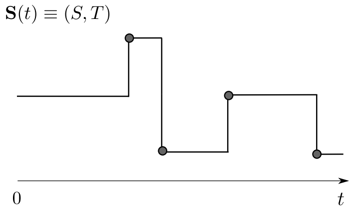
MCMC for Continuous-Time Discrete-State Systems

Vinayak Rao and Yee Whye Teh

Gatsby Computational Neuroscience Unit,
University College London

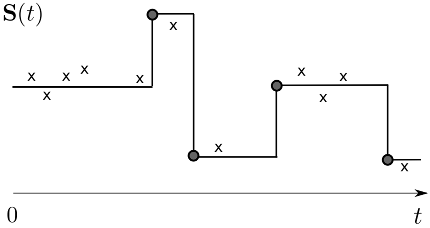
Overview

- Continuous-time discrete-state systems:



- applications in physics, chemistry, genetics, ecology, neuroscience etc.
- Examples: the Poisson process, renewal processes, Markov jump processes, continuous time Bayesian networks, semi-Markov processes etc.
- Our focus: efficient posterior inference via MCMC

Posterior inference

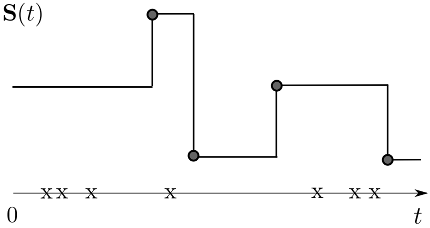


Typically, we have partial (and noisy) observations:

- State values at the end points of an interval.
- Observations $x(t) \sim F(\mathbf{S}(t))$ at a finite set of times t.

Given noisy observations of a trajectory, obtain posterior samples.

Posterior inference



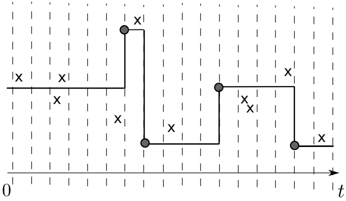
Typically, we have partial (and noisy) observations:

- State values at the end points of an interval.
- Observations $x(t) \sim F(\mathbf{S}(t))$ at a finite set of times t.
- More complicated likelihood functions that depend on the entire trajectory, e.g. Markov modulated Poisson processes

Given noisy observations of a trajectory, obtain posterior samples.

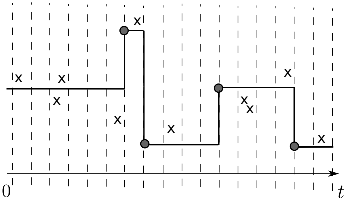
Posterior inference

One approach: discretize time.



Posterior inference

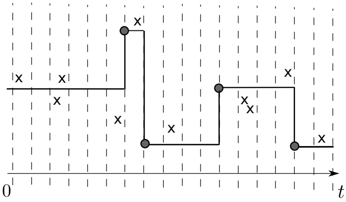
One approach: discretize time.



Pros: Can avail of vast literature on MCMC for discrete time-series models. Eg. the forward-backward algorithm.

Posterior inference

One approach: discretize time.

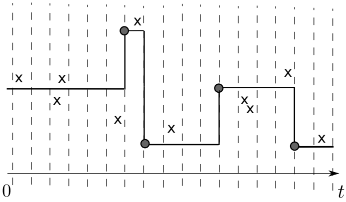


Pros: Can avail of vast literature on MCMC for discrete time-series models. Eg. the forward-backward algorithm.

Cons: Is an approximation, and introduces bias: the system can now only change state at times on a fixed grid.
To control the bias, we need a fine time-discretization, resulting in long chains.

Posterior inference

One approach: discretize time.

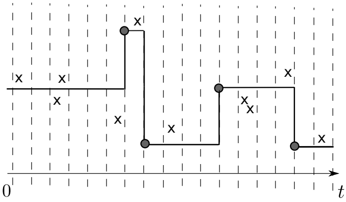


In this talk:

- Eliminate bias altogether, by devising an *exact* MCMC sampler.
- Still can use MCMC techniques from discrete time-series models.

Posterior inference

One approach: discretize time.

