_r|X_{t_r})\) for some mass/density function \(f(\cdot)\) over an arbitrary support set. A conditional probability mass function for an augmented sequence of states \(\hat{\mathbf{x}}\) is given by
\[
f_{\hat{\mathbf{x}}}(\hat{x}_0, \ldots, \hat{x}_m|\hat{t}, \hat{s}_1, \ldots, \hat{s}_m, Q(\lambda), \mathbf{O}, \Omega) \propto \pi(\hat{x}_0) \prod_{i=1}^{m} \tilde{P}_{\hat{x}_{i-1},\hat{x}_i}(\hat{t}_i) \cdot \prod_{s \in \hat{\mathbf{x}}} \left(1 + \frac{Q_{\hat{x}_i}(s) - \psi(s, \hat{x}_i)}{\Omega}\right) \prod_{r:t_r \in [\hat{t}_i, \hat{t}_{i+1})} f(O_r|\hat{x}_i),
\]
Chains, let \( \psi \) population process is assumed stationary and to jump models subject to measurement error (cf. Golightly and Wilkinson, 2015). Finally, if the transitions (by means of identity functions); thus, this offers an adaptable exact framework not restricted Conditioning on system jump observations. Relevant to epidemics, network queues and genetic chains, let \( O = \{O_r\}_{r \geq 1} \) be a sequence of jump observations at time points \( t_r \in [0, T], \ r \geq 1 \), s.t. 

\[
\mathcal{L}(O|X) = \prod_{i=1}^{\ldots} \left[ (1-p_{x_{i-1}, x_i}) \cdot \prod_{r \geq 1} \mathbf{1}(t_i \neq t_r) + \sum_{r \geq 1} \mathbf{1}(t_i = t_r) \cdot p_{x_{i-1}, x_i} \cdot f(O_r|x_{i-1}, x_i) \right]
\]

for trajectories \( X = (t, x) \), where \( p_{x,x'} \in [0,1], \ x, x' \in S \) denotes the probability that a process jump \( x \to x' \) triggers an observation with a conditional density \( f(\cdot); \) and \( p_{x,x} = 0 \) for all \( x \in S \). Then, 

\[
f_{x_{\tilde{x}_0}, \ldots, \tilde{x}_m|\tilde{t}_1, \ldots, \tilde{t}_m, Q(\lambda), O, \Omega} \propto \pi(\tilde{x}_0) \cdot \prod_{i=0}^{m} \prod_{s \in S} \left( 1 + \frac{Q_{x}(s) - \psi(s, x)}{\Omega} \right) \prod_{i=1}^{m} \tilde{P}_{\tilde{x}_{i-1}, \tilde{x}_i}(\tilde{t}_i) \left[ (1-p_{\tilde{x}_{i-1}, \tilde{x}_i}) \cdot \prod_{r \geq 1} \mathbf{1}(\tilde{t}_i \neq t_r) + \sum_{r \geq 1} \mathbf{1}(\tilde{t}_i = t_r) \cdot p_{\tilde{x}_{i-1}, \tilde{x}_i} \cdot f(O_r|\tilde{x}_{i-1}, \tilde{x}_i) \right],
\]

and a sampling procedure as introduced in Algorithm 1, where \( \tilde{P} \) in (12) is replaced by a sequence
of matrices $P_i$, $i = 1, \ldots, m$, s.t.
\[
P_{x,x'}(\hat{t}_i) = Q_{x,x'}(\hat{t}_i) \cdot p_{x,x'} \cdot f(O_r|x,x')
\]
whenever $\hat{t}_i = t_r$ for some $r \geq 1$, and
\[
P_{x,x'}(\hat{t}_i) = Q_{x,x'}(\hat{t}_i) \cdot (1 - p_{x,x'}) \quad \text{if} \quad x \neq x', \quad \text{with} \quad P_{x,x}(\hat{t}_i) = \psi(\hat{t}_i, x),
\]
otherwise, defines a Markov chain over MJP trajectories in $X$, with stationary distribution $f_X(t,x|\lambda, O)$ in (3). Above, equation (14) is explained by the fact that, in common application areas, only certain types of jumps are observable. For instance, removal times of infective individuals are often the basis for inferential epidemic studies, however, infectious times are never observed.

In all cases, the associated MCMC samplers yield ergodic Markov chains over posterior MJP trajectories. This is because, since matrices in $Q$ are sparse, a conditional sequence $\hat{x|t}$ is always supported within a finite product space of $n(t)$ subsets of $S$. However, $\psi(t,x) > 0$ for all $(t,x) \in [0,T] \times S$ by definition, and any full sequence in $X$ is always accessible by sampling an appropriate number of transition times. All trajectories are thus aperiodic and positive recurrent; moreover, auxiliary variables leave the target marginal distribution unaltered and the sampler will reach the desired invariant distribution.

### 3.2 Accelerating performance with stationary processes

In the reduction to a strictly stationary system (so rates are independent of time), forward-filtering steps in (9) reduce to
\[
\mathbb{P}(\hat{x}_{i+k} = x|u_1, \ldots, u_i, u_{i+1} = \bar{\phi}, \ldots, u_{i+k} = \bar{\phi}; \hat{t}, \hat{t}) \propto 1 + \frac{Q_{x} - \psi(x)}{\Omega} \cdot \mathbb{P}(\hat{x}_i = x|u_1, \ldots, u_i; \hat{t})
\]
whenever $u_j = \bar{\phi}$, $j = i + 1, \ldots, i + k$. Hence, times in $\hat{t}$ generated by a ‘compensating’ process in Proposition 3.1 are of no relevance; only Poisson counts $k_i, i = 0, \ldots, m$ in Algorithm 1 must be retained. Thus, a sampling scheme for stationary MJPs, similarly adaptable to observations, reduces to Algorithm 2. To aid the understanding of these results, Figure 4 shows an example with a graphical overview of a two-step data augmentation leading to count variables (15). On the left, we find an augmented trajectory ($\hat{t}, \hat{x}$); this includes the real MJP $(t, x)$ along with two virtual jumps sampled from a ‘controlled’ Poisson process with rate $\psi(\cdot) > 0$. In the centre, further virtual epochs are added from a ‘compensating’ process. This joint procedure corresponds to steps 1-2 within Algorithm 1, and is equivalent to splitting augmentation steps for stationary processes outlined in Rao and Teh (2013), where a larger sequence $\tilde{t}$ is directly sampled from (5). In the right diagram, the compensating virtual epochs are re-assigned as weights $k_i, i \geq 0$ over their corresponding nodes; within Algorithm 2, the times may be ignored for the purpose of re-sampling a new trajectory $\hat{x}$.

