Statistics and Probability Letters 113 (2016) 1–6

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Geometric ergodicity of Rao and Teh's algorithm for homogeneous Markov jump processes

Błażej Miasojedow *, Wojciech Niemiro

Institute of Applied Mathematics and Mechanics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland

ARTICLE INFO

Article history:
Received 2 December 2015
Received in revised form 8 February 2016
Accepted 8 February 2016
Available online 20 February 2016

Keywords:
Continuous time Markov processes
MCMC
Hidden Markov models
Geometric ergodicity
Drift condition
Small set

ABSTRACT

Rao and Teh (2013) introduced an efficient MCMC algorithm for sampling from the posterior distribution of a hidden Markov jump process. The algorithm is based on the idea of sampling virtual jumps. In the present paper we show that the Markov chain generated by Rao and Teh's algorithm is geometrically ergodic. To this end we establish a geometric drift condition towards a small set. We work under the assumption that the parameters of the hidden process are known and the goal is to restore its trajectory.

1. Introduction

Markov jump processes (MJP) are natural extension of Markov chains to continuous time. They are widely applied in modelling of the phenomena of chemical, biological, economic and other sciences.

In many applications it is necessary to consider a situation where the trajectory of a MJP is not observed directly, only partial and noisy observations are available. Typically, the posterior distribution over trajectories is then analytically intractable. In the literature there exist several approaches to the above mentioned problem: based on sampling (Boys et al., 2008; El-Hay et al., 2008; Fan and Shelton, 2008; Golightly and Wilkinson, 2011, 2014; Golightly et al., 2015; Nodelman et al., 2002; Rao and Teh, 2013, 2012), and also based on numerical approximations. To the best of our knowledge the most general efficient method for a finite state space is that proposed by Rao and Teh (2013), and extended to a more general class of continuous time discrete systems in Rao and Teh (2012). Although the method proposed by Fearnhead and Sherlock (2006), after a minor modification, can be used to sample exactly from the posterior distribution of a homogeneous MJP, their algorithm involves calculating matrix exponentials, which is computationally expensive.

In the present paper we establish geometric ergodicity of Rao and Teh's algorithm for homogeneous MJPs. Geometric ergodicity is a key property of Markov chains which implies Central Limit Theorem for sample averages.

Note that in practice the parameters of the hidden MJP may be unknown and have to be estimated. Then the Rao and Teh's algorithm can be combined with an additional step (Gibbs or Metropolis–Hastings) updating these parameters, according to some posterior distribution. Such extended versions of the Rao and Teh's algorithm are not considered in our paper. We assume that the probability law of a hidden MJP is known.

* Corresponding author.
E-mail addresses: bmia@mimuw.edu.pl (B. Miasojedow), wniem@mimuw.edu.pl (W. Niemiro).

http://dx.doi.org/10.1016/j.spl.2016.02.002
0167-7152/

The rest of the paper is organized as follows. In Section 2 we briefly introduce hidden Markov jump processes, next in Section 3 we recall the Rao and Teh's algorithm. The main result is proved in Section 4.

2. Hidden Markov jump processes

Consider a continuous-time homogeneous Markov process $\{X(t), t^{\min} \le t \le t^{\max}\}$ on a finite state space $\mathcal{S}$. Its probability law is defined via the initial distribution $\nu(s) = \mathbb{P}(X(t^{\min}) = s)$ and the transition intensities

$$Q(s, s') = \lim_{h \to 0} \frac{1}{h} \mathbb{P}(X(t+h) = s' | X(t) = s)$$

for $s, s' \in \mathcal{S},\ s \ne s'$. Let $Q(s) = \sum_{s' \ne s} Q(s, s')$ denote the intensity of leaving state s. For definiteness, assume that X has right-continuous trajectories. We say X is a Markov jump process (MJP).

Suppose that process X cannot be directly observed but we can observe some random quantity Y with probability distribution $L(Y|X)$. Let us say Y is the evidence and L is the likelihood. The problem is to restore the hidden trajectory of X given Y. From the Bayesian perspective, the goal is to compute/approximate/sample from the posterior

$$p(X|Y) \propto p(X)L(Y|X).$$

Function L, transition probabilities Q and initial distribution ν are assumed to be known. To get the explicit form of posterior distribution we consider a typical form of noisy observation. Assume that the trajectory $X([t^{\min}, t^{\max}])$ is observed independently at k deterministic time points with some random errors. Formally, we observe $Y = (Y_1, \dots, Y_k)$ where

$$L(Y|X) = \prod_{j=1}^{k} L_j(Y_j | X(t_j^{\mathrm{obs}})), \tag{1}$$

for some fixed known points $t^{\min} \le t_1^{\mathrm{obs}} < \cdots < t_k^{\mathrm{obs}} \le t^{\max}$.