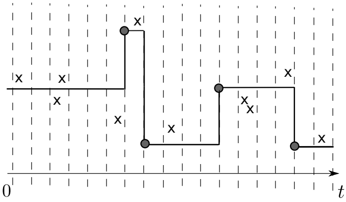


In this talk:

- Eliminate bias altogether, by devising an *exact* MCMC sampler.
- Still can use MCMC techniques from discrete time-series models.
- We proceed by constructing a *random* discretization of time.

Posterior inference

One approach: discretize time.



In this talk:

- Eliminate bias altogether, by devising an *exact* MCMC sampler.
- Still can use MCMC techniques from discrete time-series models.
- We proceed by constructing a *random* discretization of time.

We start by constructing this discretization from a Poisson process. [Rao and Teh, 2011]

The Poisson process (on the real line)

The homogeneous Poisson process with rate λ:

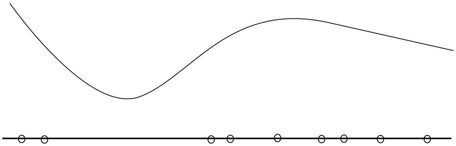
- P(an event in a small interval Δt) $\approx \lambda \Delta t$
- 'time' between successive events has distribution $\exp(\lambda)$

The Poisson process (on the real line)

The homogeneous Poisson process with rate λ:

- P(an event in a small interval Δt) $\approx \lambda \Delta t$
- 'time' between successive events has distribution $\exp(\lambda)$

The *inhomogeneous* Poisson process with rate $\lambda(t)$:



- the probability of an event in a small interval Δt is $\lambda(t)\Delta t$

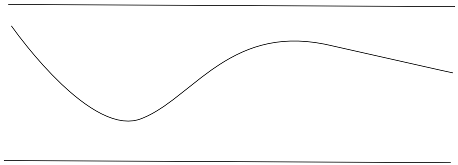
Thinning [Lewis and Shedler, 1979]

Thinning: to sample from a Poisson process with rate $\lambda(t)$.

Thinning [Lewis and Shedler, 1979]

Thinning: to sample from a Poisson process with rate $\lambda(t)$.

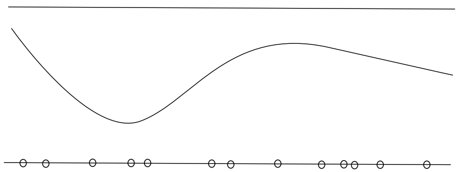
- Choose $\Omega > \lambda(t)\ \forall t$.



Thinning [Lewis and Shedler, 1979]

Thinning: to sample from a Poisson process with rate $\lambda(t)$.

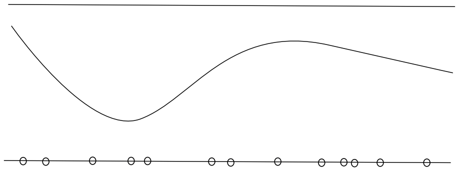
- Choose $\Omega > \lambda(t)\ \forall t$.
- Sample from a Poisson process with rate Ω.



Thinning [Lewis and Shedler, 1979]

Thinning: to sample from a Poisson process with rate $\lambda(t)$.

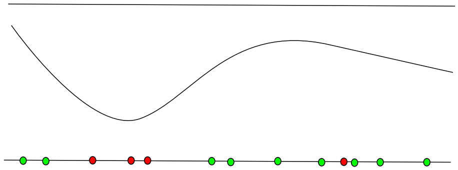
- Choose $\Omega > \lambda(t)\ \forall t$.
- Sample from a Poisson process with rate Ω.
- Keep each point with probability $\frac{\lambda(t)}{\Omega}$, otherwise 'thin'.



Thinning [Lewis and Shedler, 1979]

Thinning: to sample from a Poisson process with rate $\lambda(t)$.

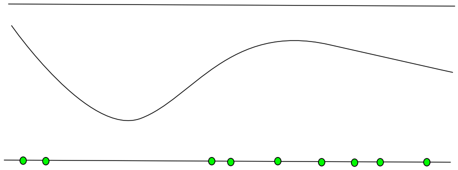
- Choose $\Omega > \lambda(t)\ \forall t$.
- Sample from a Poisson process with rate Ω.
- Keep each point with probability $\frac{\lambda(t)}{\Omega}$, otherwise 'thin'.



Thinning [Lewis and Shedler, 1979]

Thinning: to sample from a Poisson process with rate $\lambda(t)$.