These diagrams help to depict major shortcomings behind traditional uniformization schemes (see Hobolth and Stone, 2009; Rao and Teh, 2013) for inference with stationary systems. Note that any

\[\text{Diagram: Schematic of augmentation procedure to } (\hat{t}, \hat{x}, k). \text{ Left, a trajectory } (\hat{t}, \hat{x}) \text{ with virtual jumps. Centre, } \text{compensating virtual epochs are superimposed. Right, superimposed epochs assigned as weights; times ignored.} \]
We show that 

\[ \int \text{Proof.} \]

and unbounded index set \( I \) to \( \Omega \). Thus, the majority of candidate times in \( \hat{\kappa} \)

candidate sets \( \hat{\kappa} \) for all \( \kappa \).

Lemma 3.2. Let \( k = \{k_0, \ldots, k_m\} \) of Poisson count variables with rates

\[ \Omega + Q \hat{x}_i - \psi(\hat{x}_i)) \cdot (\hat{t}_{i+1} - \hat{t}_i), \quad i = 0, \ldots, m, \quad \text{s.t.} \quad \hat{t}_{m+1} = T. \quad (15) \]

3. Draw a new sequence \( \hat{x} = \{\hat{x}_0, \ldots, \hat{x}_m\} \) with a forward-backward procedure; given initial distribution \( \pi(x) \), transition weight matrix

\[ \hat{P} = \text{diag}\{\psi(x) - Q x : x \in S\} + Q, \]

and (random) importance weights

\[ w_i(x) = \left( 1 + Q x - \psi(x) \right)^{k_i} \] \( i = 0, \ldots, m. \]

4. Remove self-transitions on \( (\hat{t}, \hat{x}) \) to produce \( (t, x)_{\text{new}} \).

(augmented) sequence \( \hat{t} \) is effectively a random discretization of a time-interval \([0, T]\), and serves as a basis for forward-backward procedures. Yet, population models are always governed by large/infinite generator matrices \( Q \), and are tied to large dominating rates \( \Omega > \max_{x \in S} |Q x| \). This leads to sizeable candidate sets \( \hat{t} \) with associated overheads during forward-backward procedures. However, underlying trajectories in \( X \) are likely to consistently transition states in \( S \) whose departure rates are ‘close’ to \( \Omega \). Thus, the majority of candidate times in \( \hat{t} \) will require thinning anyway. As observed in Figure 4, this paper builds over data augmentation techniques that restrict the cardinality of \( \hat{t} \), and correspondingly penalise self-transitions in order to preserve asymptotic exactness.

3.3 Limiting properties and arbitrarily large bounds

We begin with a preliminary result regarding convergence of sequences of random variables.

\textbf{Lemma 3.2.} Let \( a \in \mathbb{R} \) and \( b, c \in \mathbb{R}_{>0} \) be some fixed constant values, and define random variables \( u_\kappa \) by non-linear transformations

\[ u_\kappa = \left( 1 + \frac{a}{\kappa} \right)^v, \quad v_\kappa \sim \mathcal{P}\left( \lfloor \kappa + b \rfloor \cdot c \right) \]

for all \( \kappa \in \mathbb{R}_{>0}, \) s.t. every \( v_\kappa \) denotes a Poisson random variable with mean rate \( (\kappa + b) \cdot c \). Then, \( u_\kappa \xrightarrow{L^2} e^{a \cdot c} \) as \( \kappa \to \infty \), and any sequence of random variables \( u_i, i \in I \) defined over an increasing and unbounded index set \( I \) converges in mean square to the same constant value.

\textbf{Proof.} We show that \( \mathbb{E}[u_\kappa^2] \) exists for all \( \kappa \in \mathbb{R}_{>0}, \) and \( \lim_{\kappa \to \infty} \mathbb{E}\left( u_\kappa - e^{a \cdot c} \right)^2 = 0. \) First, note that

\[ \mathbb{E}[u_\kappa^2] = \sum_{x=0}^{\infty} \frac{[\kappa + b] \cdot c) e^{-(\kappa + b)c}}{x!} \left( 1 + \frac{a}{\kappa} \right)^x = e^{2ac + (a^2c + 2abc)/\kappa + a^2bc/\kappa^2}, \]
Algorithm 3 Reduced construction of non-stationary correlated MJP trajectories on $\mathcal{X}$.

**Input:** Sequence of intensity matrices $Q = Q(\lambda)$ parametrized by $\lambda$.

An MJP trajectory $(t, x) \in \mathcal{X}$ with $t = \{t_0, \ldots, t_n\}$ and $x = \{x_0, \ldots, x_n\}$.

Arbitrary intensity operator $\psi : [0,T] \times S \to \mathbb{R}_+$ for candidate times.

**Output:** A new MJP trajectory $(t, x)_{\text{new}} \in \mathcal{X}$ sampled from the density $f_X(t, x|Q)$ in (1).

1: Create an (ordered) set of candidate times $\hat{t} = \{\hat{t}_0, \ldots, \hat{t}_m\}$, $m \geq n$, attaching to $t$ auxiliary events from a Poisson process; rate $\psi(t, x_i) > 0$, within intervals $(t_i, t_{i+1})$, with $t_{n+1} = T$.

2: Draw a new sequence $\hat{x} = \{\hat{x}_0, \ldots, \hat{x}_m\}$ with a forward-backward procedure; given initial distribution $\pi(x)$, transition weight matrices $P(\hat{t}_i) = \text{diag}(\{\psi(\hat{t}_i, x) - Q_x(\hat{t}_i) : x \in S\}) + Q(\hat{t}_i)$, and importance weights $w_i(x) = \exp\left(\int_{\hat{t}_i}^{\hat{t}_{i+1}} [Q_x(s) - \psi(s, \hat{x}_i)] ds\right)$, imposed over epochs $i \in \{0, \ldots, m\}$.

