ARTICLE IN PRESS

Statistics and Probability Letters ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Piecewise deterministic Markov processes for scalable Monte Carlo on restricted domains

Joris Bierkens [a,*], Alexandre Bouchard-Côté [b], Arnaud Doucet [c], Andrew B. Duncan [d], Paul Fearnhead [e], Thibaut Lienart [c], Gareth Roberts [f], Sebastian J. Vollmer [f]

[a] Delft Institute of Applied Mathematics, TU Delft, Netherlands
[b] Department of Statistics, University of British Columbia, Canada
[c] Department of Statistics, University of Oxford, UK
[d] School of Mathematical and Physical Sciences, University of Sussex, UK
[e] Department of Mathematics and Statistics, Lancaster University, UK
[f] Department of Statistics, University of Warwick, UK

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
MCMC
Bayesian statistics
Piecewise deterministic Markov processes
Logistic regression

ABSTRACT

Piecewise Deterministic Monte Carlo algorithms enable simulation from a posterior distribution, whilst only needing to access a sub-sample of data at each iteration. We show how they can be implemented in settings where the parameters live on a restricted domain.

1. Introduction

Markov chain Monte Carlo (MCMC) methods have been central to the wide-spread use of Bayesian methods. However their applicability to some modern applications has been limited due to their high computational cost, particularly in big-data, high-dimensional settings. This has led to interest in new MCMC methods, particularly non-reversible methods which can mix better than standard reversible MCMC (Diaconis et al., 2000; Turitsyn et al., 2011), and variants of MCMC that require accessing only small subsets of the data at each iteration (Welling and Teh, 2011).

One of the main technical challenges associated with likelihood-based inference for big data is the fact that likelihood calculation is computationally expensive (typically $O(N)$ for data sets of size N). MCMC methods built from piecewise deterministic Markov processes (PDMPs) offer considerable promise for reducing this $O(N)$ burden, due to their ability to use sub-sampling techniques, whilst still being guaranteed to target the true posterior distribution (Bierkens et al., 2016; Bouchard-Côté et al., 2015; Fearnhead et al., 2016; Galbraith, 2016; Pakman et al., 2016). Furthermore, factor graph decompositions of the target distribution can be leveraged to perform sparse updates of the variables (Bouchard-Côté et al., 2015; Nishikawa et al., 2015; Peters and De With, 2012).

PDMPs explore the state space according to constant velocity dynamics, but where the velocity changes at random event times. The rate of these event times, and the change in velocity at each event, are chosen so that the position of the resulting process has the posterior distribution as its invariant distribution. We will refer to this family of sampling methods as Piecewise Deterministic Monte Carlo methods (PDMC).

* Corresponding author.
E-mail address: joris.bierkens@tudelft.nl (J. Bierkens).

https://doi.org/10.1016/j.spl.2018.02.021
0167-7152/

Existing PDMC algorithms can only be used to sample from posteriors where the parameters can take any value in $\mathbb{R}^d$. In this paper (Section 2) we show how to extend PDMC methodology to deal with constraints on the parameters. Such models are ubiquitous in machine learning and statistics. For example, many popular models used for binary, ordinal and polychotomous response data are multivariate real-valued latent variable models where the response is given by a deterministic function of the latent variables (Albert and Chib, 1993; Fahrmeir and Tutz, 2001; Train, 2009). Under the posterior distribution, the domain of the latent variables is then constrained based on the values of the responses. Additional examples arise in regression where prior knowledge restricts the signs of marginal effects of explanatory variables such as in econometrics (Geweke, 1986), image processing and spectral analysis (Bellavia et al., 2006; Guo and Berman, 2012) and non-negative matrix factorization (Kim and Park, 2007). A few methods for dealing with restricted domains are available but these either target an approximation of the correct distribution (Patterson and Teh, 2013) or are limited in scope (Pakman and Paninski, 2014).

2. Piecewise deterministic Monte Carlo on restricted domains

Here we present the general PDMC algorithm in a restricted domain. Specific implementations of PDMC algorithms can be derived as continuous-time limits of familiar discrete-time MCMC algorithms (Bierkens and Roberts, 2017; Peters and De With, 2012), and these derivations convey much of the intuition behind why the algorithms have the correct stationary distribution. Our presentation of these methods is different, and more general. We first define a simple class of PDMPs and show how these can be simulated. We then give simple recipes for how to choose the dynamics of the PDMP so that it will have the correct stationary distribution.