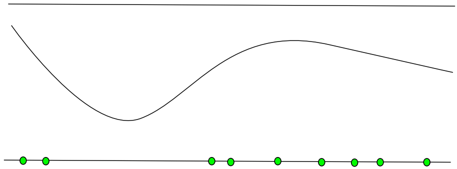
- Choose $\Omega > \lambda(t)\ \forall t$.
- Sample from a Poisson process with rate Ω.
- Keep each point with probability $\frac{\lambda(t)}{\Omega}$, otherwise 'thin'.



Thinning [Lewis and Shedler, 1979]

Thinning: to sample from a Poisson process with rate $\lambda(t)$.

- Choose $\Omega > \lambda(t)\ \forall t$.
- Sample from a Poisson process with rate Ω.
- Keep each point with probability $\frac{\lambda(t)}{\Omega}$, otherwise 'thin'.

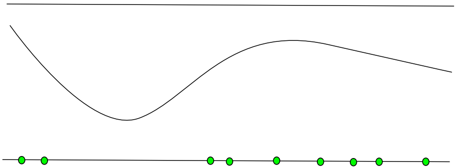


Follows from the *complete randomness* of the Poisson process.

Thinning [Lewis and Shedler, 1979]

Thinning: to sample from a Poisson process with rate $\lambda(t)$.

- Choose $\Omega > \lambda(t)\ \forall t$.
- Sample from a Poisson process with rate Ω.
- Keep each point with probability $\frac{\lambda(t)}{\Omega}$, otherwise 'thin'.



Follows from the *complete randomness* of the Poisson process.

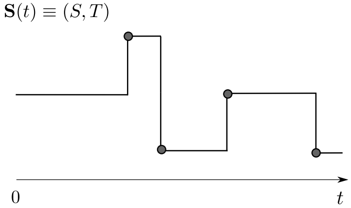
We consider pure-jump processes with temporal dependencies: thin points by running a *Markov chain*.

Uniformization (at a high level)

- Define Ω larger than the fastest rate at which 'events occur'.
- Draw from a Poisson process with rate Ω.
- Construct a Markov chain with transition times given by the drawn point set.
- The Markov chain is *subordinated* to the Poisson process.
- Keep a point t with probability $\lambda(t|state)/\Omega$.

Markov jump processes (MJPs)

An MJP $\mathbf{S}(t),\ t \in \mathbb{R}_+$ is a right-continuous piecewise-constant stochastic process taking values in some finite space $\mathcal{S} = \{1, 2, ...n\}$. It is parametrized by an *initial distribution* π and a *rate matrix* A.



$$\begin{bmatrix} -A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & -A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & -A_{nn} \end{bmatrix}$$

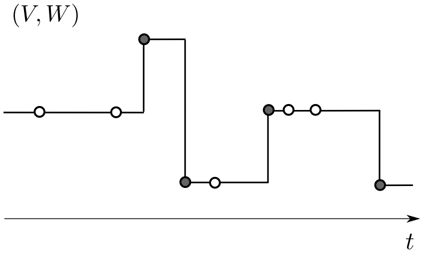
A_{ij} : rate of leaving state i for j

$$A_{ii} = \sum_{j=1, j \neq i}^{n} A_{ij}$$

A_{ii} : rate of leaving state i

Uniformization for MJPs [Jensen, 1953]

- Alternative to Gillespie's algorithm.
- Sample a set of times from a Poisson process with rate $\Omega \geq \max_i A_{ii}$ on the interval $[t_{start}, t_{end}]$.
- Run a discrete time Markov chain with initial distribution π and transition matrix $B = (I + \frac{1}{\Omega} A)$ on these times.



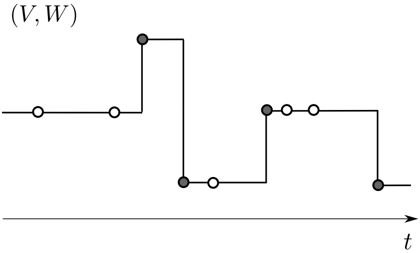
The matrix B allows self-transitions.