3: Remove self-transitions on $(\hat{t}, \hat{x})$ to produce $(t, x)_{\text{new}}$.

which is well defined for all $\kappa \in \mathbb{R}_{\geq 0}$. Similarly $\mathbb{E}[u_k] = e^{ac + abc/\kappa}$, and it follows that

$$
\mathbb{E}\left[u_k - e^{ac}\right]^2 = \mathbb{E}[u_k]^2 - 2 \cdot e^{ac} \cdot \mathbb{E}[u_k] + e^{2ac} \xrightarrow{\kappa \to \infty} 0.
$$

Next, for each epoch $i = 0, \ldots, m$ within an (augmented) sequence $\hat{t}$, define a further partition of the interval $[\hat{t}_i, \hat{t}_{i+1}]$, into $\nu$ equally spaced subintervals of step size $\Delta t = \frac{\hat{t}_{i+1} - \hat{t}_i}{\nu}$, s.t.

$$
\int_{\hat{t}_i}^{\hat{t}_{i+1}} [\Omega + Q_x(s) - \psi(s, \hat{x}_i)] ds \approx \sum_{j=0}^{\nu-1} \Delta t \cdot [\Omega + Q_x(\hat{t}_i + j \cdot \Delta t) - \psi(\hat{t}_i + j \cdot \Delta t, \hat{x}_i)]
$$

(16)

offers a Riemann approximation (exact as $\Delta t \to 0$) to the intensity of compensating jumps in Proposition 3.1 and variables $k_i$ in Algorithm 1 (Step 2). The approximating rate is piecewise constant; s.t. compensating jumps under (16) are uniformly distributed in each tagged subinterval $j = 0, \ldots, \nu - 1$ of $[\hat{t}_i, \hat{t}_{i+1}]$. Thus, for all $i = 0, \ldots, m$ Poisson counts $k'_i$ respond to rates $\Delta t \cdot [\Omega + Q_x(\hat{t}_i + j \cdot \Delta t) - \psi(\hat{t}_i + j \cdot \Delta t, \hat{x}_i)]$; and $w_i(x)$ in (13) is approximated by

$$
w_i(x) \approx \prod_{j=0}^{\nu-1} \left(1 + \frac{Q_x(\hat{t}_i + j \cdot \Delta t) - \psi(\hat{t}_i + j \cdot \Delta t, \hat{x}_i)}{\Omega}\right)^{k'_i}.
$$

By Lemma 3.2, as the dominating rate $\Omega \to \infty$, and thus the inferential framework accommodates arbitrarily large rates within $Q(\lambda)$, it further holds

$$
w_i(x) \approx \exp\left(\int_{\hat{t}_i}^{\hat{t}_{i+1}} [Q_x(s) - \psi(s, \hat{x}_i)] ds\right),
$$

which leads to a simplified sampler design for non-stationary systems as shown in Algorithm 3 (similarly amendable to observations). There, note that the dominating rate $\Omega$ and compensating jumps are no longer relevant. The result retrieves an
analogue construction to algorithmic propositions for semi-Markov processes in Rao and Teh (2012); however, it requires iterative calculations of exponential functionals, notoriously resource-demanding in computational implementations.

4 Scalable sampling of deviations from mean-average dynamics

We continue with novel integrations of auxiliary variables to sample MJP paths in \( f_X(t, x|\lambda, O) \) in (3) as controlled deviations from approximate mean-average dynamics. As reference, we use a time functional \( \xi(t)_{0 \leq t \leq T} \) supported on an arbitrary set, so that a distance to population levels in \( S \) is quantifiable. We require that \( \xi(t) \) is close to a region of high density in the posterior distribution of \( X_t|O \), for all \( t \in [0, T] \). Under reasonably mild conditions, limiting theorems in Kurtz (1970, 1971) guarantee that every stochastic jump process accepts a real-valued deterministic approximation, as a solution to a system of ordinary differential equations (ODEs). This can be further calibrated to observed data in a computationally inexpensive manner, and we observe an example on the left hand side diagram within Figure 5.

Given \( \xi(t)_{0 \leq t \leq T} \), we complement each iteration in Algorithms 1-3 with a further auxiliary sequence \( u \) of the form (4). The goal is to form an informative set that restricts the explorable space of \( \hat{x} \) in sampling steps for (6). Importantly, this must be completed within Gibbs procedures and thus not compromise the mixing properties of the MCMC sampler (cf. Georgoulas et al., 2017). For simplicity in the presentation, we restrict the following formulations to integer-valued univariate population systems, where \( \xi(t)_{0 \leq t \leq T} \) is a real-valued function; however, the various definitions are readily amendable to multivariate models with various support sets.

**Auxiliary truncated normal random variables.** For a current (augmented) trajectory \( (\hat{t}, \hat{x}) \), define

\[
g(u_i|u_{i-1}, \hat{x}_i, \xi(\hat{t}_i)) = \phi \left( \frac{u_i - \mu_i}{\sigma} \right) / \sigma \left[ 1 - \Phi \left( \frac{|\hat{x}_i - \xi(\hat{t}_i)| - \mu_i}{\sigma} \right) \right],
\]

whenever \( \hat{x}_i \in (\xi(\hat{t}_i) - u_i, \xi(\hat{t}_i) + u_i), i = 1, \ldots, m \); where

\[
\mu_i = \max(\mu, u_{i-1} - \kappa), \quad \text{with} \quad \mu_1 = \mu \in \mathbb{R}_+, \kappa \in (0, 1),
\]

and \( \phi(\cdot), \Phi(\cdot) \) denote the standard normal density/cumulative distribution functions, respectively. Each \( u_i \) is thus normally distributed (mean \( \mu_i \), standard deviation \( \sigma \)) and truncated to a space...
over a subset of epochs $i \in \mathcal{I} \subseteq \{1, \ldots, m\}$; i.e. auxiliary variables are undefined for $i = 1, \ldots, m$ s.t. $i \not\in \mathcal{I}$, and $v_i, i \in \mathcal{I}$ are gamma distributed random variables. Here, $\alpha \in \mathbb{N}$ will secure a fast evaluation
of the associated densities; and rate parameters are subordinated to a random autoregressive process \( \mathbf{\mu} \) (stationary mean \( \mu \), lag-1 deviation \( \sigma \)), s.t. \( \beta_i = \alpha \cdot e^{-\mu_i} \) for all \( i \in \mathcal{I} \), with