Our objective is to compute expectations with respect to a probability distribution π on $\mathcal{O} \subseteq \mathbb{R}^d$ which is assumed to have a smooth density, also denoted $\pi(x)$, with respect to the Lebesgue measure on $\mathcal{O}$. With this objective in mind, we will construct a continuous-time Markov process $Z_t = (X_t, V_t)_{t\geq 0}$ taking values in the domain $E = \mathcal{O} \times \mathcal{V}$, where $\mathcal{O}$ and $\mathcal{V}$ are subsets of $\mathbb{R}^d$, such that $\mathcal{O}$ is open, pathwise connected and with Lipschitz boundary $\partial\mathcal{O}$. In particular, if $\mathcal{O} = \mathbb{R}^d$ then $\partial\mathcal{O} = \emptyset$. The dynamics of Z_t are easy to describe if one views X_t as position and V_t as velocity. The position process X_t moves deterministically, with constant velocity V_t between a discrete set of *switching times* which are simulated according to N inhomogeneous Poisson processes, with respective intensity functions $\lambda_i(X_t, V_t)$, $i = 1, \ldots, N$, depending on the current state of the system. At each switching time the position stays the same, but the velocity is updated according to a specified transition kernel. More specifically, suppose the next switching event occurs from the ith Poisson process, then the velocity immediately after the switch is sampled randomly from the probability distribution $Q_i(x, v, \cdot)$ given the current position x and velocity v. The switching times are random, and designed in conjunction with the kernels $(Q_i)_{i=1}^N$ so that the invariant distribution of the process coincides with the target distribution π.

To ensure that X_t remains confined within $\mathcal{O}$ the velocity of the process is updated whenever X_t hits $\partial\mathcal{O}$ so that the process moves back into $\mathcal{O}$. We shall refer to such updates as *reflections* even though they need not be specular reflections.

The resulting stochastic process is a Piecewise Deterministic Markov Process (PDMP, Davis, 1984). For it to be useful as the basis of a Piecewise Deterministic Monte Carlo (PDMC) algorithm we need to (i) be able to easily simulate this process; and (ii) have simple recipes for choosing the intensities, $(\lambda_i)_{i=1}^N$, and transition kernels, $(Q_i)_{i=1}^N$, such that the resulting process has $\pi(x)$ as its marginal stationary distribution. We will tackle each of these problems in turn.

2.1. Simulation

The key challenge in simulating our PDMP is simulating the event times. The intensity of events is a function of the state of the process. But as the dynamics between event times are deterministic, we can easily represent the intensity for the next event as a deterministic function of time. Suppose that the PDMP is driven by a single inhomogeneous Poisson process with intensity function

$$\widetilde{\lambda}(u; X_t, V_t) = \lambda(X_t + uV_t, V_t), \quad u \geq 0.$$

We can simulate the first event time directly if we have an explicit expression for the inverse function of the monotonically increasing function

$$u \mapsto \int_0^u \widetilde{\lambda}(s; X_t, V_t)\, ds. \tag{1}$$

In this case the time until the next event is obtained by (i) simulating a realization, y say, of an exponential random variable with rate 1; and (ii) setting the time until the next event as the value τ that solves $\int_0^\tau \widetilde{\lambda}(s; X_t, V_t)\, ds = y$.

Inverting (1) is often not practical. In such cases simulation can be carried out via *thinning* (Lewis and Shedler, 1979). This requires finding a tractable upper bound on the rate, $\bar{\lambda}(u) \geq \widetilde{\lambda}(u; X_t, V_t)$ for all $u > 0$. Such an upper bound will typically take the form of a piecewise linear function or a step function. Note that the upper bound $\bar{\lambda}$ is only required to be valid along the trajectory $u \mapsto (X_t + uV_t, V_t)$ in $\mathcal{O} \times \mathcal{V}$. Therefore the upper bound may depend on the starting point (X_t, V_t) of the line segment we are currently simulating. We then propose potential events by simulating events from an inhomogeneous Poisson process with rate $\bar{\lambda}(u)$, and accept an event at time u with probability $\widetilde{\lambda}(u; X_t, V_t)/\bar{\lambda}(u)$. The time of the first accepted event will be the time until the next event in our PDMP.

Please cite this article in press as: Bierkens J., et al., Piecewise deterministic Markov processes for scalable Monte Carlo on restricted domains. Statistics and Probability Letters (2018), https://doi.org/10.1016/j.spl.2018.02.021.