Uniformization for MJPs [Jensen, 1953]

Proposition

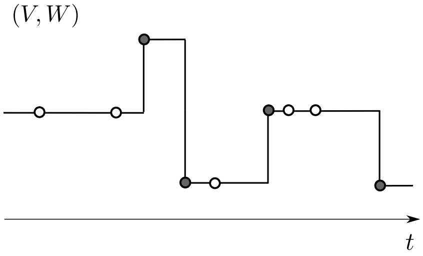
For any $\Omega \geq \max_i |A_{ii}|$, *the (continuous time) sequence of states obtained by the uniformized process is a sample from a MJP with initial distribution* π *and rate matrix* A.

Auxiliary variable Gibbs sampler



Inference via MCMC.

Auxiliary variable Gibbs sampler



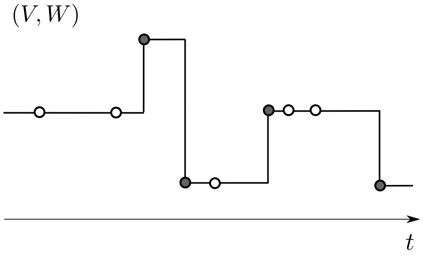
Inference via MCMC.

State space of the sampler consist of:

- Trajectory of MJP $\mathbf{S}(t)$.
- Auxiliary set of points rejected via self-transitions.

[Rao and Teh, 2011]

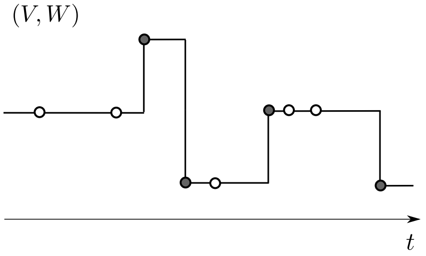
Auxiliary variable Gibbs sampler



Inference via MCMC.

- Given current MJP path, we need to resample the set of rejected points. Conditioned on the path, these are:
  - *independent of the observations*,
  - produced by 'thinning' a rate Ω Poisson process with probability $1 - \frac{A_{\mathbf{s}(t)\mathbf{s}(t)}}{\Omega}$ (diagonal of the transition matrix $B = (I + \frac{1}{\Omega}A)$),
  - thus, distributed according to a inhomogeneous Poisson process with piecewise constant rate $(\Omega - A_{\mathbf{s}(t)\mathbf{s}(t)})$.

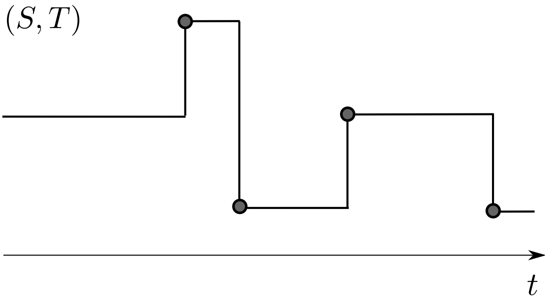
Auxiliary variable Gibbs sampler



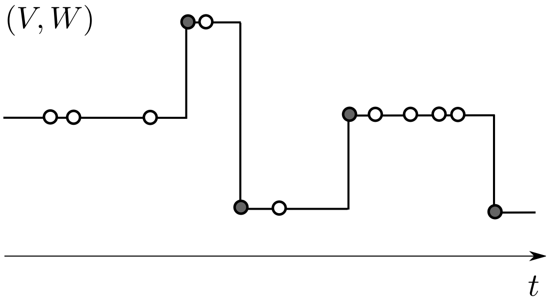
Inference via MCMC.

- Given all potential transition points, the MJP trajectory is resampled using the forward-filtering backward-sampling algorithm.
- The likelihood of the state between 2 successive points must include all observations in that interval.