\[
\mu_i | \mu_{i-1} \sim \mathcal{N}\left(\mu + (\mu_{i-1} - \mu)(1 - \kappa)^l, \sigma^2 \cdot (1 - (1 - \kappa)^2)/(1 - (1 - \kappa)^2)\right),
\]

where \( l \in \mathbb{N} \) denotes the lag between subsequent time points in \( \mathcal{I} \), and \( \kappa \in (0, 1) \). Thus, \( \mu_i \sim \mathcal{N}(\kappa \mu + (1 - \kappa)\mu_{i-1}, \sigma^2) \) whenever \( l = 1 \) and \( \mathcal{I} = \{1, \ldots, m\} \) is associated to all times in \( t \). This construct ensures \( \mathbb{E}[u_i | x_i, \xi(i)] = |x_i - \xi(i)| + e^{\mu_i} \) and \( \mathbb{V}[u_i] = e^{2\mu_i} / \alpha \); well calibrated, it allows for \( \hat{x} \) to significantly deviate from \( \xi(t_{i \leq T}) \) over restricted time-intervals. A diagram depicting such structure of auxiliary variables is shown in Figure 7; there, \( u_i, i \in \mathcal{I} \) are represented by coloured dots, placed over equally spaced epochs with lag \( l = 2 \); means assigned to Gamma variables (grey dots) are random and transition according to log-normal distributions.

Next, assume process jumps are of unit length. In order to sample a compatible sequence \( \hat{x} | \mathbf{t}, \mathbf{u} \) within Gibbs steps in Algorithm 1, the analogue to forward filtering procedures in (18) is given by

\[
\mathbb{P}(\hat{x}_i = x | \{u_j : j \leq i, j \in \mathcal{I}\}; \mathbf{\mu}, \mathbf{t}) \propto \prod (x \in \mathcal{S}_i) \cdot w_i(x | \mathbf{\mu}) \cdot \sum_{x' \in \mathcal{S}_{i-1}} \tilde{P}_{x'x}(\hat{t}_i) \cdot \mathbb{P}(\hat{x}_{i-1} = x' | \{u_j : j < i, j \in \mathcal{I}\}; \mathbf{\mu}, \mathbf{t})
\]

for \( i = 1, \ldots, m \), with restricted subsets defined s.t.

\[
\mathcal{S}_i = \begin{cases} 
\{x \in \mathcal{S} : \xi(\hat{t}_i) - u_i \leq x \leq \xi(\hat{t}_i) + u_i\} & \text{if } i \in \mathcal{I}, \\
\{x \in \mathcal{S} : \min \mathcal{S}_{i-1} - 1 \leq x \leq \max \mathcal{S}_{i-1} + 1\} & \text{otherwise}.
\end{cases}
\]

This assumes that \( X \) is supported over an unbounded set of integers (but may be suitably redefined otherwise). Importance weights are given by

\[
w_i(x | \mathbf{\mu}) = (u_i - |x - \xi(\hat{t}_i)|)^{\nu_i-1}e^{-\beta_i(u_i - |x - \xi(\hat{t}_i)|)}\prod_{s \in \mathcal{S}_i} \left(1 + \frac{Q_x(s) - \psi(s, x)}{\Omega}\right),
\]

whenever \( i \in \mathcal{I} \) and \( w_i(x | \mathbf{\mu}) = \prod_{s \in \mathcal{S}_i} (1 + (Q_x(s) - \psi(s, x)) / \Omega) \) otherwise. This suggests a forward implementation with dynamic vectors, since the explorable space of MJP trajectories expands across jump epochs \( i \notin \mathcal{I} \), while contracting again towards mean-average dynamics in the presence of auxiliary evidence. Backward sampling steps still correspond to (19) above. Again, similar amendments may be realized over Algorithms 2-3; also, further conditioning this procedure on observations corresponds to including alterations on \( \tilde{P}, w(\cdot) \) as listed in Subsection 3.1.
Below, we discuss results of algorithmic implementations of these methods, on two instances of popular jump processes, and we draw comparisons on efficiency with current benchmark methodologies for inferential tasks. Results and comparisons reported are produced by C++ implementations; code and data can be found on github.com/IkerPerez/scalableSamplingMJPs.

4.1 Example 1: A pure birth-death process

A birth-death process is a population model with applications in queueing theory and performance engineering tasks. In its simplest form, $X$ refers to a population supported within the set of non-negative integers $S = \mathbb{N}_0$, and state transitions involve both births and deaths. Infinitesimal rates for jumps are denoted by $\{\lambda_x(t)\}_{x \in S}$ and $\{\mu_x(t)\}_{x \in S}$, respectively, for all $t \geq 0$, so that

$$Q_{x,x'}(t) = \begin{cases} 
\lambda_x(t) & \text{if } x' = x + 1, \\
\mu_x(t) \cdot 1(x > 0) & \text{if } x' = x - 1, \\
-\lambda_x(t) - \mu_x(t) \cdot 1(x > 0) & \text{if } x' = x,
\end{cases}$$

and $Q_{x,x'}(t) = 0$ otherwise. Hence, the process increases its population by 1 whenever a birth occurs; alternatively, it decreases its population by 1 during a death event.

A finite capacity immigration-death process. In this variant, $S$ is bounded from above by some positive constant $N \in \mathbb{N}_0$; thus, it is a system equivalent to a closed queueing network with an infinite processor (cf. Perez and Casale, 2018), or a truncated $M/M/N/N$ queue (Gross et al., 2008). Importantly, for all states $x \in \{0, \ldots, N\}$, death rates scale along with population levels, s.t. $\mu_x(t) = x' \cdot \mu(t)$ for some time dependent function $\mu(t)$. Here, we assume that arrivals enter the system with a constant birth intensity $\lambda_x(t) = \lambda \cdot 1(x < N)$, $\lambda \in \mathbb{R}_+$, and death rates respond to seasonal patterns, s.t. $\mu(t) = \mu \cdot r(t)$ for some positive functional $r(t) \in [1, 2]$, $t \geq 0$. We further assume that $x_0 = N$, and note that the model is fully parametrized by $\lambda$ and $\mu$.

