BOQIAN MJPS*

AUTHOR†

Abstract.

Key words. MCMC, Markov Jump Process, Slice Sampling, infinite state space

1. Beam Sampling for continuous time Infinite Hidden Markov Models.

Model Assumptions:
We are considering a continuous time Infinite Hidden Markov Model (iHMM), with transition matrix $A = (A_{ij})_{i,j\geqslant 1}$.

$$A_i \doteq A_{ii} = -\sum_{j\neq i} A_{ij}$$
$$A_{ij} > 0 \, , i \neq j$$

Now, we are using the way in reference 1 to construct a MJPs with virtual jumps.

PROPOSITION 1.1. *The path (W, V) returned by the thinning procedure described in algorithm 3 is equivalent to a sample (S, T) from the $MJP(\pi_0, A)$.*

Proof. $S = (s_0, s_1, ..., s_N)$,$T = (t_0, t_1, ..., t_N, t_{N+1})$. And let's call the virtual jumps as U. Denote the virtual jump times between (t_i, t_{i+1}) as n_i. Then the density function of (W, V) will be as follows.

$$\begin{aligned} P(W,V) &= \pi_0(s_0)\prod_{i=0}^{N-1}\exp(-B_{s_i}(t_{i+1}-t_i))B_{s_i}^{n_i}(1-\frac{A_{s_i}}{B_{s_i}})^{n_i}B_{s_i}\frac{A_{s_i}}{B_{s_i}}\cdot\exp(-B_{s_N}(t_{N+1}-t_N))B_{s_N}^{n_N}(1-\frac{A_{s_N}}{B_{s_N}})^{n_N} \\ &= \pi_0(s_0)\exp(-\int_{t_0}^{t_{N+1}} B_{S_{(t)}}dt)\prod_{i=0}^{N}(B_{s_i}-A_{s_i})^{n_i}\prod_{i=0}^{N-1}A_{s_is_{i+1}} \end{aligned}$$

So after integrating with respect to virtual jump times and the numbers of virtual jumps, we can get the following.

$$\begin{aligned} P(S,T) &= \sum_{n_1,n_2,...,n_N\geqslant 0}\int_{t_0\leqslant\tau_1^1\leqslant...\leqslant\tau_{n_1}^1\leqslant t_1}...\int_{t_N\leqslant\tau_1^N\leqslant...\leqslant\tau_{n_N}^N\leqslant t_{N+1}} P(W,V)d\tau_1^1...d\tau_{n_1}^1...d\tau_1^N...d\tau_{n_N}^N \\ &= \pi_0(s_0)\exp(-\int_{t_0}^{t_{N+1}} A_{S_{(t)}}dt)\prod_{i=0}^{N-1}A_{s_is_{i+1}} \end{aligned}$$

So the proposition is proved.
□
The main idea of beam sampler for infinite-state continuous time Hidden Markov

*This work was supported by the Society for Industrial and Applied Mathematics
†Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania. (tex@siam.org). Questions, comments, or corrections to this document may be directed to that email address.

Algorithm 1 State-dependent thinning for MJPs

Input: Transition matrix $A_{ss'}$, and an initial distribution over states π_0.
Dominating Transition Rate Vector $B_s \geqslant A_s$.
Output: A piecewise constant trajectory $(V, W) = ((v_i, w_i))$ on the time interval $[t_{start}, t_{end}]$.
Initialize,
Draw $v_0 \sim \pi_0$ and set $w_0 = t_{start}$. Set $i = 0$.
while $w_i < t_{end}$ **do**
  (a) Sample $\tau_i \sim B_{v_i}$.
  (b) Set $v_{i+1} = v_i$ with probability $1 - \frac{A_{v_i}}{B_{v_i}}$ and set $w_{i+1} = w_i + \tau_i$.
  (c) **Else:** Set $w_{i+1} = w_i + \tau_i$ and sample v_{i+1} with $P(v_{i+1} = s|v_i) = A_{v_i,s}/A_{v_i}$.
  (d) Incresement i.
end while

Model is to introduce auxiliary variables μ such that conditioned on μ, the number of trajectories with positive probability is finite. Then dynamic programming can be used to compute the conditional probabilities efficiently.
Assume $W = (w_0, w_1, ..., w_{N'}, w_{N'+1})$, $V = (v_0, v_1, ..., v_{N'})$, $\mu = (\mu_1, \mu_2, ..., \mu_{N'})$.

$$P(\mu|W, V) = \prod_{i=1}^{N'} \frac{\mathbb{I}(0 \leqslant \mu_i \leqslant \frac{A_{v_{i-1}v_i}}{A_{v_{i-1}}})}{\frac{A_{v_{i-1}v_i}}{A_{v_{i-1}}}}$$

It indicates that conditioned on the trajectory(V, W), μ_i is depending on A, v_i, and v_{i-1} and $\mu_i \sim Uniform(0, \frac{A_{v_{i-1}v_i}}{A_{v_{i-1}}})$.