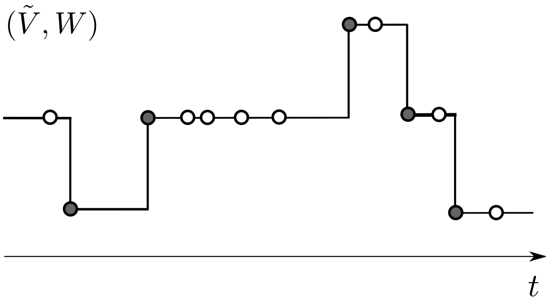
Inference via MCMC



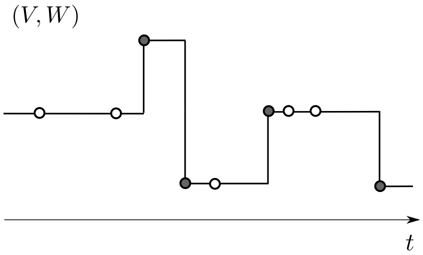
Inference via MCMC



Inference via MCMC



Comments



- Complexity: $O(n^2 P)$, where P is the (random) number of points.
- Can take advantage of sparsity in transition rate matrix A.
- Only dependence between successive samples is via the transition times of the trajectory.
- Sampler is ergodic for any $\Omega > \max_i |A_{ii}|$.
- Increasing Ω reduces this dependence, but increases computational cost.

Experiments

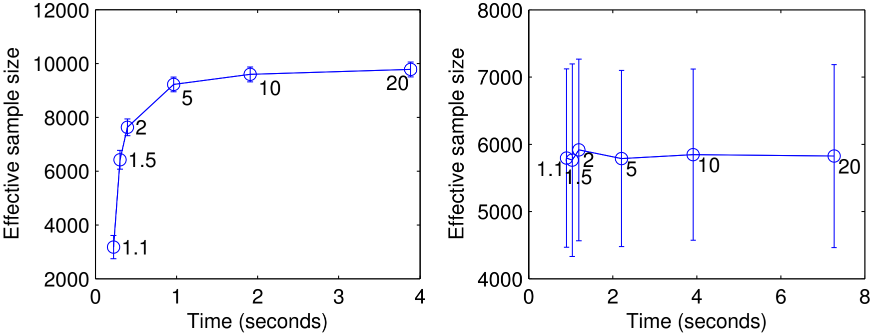


Figure: Effective sample sizes vs computation times for different settings of Ω for (left) a fixed rate matrix A and (right) Bayesian inference on the rate matrix

Experiments

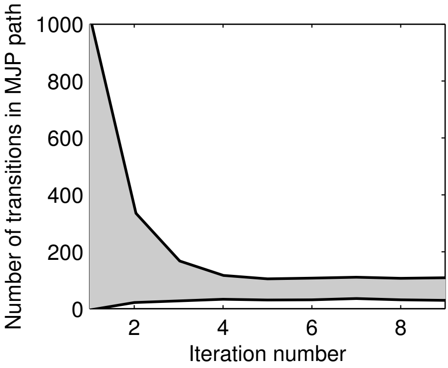


Figure: Traceplot of the number of MJP jumps for different initializations

Existing approaches to sampling

[Fearnhead and Sherlock, 2006, Hobolth and Stone, 2009] produce *independent* posterior samples, marginalizing over the infinitely many MJP paths using matrix exponentiation.

- scale as $O(n^3 + n^2 P)$.
- any structure, e.g. sparsity, in the rate matrix A cannot be exploited in matrix exponentiation.
- cannot be easily extended to complicated likelihood functions (e.g. Markov modulated Poisson processes, continuous time Bayesian networks).

Experiments

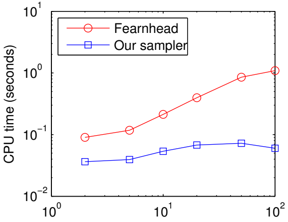


Figure: CPU time vs number of Poisson events.

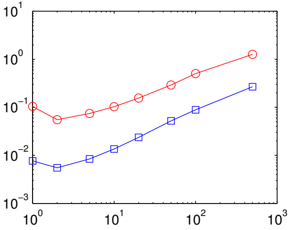


Figure: CPU time vs interval length (fixed number of events).

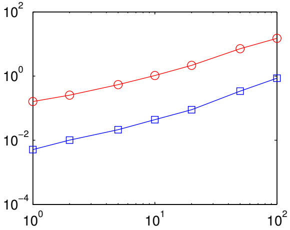


Figure: CPU time vs interval length (fixed rate).