Noisy state observations and inference. Let $O_t = \{O_r\}_{r \geq 1}$ be state observations subject to measurement error, s.t. $O_t \sim \mathcal{N}(X_{t'}, \sigma^2)$ reflect normal random variables at times $t_r \in [0, T], r \geq 1$. This, along with a finite population set-up, allows for the implementation (for comparison purposes) of benchmark uniformization-based inferential techniques. We find sample observations within the left diagram in Figure 5 (black dots), for a latent process realisation with capacity $N = 50$ and seasonality $r(t) = 3/2 + \cos(2\pi \cdot t/T)/2$. The dark line in the figure corresponds to the deterministic approximation $\xi(t)$ with death rate parameter

$$\mu = \arg\min_{\mu \in \mathbb{R}_+} \sum_{r \geq 1} (\xi(t_r) - O_r)^2 \quad \text{subject to } \frac{d\xi(t)}{dt} = 1(\xi(t) < 50) \cdot \lambda - r(t) \cdot \xi(t) \cdot \mu,$$

and a (known) birth rate $\lambda$ fixed to an arbitrary value (ensuring model identifiability).

From (2), notice that the posterior density $f(\mu|O)$ requires integrating the observation likelihood, over all possible trajectories with associated density

$$f_X(t, x|Q) = \pi(x_0) e^{-\sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} f_i(x_r r(s) \mu + 1(x < N) \lambda) ds} \prod_{i=1}^{n} \lambda^{i(x_i = x_{i-1} + 1)} [\mu \cdot x_{i-1} \cdot r(t_i)]^{i(x_i = x_{i-1} - 1)},$$

and is thus intractable. In this task, we carry posterior MCMC inference on the death rate by iterating between sampling trajectories and $\mu$ from its conditional density

$$f(\mu|t, x) \propto \mu \sum_{i=1}^{n} (x_i = x_{i-1} - 1) e^{-\mu \sum_{i=1}^{n} x_i \int_{t_{i-1}}^{t_i} r(s) ds} \cdot \pi(\mu),$$

with some loosely uninformative prior $\pi(\mu)$. A sample trace output is shown in Figure 8, corresponding to the data displayed within Figure 5. There, the 3 different traces and densities correspond to (i) a traditional uniformization-based implementation (Rao and Teh, 2013), (ii) Algorithm 3 and (iii) a
variant centred around mean-average dynamics and normal auxiliary variables in (17). All alternatives yield equivalent estimates for $\mu$, of a seemingly similar quality.

We repeat the process for multiple simulated birth-death trajectories, at increasing population sizes $N$. In all cases, $T = 100$ and we produce 50 noisy observations over equally spaced intervals. Comparisons on efficiency across various methods are offered in Figure 14, which further includes a summary of population sizes tested, birth rates used and deviation associated with the observations. In the diagram, we find ratios in effective sample sizes (scaled for computation time) against the benchmark algorithm of Rao and Teh (2013) (black line) with dominating rate $\Omega = 1.5 \cdot \max_{x \in S} \sup_{t \in [0,T]} |Q_x(t)|$. Whiskers represent 95% confidence intervals. There, (i) the blue line corresponds to Algorithm 3, with $\psi(t, x) = |Q_x(t)|$, $(t, x) \in [0,T] \times S$, s.t. auxiliary jumps attached to $t$ are generated in proportion to diagonal elements of $Q(t)$, $t \in [0,T]$, and $\hat{t}$ is of approximately double the size of $t$ in each MCMC iteration, (ii) the green line is for further restricting MJP samples to deviations from $\xi(\cdot)$, using auxiliary variables (17) with mean $\mu = N/10$, autoregressive coefficient $\kappa = 1$ and Gaussian deviation $\sigma = 0.65 \cdot (1 + \kappa)$; this offers a good heuristic, noting that birth-death jumps are of magnitude one s.t. truncated normal densities are always substantial, (iii) the red line finally assigns $\mu = \sqrt{N}$, $\kappa = 0.05$ and $\sigma = 1.5 \cdot (1 + \kappa)$, s.t. explorable spaces for $\hat{x}$ are very restricted around the mean-average solution, yet, randomness and autoregressive effects are strong and can accommodate sudden short-timed deviations from $\xi(t)_{t \in [0,T]}$.

---

**Figure 8:** MCMC traces and kernel density representations for death rates, corresponding to uniformization (red), Algorithm 3 (blue) and Algorithm 3 combined with variables (17) (black).

**Figure 9:** Left, diagram with ratios in effective sample sizes versus a benchmark uniformization-based inference algorithm, for different algorithmic implementations. Intervals around points represent confidence intervals. Right, table summarizing the population levels, birth-rates and observation error used in simulations.
Overall, Algorithm 3 does not pose big gains over traditional uniformization, since rates for jumps in a birth-death system scale linearly with the population. Yet, from confidence metrics across both green and red scalings, we conclude that well-tuned auxiliary-variable techniques presented in this section yield very significant efficiency gains; due to the approach naturally integrating within Gibbs steps for posterior paths of $X$.

### 4.2 Example 2: Markovian stochastic epidemic models

Next, we address an inferential task with a time-homogeneous Susceptible-Infected-Removed (SIR) stochastic epidemic model (Bailey et al., 1975). Here, $X = (S_t, I_t, R_t)_{t \in [0,T]}$ tracks a population of $N$ individuals s.t. $S = \{0, \ldots, N\}^3$. At any time $t \in [0, T]$ each member of the population is either susceptible (capable of contracting a disease), infective (able to pass the disease to others) or removed (immune to infection and unable to infect others). Since $S_t + I_t + R_t = N$, then $X \equiv (S_t, I_t)_{t \in [0,T]}$ corresponds to a bivariate jump process. In common applications, $X$ begins with a susceptible population of $N-1$ individuals, along with an infective member whose disease contraction time is unknown; the infinitesimal generator matrix $Q$ is such that

$$Q \{s,i\}, \{s',i'\} = \begin{cases} \beta s i & \text{if } s' = s-1, i' = i+1, \\ \gamma i & \text{if } s' = s, i' = i-1, \\ 0 & \text{otherwise.} \end{cases}$$

and $Q \{s,i\}, \{s',i'\} = 0$ otherwise. Therefore, infective individuals become removed (by death or recovery) after an independent infectious period with removal rate $\gamma$. While infected, they may further transfer the disease to members of the susceptible population through Poisson contacts with infection rate $\beta$. When the epidemic has ceased at some terminal time $T$, then the entire population is divided between susceptibles (who avoided infection) and removed members.