Proposition 1.2. *Conditioned on a trajectory (S, T) of the MJP, the virtual jump times U are distributed as a Poisson process with density $B_{s(t)} - A_{s(t)}$.*

Proof. $S = (s_0, s_1, ..., s_N)$,$T = (t_0, t_1, ..., t_N, t_{N+1})$. And let's call the virtual jumps as U. Denote the virtual jump times between (t_i, t_{i+1}) as n_i. Then the density function of (W, V) will be as follows.

$$\begin{aligned} P(W, V) &= P(U, S, T) \\ &= \pi_0(s_0) \exp(-\int_{t_0}^{t_{N+1}} B_{S_{(t)}} dt) \prod_{i=0}^{N} (B_{s_i} - A_{s_i})^{n_i} \prod_{i=0}^{N-1} A_{s_i s_{i+1}} \end{aligned}$$

$$P(S, T, n_0, ..., n_N) = \pi_0(s_0) \exp(-\int_{t_0}^{t_{N+1}} B_{S_{(t)}} dt) \prod_{i=0}^{N} \frac{((B_{s_i} - A_{s_i})(t_{i+1} - t_i))^{n_i}}{n_i!} \prod_{i=0}^{N-1} A_{s_i s_{i+1}}$$

So the conditional probability $P(n_0, n_1, ..., n_N|S,T)$ will be as follows.

$$P(n_0, ..., n_N|S,T) = \exp(-\int_{t_0}^{t_{N+1}} (B_{S_{(t)}} - A_{S_{(t)}})dt) \prod_{i=0}^{N} \frac{((B_{s_i} - A_{s_i})(t_{i+1} - t_i))^{n_i}}{n_i!}$$

So it indicates that conditioned on the trajectory (S,T), the virtual jump U is distributed as a non-homogeneous Poisson process with density $B_{s(t)} - A_{s(t)}$. □
Sampling v: Using the same trick used in Beam Sampling for the Infinite HMM, we can sample $P(v_t|y, \mu, W)$. So can we sample $P(v_t|v_{t+1}, y, W, u)$.

First of all, consider $P(v_i|y_{w_0,w_{i+1}}, w_{0:i}, \mu_{0:i})$.

$$\begin{aligned}
P(v_i, w_{0:i}, \mu_{0:i}, y_{[w_0,w_{i+1})}) &= \sum_{v_{i-1}} P(v_i, y_{[w_i,w_{i+1})}, w_i, \mu_i, v_{i-1}, w_{0:i-1}, \mu_{0:i-1}, y_{[w_0,w_i)}) \\
&= \sum_{v_{i-1}} P(v_i, y_{[w_i,w_{i+1})}, w_i, \mu_i|v_{i-1}, w_{0:i-1}, \mu_{0:i-1}, y_{[w_0,w_i)})P(v_{i-1}, w_{0:i-1}, \mu_{0:i-1}, y_{[w_0,w_i)}) \\
&= \sum_{v_{i-1}} P(y_{[w_i,w_{i+1})}|v_i, w_i, w_{i+1})P(\mu_i|v_i, v_{i-1})P(v_i, w_i|v_{i-1}, w_{i-1})P(v_{i-1}, w_{0:i-1}, \mu_{0:i-1}, y_{[w_0,w_i)}) \\
&= P(y_{[w_i,w_{i+1})}|v_i, w_i, w_{i+1}) \sum_{v_{i-1}} \frac{\mathbb{I}(0 \leqslant \mu_i \leqslant \frac{A_{v_{i-1}v_i}}{A_{v_{i-1}}})}{\frac{A_{v_{i-1}v_i}}{A_{v_{i-1}}}} \exp(-B_{v_{i-1}}(w_i - w_{i-1})) \\
&(B_{v_{i-1}} - A_{v_{i-1}})^{\mathbb{I}(v_i=v_{i-1})} A_{v_{i-1}v_i}^{\mathbb{I}(v_i v_{i-1})} P(v_{i-1}, w_{0:i-1}, \mu_{0:i-1}, y_{[w_0,w_i)}) \\
&= P(y_{[w_i,w_{i+1})}|v_i, w_i, w_{i+1}) \sum_{\Im_{i-1}} \frac{\mathbb{I}(0 \leqslant \mu_i \leqslant \frac{A_{v_{i-1}v_i}}{A_{v_{i-1}}})}{\frac{A_{v_{i-1}v_i}}{A_{v_{i-1}}}} \exp(-B_{v_{i-1}}(w_i - w_{i-1})) \\
&(B_{v_{i-1}} - A_{v_{i-1}})^{\mathbb{I}(v_i=v_{i-1})} A_{v_{i-1}v_i}^{\mathbb{I}(v_i v_{i-1})} P(v_{i-1}, w_{0:i-1}, \mu_{0:i-1}, y_{[w_0,w_i)})
\end{aligned}$$

Although the summation over v_{i-1} is an infinite sum, the auxiliary variable μ_i truncates this summation to the finitely many v_{i-1}'s and v_i's that satisfy both constrains $\mu_i \leqslant \frac{A_{v_{i-1}v_i}}{A_{v_{i-1}}}$ and $P(v_{i-1}|y_{[w_0,w_i)}, \mu_{0:i-1}) > 0$. This means that $|\Im_{i-1}| < +\infty$.