Experiments

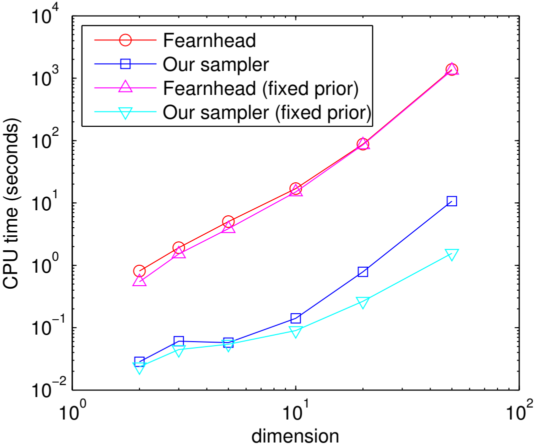


Figure: CPU time required to produce 100 effective samples as the state space of the MJP is increased

Conclusions (for part 1)

- Uniformization: sample an MJP by first sampling a Poisson process and then running a Markov chain subordinated to it.
- Inverting this generative process allows flexible posterior inference via an auxiliary variable Gibbs sampler.

The M/M/∞ queue (immigration-death process)

- M/M/∞ queue: an infinte state MJP.
- The state at any time represents the size of a population.
- The population increases with rate
  $A_{s,s+1} = \alpha, \mathcal{S} = \{1, \cdots, \infty\}$ (immigration)
- The population decreases with rate
  $A_{s,s-1} = s\beta, \mathcal{S} = \{1, \cdots, \infty\}$ (death).
- All other rates are 0.

The M/M/∞ queue (immigration-death process)

- M/M/∞ queue: an infinte state MJP.
- The state at any time represents the size of a population.
- The population increases with rate
  $A_{s,s+1} = \alpha, \mathcal{S} = \{1, \cdots, \infty\}$ (immigration)
- The population decreases with rate
  $A_{s,s-1} = s\beta, \mathcal{S} = \{1, \cdots, \infty\}$ (death).
- All other rates are 0.

- There is no upper bound on events rates in the system. This makes uniformization inapplicable.

The M/M/∞ queue (immigration-death process)

- M/M/∞ queue: an infinte state MJP.
- The state at any time represents the size of a population.
- The population increases with rate
  $A_{s,s+1} = \alpha, \mathcal{S} = \{1, \cdots, \infty\}$ (immigration)
- The population decreases with rate
  $A_{s,s-1} = s\beta, \mathcal{S} = \{1, \cdots, \infty\}$ (death).
- All other rates are 0.

- There is no upper bound on events rates in the system. This makes uniformization inapplicable.

One approach: apply uniformization to the system truncated to 50 states (the M/M/50/50 queue), with $\Omega = 2 \max |A_{ss}|$.

The M/M/∞ queue (immigration-death process)

For each leaving rate A_{ss}, define a dominating $B_{ss} \geq A_{ss} \quad \forall s$.

- We produce *candidate* event times W from B_{ss} at a higher rate than actual event rates in the system.
- We probabilistically reject (or thin) these events with probability A_{ss}/B_{ss}.

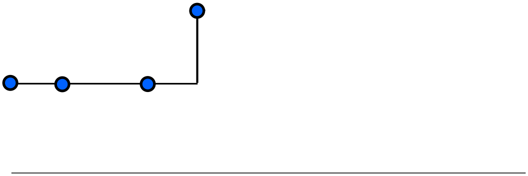
The M/M/∞ queue (immigration-death process)

The M/M/∞ queue (immigration-death process)

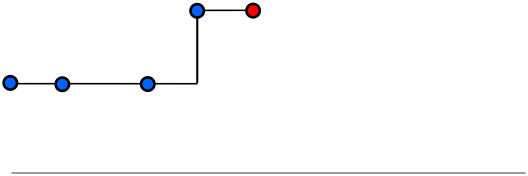
The M/M/∞ queue (immigration-death process)

The M/M/∞ queue (immigration-death process)

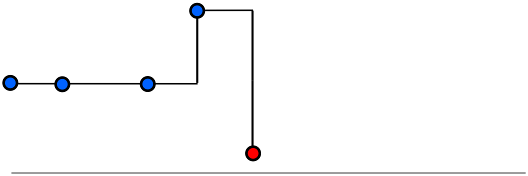
The M/M/∞ queue (immigration-death process)



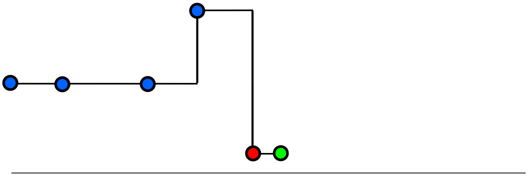
The M/M/∞ queue (immigration-death process)