**Observations removed and inference.** Here, $X$ can further be represented by a triplet $(t, s, i)$ of transition times along with corresponding susceptible/infected population vectors. In common inferential settings, all removal observations $t^R$ are available; that is, some $k < N$ times $0 \leq t^R_1 < \cdots < t^R_k < T$ when infective individuals have either died or recovered from a disease. Thus,

$$\mathcal{L}(t^R|t, s, i) = \prod_{j=1}^{[t^R]} I\{\lim_{t \nearrow t^R_j} I_t = I^R_{j-1} + 1\},$$

which is a simplified, analogue expression to (14), where removal jump observations are always observed. The term $\mathcal{L}(t^R|\beta, \gamma)$, key for inference tasks on the rates, requires integrating over a space of full infection and removal times with associated density

$$f_X(t, s, i|\beta, \gamma) = \pi(i_0)e^{-\sum_{j=0}^{[t]} (\beta s_j i_j + \gamma i_j)(t_j - t_{j-1})} \prod_{j=0}^{[t]-1} (\beta s_{j+1} i_j) \gamma i_j, \gamma i_{j+1} = i_{j+1} - i_j,$$

where $t_{[t+1]} = T$, and $\pi(i_0)$ denotes the distribution of the initial infection time. The expression offers a basis for an MCMC approach to inference, through augmentation of the MJP trajectory with missing infection times; in combination with samples from rate posteriors

$$f(\beta|t, s, i) \propto \beta^{|t^R|-1}e^{-\sum_{j=0}^{[t]} \beta s_j i_j(t_{j+1} - t_j)} \cdot \pi(\beta) \quad \text{and} \quad f(\gamma|t, s, i) \propto \beta^{|t^R|}e^{-\sum_{j=0}^{[t]} \gamma i_j(t_{j+1} - t_j)} \cdot \pi(\gamma)$$

and time-intervals $[t_0, T]$ with initial infection $\pi(t_0|t_1) \propto e^{-\beta(N-1) + \gamma(t_1 - t_0)} \cdot \pi(t_0)$, where $\pi(\beta), \pi(\gamma)$ and $\pi(t_0)$ denote priors. This is usually achieved with Metropolis-Hastings steps (O’Neill and Roberts, 1999; Jewell et al., 2009), where updates proceed by proposing additions, deletions or moves of a proportion (usually half) of infection times; however, the scalability of the algorithm is reportedly poor. This is displayed in Figure 10, where we find output traces for parameters in a small epidemic ($N = 50$). In all density and autocorrelation (ACF) diagrams, black/grey representations correspond...
to an auxiliary-variable algorithm as introduced in this paper; red coloured counterparts relate to a benchmark M-H implementation (O’Neill and Roberts, 1999). Severe efficiency differences may be observed within the ACF plot. On the left, we find posterior mean dynamics and a 95% credible interval for \((I_t + R_t)_{t \in [0,T]}\); the dashed blue line corresponds to the real unobserved value, and the green line is the (observed) removal process \((R_t)_{t \in [0,T]}\) with jump times \(t^R\).

Next, we compare efficiency metrics in the procedures across increasing populations; and further analyze an adaptation of uniformization methods in Rao and Teh (2013) to system-jump observations (see definitions in Subsection 3.1). Removal data is simulated with rates \(\gamma = 1, \beta = 2/N\) and securing a final removed population \(R_T = N \cdot 80\) (most representative outcome). Ratios on effective samples are reported within Figure 11; there, benchmark lines (in blue) correspond to Algorithm 2 (not requiring exponential evaluations), with operator \(\psi(x) = |Q_x/2|, x \in S\) (candidate jumps attached to \(t\) with half intensity of diagonal in \(Q\)). We make this choice because (i) the model is stationary and (ii) existing alternative augmentation schemes do not scale (i.e. they do not work) with large populations. Also, (i) the green line represents the afore-mentioned benchmark epidemics Metropolis algorithm, (ii) the black line is for vanilla uniformization; with dominating rate \(\Omega = 1\) and (iii) the red line corresponds to sampling paths as deviations from mean-average dynamics \(\xi(\cdot)\), given by the solution to a multivariate system

\[
\frac{d\xi_S(t)}{dt} = -\beta \cdot \xi_S(t) \cdot \xi_t(t), \quad \frac{d\xi_t(t)}{dt} = \beta \cdot \xi_S(t) \cdot \xi_t(t) - \gamma \cdot \xi_t(t), \quad \text{and} \quad \frac{dR(t)}{dt} = \gamma \cdot \xi_t(t),
\]

with infection/removal parameters set to optimize \(\min\beta,\gamma \in \mathbb{R}_+ \sum_{t \in D[0,T]} (\xi_R(t) - R_t)^2\) over an arbitrary discretization \(D[0,T]\) of the time interval. This is achieved incorporating Gamma variables in (20) over Algorithm 2, with lag \(l = 25\), stationary mean \(\mu = \log(N/10)\), autoregressive coefficient \(\kappa = 0.5\) and deviation \(\sigma = 0.25\).

Existing inferential methods (green and black lines) do not scale to sizeable populations and perform poorly even within small ones. Noticeably, vanilla uniformization is bound to be inefficient in systems where generator rates scale quadratically; in epidemics, the data-augmentation procedure is associated with large dominating rates, often s.t. \(\Omega > \beta \cdot (N/2)^2 + \gamma \cdot N\).

5 Splitting the problem by mapping states or transitions

Finally, we discuss mappings to reduce full MJP augmentations into families of smaller end-point conditioned tasks. A fixed \(l \in \mathbb{N}\) will again define a lag for auxiliary variables in (4), among the
discretization epochs in \( \hat{t} \). We thus employ a reduced (deterministic) sequence \( \{u_i\}_{i=\hat{l},2\hat{l},...} \) at times \( \hat{l},\hat{2l},... \) s.t. \( u_i = T(\hat{x}_{i-1},\hat{t}_i) \) for some surjective mapping \( T : S^2 \to J \); and variables in \( \mathbf{u} \) are undefined other than for lagged times. Through \( T \), we map pairs of states in \( S^2 \) to elements of the power set \( \Sigma_S \). A particular case of such construct was first discussed in Perez et al. (2018); there, the authors simplify augmentation tasks for networked queueing systems by mapping MJP state transitions to job orderings across queues. Importantly, within the following examples, a lag \( \hat{l} \) must be (randomly) re-instantiated (or drifted) within every MCMC iteration, in order to ensure that trajectories \( X \) are sampled from within their full support \( \mathcal{X} \).