3. Uniformization and Rao and Teh's algorithm

In this section we describe a classical and well-known technique of uniformization (Jensen, 1953).

A Markov jump process can be represented in terms of potential times of jumps and the corresponding states. Every trajectory $X([t^{\min}, t^{\max}])$ is right continuous and piecewise constant: $X(t) = S_{i-1}$ for $T_{i-1} \le t < T_i$, where random variables T_i are such that $t^{\min} < T_1 < \cdots < T_N < t^{\max}$ (by convention, $T_0 = t^{\min}$ and $t^{\max} < T_{N+1}$). The random sequence of states $S = (S_0, S_1, \dots, S_N)$ such that $S_i = X(T_i)$ is called a skeleton. We do not assume that $S_{i-1} \ne S_i$, and therefore the two sequences

$$\begin{pmatrix} T \\ S \end{pmatrix} = \begin{pmatrix} t^{\min} & T_1 & \cdots & T_i & \cdots & T_N & t^{\max} \\ S_0 & S_1 & \cdots & S_i & \cdots & S_N & \end{pmatrix}$$

represent the process X in a redundant way: many pairs (T, S) correspond to the same trajectory $X([t^{\min}, t^{\max}])$. Let $J = \{i \in [1:N] : S_{i-1} \ne S_i\} \cup \{0\}$, so that $T_J = (T_i : i \in J)$ are moments of *true* jumps and $T_{-J} = T \setminus T_J = (T_i : i \notin J)$ are *virtual* jumps. We write $[l:r] = \{l, l+1, \dots, r\}$. By a harmless abuse of notation, we identify increasing sequences of points in $[t^{\min}, t^{\max}]$ with finite sets. Note that the trajectory of X is uniquely defined by (T_J, S_J). Let us write $X \equiv (T_J, S_J)$ and also use the notation $J(X) = T_J$ for the set of true jump times.

Uniformization obtains if T is a sequence of consecutive points of a homogeneous Poisson process with intensity λ, where $\lambda \ge Q^{\max} = \max_s Q(s)$. The skeleton S is then (independently of T) a discrete-time, homogeneous Markov chain with the initial distribution ν and the transition matrix

$$P(s, s') = \begin{cases} \dfrac{Q(s, s')}{\lambda} & \text{if } s \ne s'; \\[2ex] 1 - \dfrac{Q(s)}{\lambda} & \text{if } s = s'. \end{cases} \tag{2}$$

Rao and Teh (2013) exploit uniformization to construct a special version of Gibbs sampler which converges to the posterior $p(X|Y)$. The key facts behind their algorithm are the following. First, given the trajectory $X \equiv (T_J, S_J)$ the conditional distribution of virtual jump times T_{-J} is that of the non-homogeneous (actually piecewise homogeneous) Poisson process with intensity $\lambda - Q(X(t)) \ge 0$. Second, this distribution does not change if we introduce the likelihood. Indeed, $L(Y|X) = L(Y|T_J, S_J)$, so Y and T_{-J} are conditionally independent given (T_J, S_J) and thus $p(T_{-J}|T_J, S_J, Y) = p(T_{-J}|T_J, S_J)$. Third, the conditional distribution $p(S|T, Y)$ is that of a hidden discrete time Markov chain and can be efficiently sampled from using the algorithm FFBS (Forward Filtering–Backward Sampling, Carter and Kohn (1994) and Frühwirth-Schnatter (1994)).

The Rao and Teh's algorithm generates a Markov chain $X_0, X_1, \dots, X_m, \dots$ (where $X_m = X_m([t^{\min}, t^{\max}])$ is a trajectory of an MJP), convergent to $p(X|Y)$, where $Y = (Y_1, \dots, Y_k)$ is a vector of observations with the probability distribution of the form (1). A single step, that is the rule of transition from $X_{m-1} = X$ to $X_m = X'$ is described in Algorithm 1.

Convergence of the algorithm has been shown by its authors in Rao and Teh (2013). It follows from the fact that the chain has the stationary distribution $p(X|Y)$ and is irreducible and aperiodic, provided that $\lambda > Q^{\max}$.