Secondly, consider $P(v_i|v_{i+1}, y_{w_0,w_{N'+1}}, w_{0:N'+1}, \mu_{0:N'})$.

$$\begin{aligned}
&P(v_i|v_{i+1}, y_{w_0,w_{N'+1}}, w_{0:N'+1}, \mu_{0:N'}) \propto P(v_i, v_{i+1}, y_{w_0,w_{N'+1}}, w_{0:N'+1}, \mu_{0:N'}) \\
&= P(y_{[w_{i+1},w_{N'+1})}, \mu_{i+2:N}, w_{i+2:N}|v_i, v_{i+1}, y_{[w_0,w_{i+1}), w_{0:i+1}, \mu_{0:i+1})})P(v_i, v_{i+1}, y_{[w_0,w_{i+1})}, w_{0:i+1}, \mu_{0:i+1}) \\
&= P(y_{[w_{i+1},w_{N'+1})}, \mu_{i+2:N}, w_{i+2:N}|v_{i+1}, w_{i+1})P(v_i, v_{i+1}, y_{[w_0,w_{i+1})}, w_{0:i+1}, \mu_{0:i+1}) \\
&= Const \cdot P(v_i, v_{i+1}, y_{[w_0,w_{i+1})}, w_{0:i+1}, \mu_{0:i+1}) \\
&= Const \cdot P(v_{i+1}, u_{i+1}, w_{i+1}|v_i, y_{[w_0,w_{i+1})}, w_{0:i}, \mu_{0:i}) \cdot P(v_i, w_{0:i}, \mu_{0:i}, y_{[w_0,w_{i+1})}) \\
&= Const \cdot P(v_{i+1}, u_{i+1}, w_{i+1}|v_i, w_{0:i}, \mu_{0:i}) \cdot P(v_i, w_{0:i}, \mu_{0:i}, y_{[w_0,w_{i+1})})
\end{aligned}$$

Finally, to sample the complete trajectory, we can sample $P(v_{N'}|y_{w_0,w_{N'+1}}, \mu_{0:N'})$ first, and then do a backward sampling using the above formula.

Algorithm 2 Beam Sampler for continuous time Infinite Hidden Markov Models

Input: observations $y_{[t_0,t_{k+1})}$, A, B, π_0
Initialize, $i = 0$
(a) Set current trajectory $[S,T](0)$ arbitrarily.
repeat
  for $i = 0$ **to** N **do**
    (a) Sample virtual jumps $U(i+1) \sim Poisson\ Process(B_{s(t)} - A_{s(t)})$, given $S(i), T(i)$.
    (b) Sample $\mu(i+1)_j \sim Uniform(0, \frac{A_{v_{j-1}v_j}}{A_{v_{j-1}}})$, $j = 1, 2, ..., N'$.
    (c) Sample $V(i+1) \sim P(V|W(i+1), \mu(i+1), y)$
    (d) Delete all the virtual jumps to get $S(i+1),\ T(i+1)$
  end for
until $i = N$

Theorem 1.3. *Algorithm 4 has $P(S,T,W,\mu|y)$ as a stationary distribution.*

Proof. Firstly, prove (c) step has $P(S,T|W,\mu,y)$ as a stationary distribution. It comes from the following detail balance condition.

$$\begin{aligned}
P((W,S,T,\mu) \to (W,S^*,T^*,\mu))P(S,T|W,\mu,y) &= P(V^*|W,\mu,y)P(V|W,\mu,y) \\
&= P((W,S^*,T^*,\mu) \to (W,S,T,\mu))P(S^*,T^*|W,\mu,y)
\end{aligned}$$

Secondly, prove (a) and (b) step have $P(W,\mu|S,T,y)$ as a stationary distribution.

$$\begin{aligned}
P(W,\mu|S,T,y) = P(U,\mu|S,T,y) &= \frac{P(U,\mu,S,T,y)}{P(S,T,y)} \\
&= \frac{P(y|S,T)P(U,\mu,S,T)}{P(y|S,T)P(S,T)} = P(\mu|S,T,U)P(U|S,T) \\
&= P(\mu|V,W)P(U|S,T)
\end{aligned}$$

We know the transition probability $P((S,T,W,\mu) \to (S,T,W^*,\mu^*))$ is as follows.

$$\begin{aligned}
P((S,T,W,\mu) \to (S,T,W^*,\mu^*)) &= P(\mu^*|V^*,W^*)P(U^*|S,T) \\
&= P(\mu^*|S,T,U^*)P(U^*|S,T) = P(W^*,\mu^*|S,T,y)
\end{aligned}$$

So step(a) and (b) have $P(W,\mu|S,T,y)$ as a stationary distribution.
Above all, this theorem is proved.
□

2. Figures and tables.

3. Bibliography and BibTEX.

4. Conclusion. Appendix. The use of appendices.

Appendix A. Title of appendix.

REFERENCES

[1] Vinayak Rao, Yee Whye Teh, *MCMC for continuous-time discrete-state systems*, NIPS, 2012.