**Partitioning a state space.** Here, an (augmented) MJP process is forced to transition (small) population ranges at lagged times \( \{\hat{t}_i\}_{i=\hat{l},2\hat{l},...} \). For a univariate \( S \)-valued process example, we define

\[
J \subset \Sigma_S \quad \text{s.t.} \quad \emptyset \notin J, \bigcup_{A \in J} = S \quad \text{and} \quad A \cap B = \emptyset \quad \text{for all} \ A,B \in J.
\]

Each part \( A \in J \) must be defined s.t. jumps (including virtual self-jumps) restricted among its states yield an irreducible Markov chain on \( A \). Then, for every existing (augmented) sequence \( \hat{x} \), we let \( T(x,x') = \{A \in J : x' \in A\} \) map jumps \( x \to x' \) at times \( \{\hat{t}_i\}_{i=\hat{l},2\hat{l},...} \) to parts \( \{A_i\}_{i=\hat{l},2\hat{l},...} \) that contain the arrival states \( \{\hat{x}_i\}_{i=\hat{l},2\hat{l},...} \). In order to re-sample a new compatible sequence \( \hat{x}\hat{t}, \mathbf{u} \) within Algorithms 1-3, we can split forward-backward procedures over intervals \( [\hat{t}_{i-1},\hat{t}_i) \), \( i = \hat{l},2\hat{l},... \) s.t. each forward estimation

\[
P(\hat{x}_i = x|u_0,\ldots,u_{i-1};\hat{t}_i) \quad \text{at epochs} \quad i = \hat{l},2\hat{l},...
\]

is restricted to the subset \( A_i \) of \( S \) and fed as the initial distribution \( \pi(i) \) at time \( \hat{t}_i \) during the next interval. Backward steps proceed normally within and across sub-intervals. In Figure 12 (left) we find a sample diagram depicting this partitioning of the augmentation task. There, grey circles represent the reach of a univariate birth-death jump process at time points in \( \hat{t} \); blue squares (assigned at randomly lagged times, not equally spaced) correspond to ranges the process must transit.

**Sampling end-point conditioned bridges.** To further simplify data-augmentation, the above partition may be defined s.t. \( J = S \) and \( T(x,x') = x' \), for all \( x,x' \in S \). Thus, a new compatible sequence \( \hat{x}\hat{t}, \mathbf{u} \) will be locked at times \( \{\hat{t}_i\}_{i=\hat{l},2\hat{l},...} \), and forward-backward procedures are independent across subintervals \( [\hat{t}_{i-1},\hat{t}_i) \). The approach is depicted within Figure 12 (right), where our algorithms will sample bridges across the auxiliary mapped states.

In both cases, methods easily generalise to multivariate process settings, and trajectories may straightforwardly be conditioned on data by following previously introduced conventions. Noticeably, the correlation across subsequent trajectory samples for \( X \) will be drastically increased; yet, this is compensated by considerably simplified procedures, and we below report on efficiency results with
algorithmic implementations for a predator-prey model. C++ repositories to reproduce these results may be found on github.com/IkerPerez/scalableSamplingMJP.

5.1 Example 3: An stochastic Lotka-Volterra model

A Lotka-Volterra model (Boys et al., 2008) describes predator-prey interactions among two biological species. Here, a non-stationary process \(X = (X_1^t, X_2^t)_{t \geq 0}\) evolves stochastically according to rates

\[
Q_{(x_1, x_2), (x_1 + 1, x_2)}(t) = \alpha(t) \cdot x_1, \quad Q_{(x_1, x_2), (x_1, x_2 - 1)}(t) = \gamma(t) \cdot x_2,
\]

\[
Q_{(x_1, x_2), (x_1, x_2 + 1)}(t) = \delta(t) \cdot x_1 \cdot x_2, \quad Q_{(x_1, x_2), (x_1, x_2 - 1)}(t) = \delta(t) \cdot x_1 \cdot x_2.
\]

Thus, \((X_1^t)_{t \geq 0}\) refers to the prey population and \((X_2^t)_{t \geq 0}\) is the predator counterpart. Within the following inferential task, functionals decompose between interaction parameters and a seasonality modifier; so that \(\alpha(t) = \alpha \cdot r(t), \beta(t) = \beta \cdot r(t)\) and so on, for some (known) \(r(t) \in [1, 2], t \geq 0\). Additionally, an initial state is (uniformly) randomized between (bounded) populations with capacity \(N \in \mathbb{N}_0\).

State measurements and inference. For simulated datasets (at various population bounds), we produce noisy state observations s.t. \(O_r \sim N(X_{t_r}, N/25)\) at equally spaced times \(t_r \in [0, T], r \geq 1\). Throughout, parameter choices \(\alpha = 0.125, \beta = 0.005\) and \(\gamma = 0.1\) are assigned, with \(r(t) = 3/2 + \cos(2\pi \cdot t/T)/2\). Similarly to previous examples, backwards inference on the rates proceeds by data augmentation of trajectory densities \(f_X(t, x|Q)\) in (1), along with draws from the posterior \(f(\alpha, \beta, \delta, \gamma| t, x)\) (which factors across the individual rates). In Figure 13 we find a sample representation of augmented prey and predator population trajectories (red lines), along with observations (dark circles). There, dashed lines represent posterior mean-average paths, and grey areas are for %95 credible intervals.

Efficiency results comparing different augmentation methods are shown in Figure 14. As before,
the diagrams display ratios (along with confidence intervals) in effective sample sizes (scaled for computation time). The horizontal axes represent bounds imposed over each marginal biological species; thus, the real explorable state space tested increases up to $120^2 = 14,400$. The left diagram corresponds to average effective samples across parameter rates $\alpha, \beta, \delta, \gamma$; instead, the right diagram represents ratios on the minimum effective samples across the 4 parameters. In both instances, the reference line (in red) at level 1 corresponds to sampling end-point conditioned bridges across (randomised) intervals with lag $l = 0.5 \cdot N$, built on top of Algorithm 3 with operator $\psi(t, x) = 0.5 \cdot |Q_x(t)|$, $(t, x) \in [0, T] \times S$. The green line is for the same procedure, but using a lag $l = 0.75 \cdot N$; and the blue line represents a plain implementation of Algorithm 3 without auxiliary variables driving an increase in efficiency. Finally, in black we observe efficiency results for a vanilla uniformization procedure with dominating rate $\Omega = 1.5 \cdot \max_{x \in S} \sup_{t \in [0, T]} |Q_x(t)|$.

![Graph showing ratios in effective sample sizes versus a benchmark auxiliary-variable procedure.]

Results are consistent with auxiliary-variable methods introduced in Section 4. In all cases, the various alternatives introduced in this paper can (i) scale inferential uniformization-based inferential frameworks to much larger problems, and (ii) drive significant increases in computational efficiency.

6 Discussion

This paper has presented a novel and comprehensive framework for the design of scalable data-augmentation procedures, suitable for use within exact Bayesian inferential tasks, and applicable to birth-death, epidemic or predator-prey systems, to name only a few. The need for auxiliary-variable augmentation designs as presented here is justified by the limitations in existing state-of-the-art uniformization-based approaches (see Hobolth and Stone, 2009; Rao and Teh, 2012, 2013; Miasojedow and Niemiro, 2015; Georgoulas et al., 2017; Zhang and Rao, 2018, and references therein), which are inefficient, unadaptable or unusable with mid-sized or large population systems, often associated with multiple types of observational data.

We have reported on results that apply multiple MCMC algorithm construction to problems of broad statistical interest, and demonstrated prior claims on efficiency and scalability benefits, by direct comparison to current benchmark methods in the literature. Finally, since the presented framework builds on uniformized representations of non-stationary jump processes, we note that the various techniques introduced in this paper will be only applicably to purely Markovian processes.
References

Bailey, N. T. et al. (1975). *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd.

Boys, R. J., Wilkinson, D. J., and Kirkwood, T. B. (2008). Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing*, 18(2):125–135.

Cappé, O., Robert, C. P., and Rydén, T. (2003). Reversible jump, birth-and-death and more general continuous time markov chain monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(3):679–700.

Daley, D. J. and Vere-Jones, D. (2007). *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media.

Fearnhead, P. and Sherlock, C. (2006). An exact gibbs sampler for the markov-modulated poisson process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(5):767–784.

Georgoulas, A., Hillston, J., and Sanguinetti, G. (2017). Unbiased bayesian inference for population markov jump processes via random truncations. *Statistics and computing*, 27(4):991–1002.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361.

Golightly, A. and Sherlock, C. (2018). Efficient sampling of conditioned markov jump processes. *arXiv preprint arXiv:1809.07139*.

Golightly, A. and Wilkinson, D. J. (2015). Bayesian inference for markov jump processes with informative observations. *Statistical applications in genetics and molecular biology*, 14(2):169–188.

Gross, D., Shortle, J. F., Thompson, J. M., and Harris, C. M. (2008). *Fundamentals of Queueing Theory*. Wiley-Interscience, New York, NY, USA, 4th edition.

Higdon, D. M. (1998). Auxiliary variable methods for markov chain monte carlo with applications. *Journal of the American Statistical Association*, 93(442):585–595.

Hobolth, A. and Stone, E. A. (2009). Simulation from endpoint-conditioned, continuous-time markov chains on a finite state space, with applications to molecular evolution. *The Annals of Applied Statistics*, 3(3):1204–1231.

Jensen, A. (1953). Markov chains as an aid in the study of Markov processes. *Scandinavian Actuarial Journal*, 36:87–91.

Jewell, C. P., Kypraios, T., Neal, P., and Roberts, G. O. (2009). Bayesian analysis for emerging infectious diseases. *Bayesian Analysis*, 4(3):465–496.

Kurtz, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump markov processes. *Journal of Applied Probability*, 7(1):4958.

Kurtz, T. G. (1971). Limit theorems for sequences of jump markov processes approximating ordinary differential processes. *Journal of Applied Probability*, 8(2):344356.

Miasojedow, B. and Niemiro, W. (2015). Particle gibbs algorithms for markov jump processes. *arXiv preprint arXiv:1505.01434*.

Miasojedow, B., Niemiro, W., et al. (2017). Geometric ergodicity of rao and tehs algorithm for markov jump processes and cbtms. *Electronic Journal of Statistics*, 11(2):4629–4648.

O’Neill, P. D. and Roberts, G. O. (1999). Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(1):121–129.

Opper, M. and Sanguinetti, G. (2008). Variational inference for markov jump processes. In *Advances in Neural Information Processing Systems*, pages 1105–1112.

Perez, I. and Casale, G. (2018). Approximate bayesian inference with queueing networks and coupled jump processes. *arXiv preprint arXiv:1807.08673*.

Perez, I., Hodge, D., and Kypraios, T. (2018). Auxiliary variables for bayesian inference in multi-class queueing networks. *Statistics and Computing*, 28(6):1187–1200.

Rao, V. and Teh, Y. W. (2012). MCMC for continuous-time discrete-state systems. In *Advances in Neural Information Processing Systems*, pages 701–709.

Rao, V. A. and Teh, Y. W. (2013). Fast MCMC sampling for Markov jump processes and extensions. *Journal of Machine Learning Research*, 14:3295–3320.
Sutton, C. and Jordan, M. I. (2011). Bayesian inference for queueing networks and modeling of internet services. *The Annals of Applied Statistics*, 5(1):254–282.

Van Dijk, N. M. (1992). Uniformization for nonhomogeneous markov chains. *Operations research letters*, 12(5):283–291.

Van Dijk, N. M., Van Brummelen, S. P. J., and Boucherie, R. J. (2018). Uniformization: Basics, extensions and applications. *Performance evaluation*, 118:8–32.

Zhang, B. and Rao, V. (2018). Efficient parameter sampling for markov jump processes. *arXiv preprint arXiv:1704.02369*. 