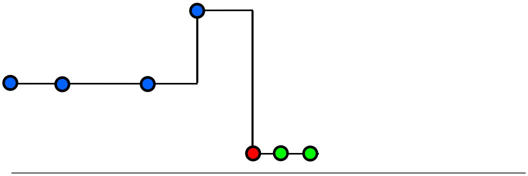
The M/M/∞ queue (immigration-death process)



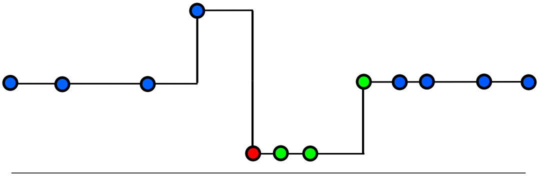
The M/M/∞ queue (immigration-death process)



The M/M/∞ queue (immigration-death process)



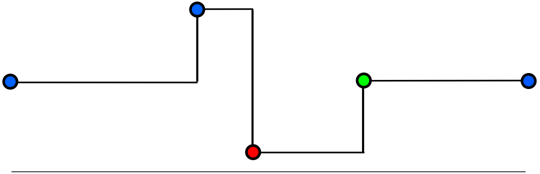
The M/M/∞ queue (immigration-death process)



Proposition

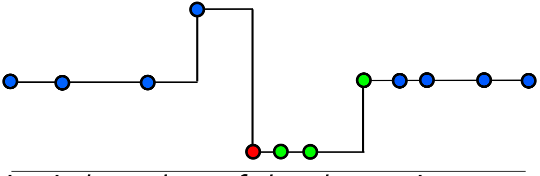
For any set of $B_{ss} \geq A_{ss}$, the (continuous time) sequence of states obtained by the process above is a sample from a MJP with initial distribution π and rate matrix A.

Resampling thinned events given system path



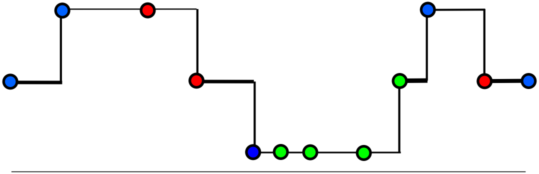
- Once again, *independent of the observations*,
- Recall that for uniformization, these are distributed according to an inhomogeneous Poisson process with piecewise constant rate $(\Omega - A_{\mathbf{s}(t)\mathbf{s}(t)})$.
- now, an inhomogeneous Poisson process with piecewise constant rate $(B_{\mathbf{s}(t)\mathbf{s}(t)} - A_{\mathbf{s}(t)\mathbf{s}(t)})$.

Resampling thinned events given system path

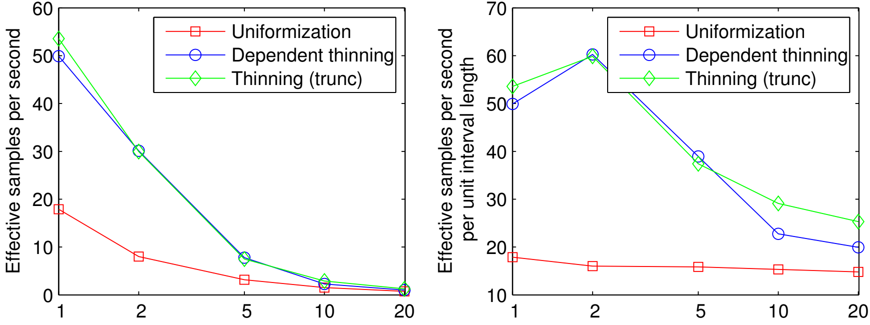


- Once again, *independent of the observations*,
- Recall that for uniformization, these are distributed according to an inhomogeneous Poisson process with piecewise constant rate $(\Omega - A_{\mathbf{s}(t)\mathbf{s}(t)})$.
- now, an inhomogeneous Poisson process with piecewise constant rate $(B_{\mathbf{s}(t)\mathbf{s}(t)} - A_{\mathbf{s}(t)\mathbf{s}(t)})$.

Resampling thinned events given system path



- Resampling a new trajectory must account for the new labels of events.
- Can easily do this be treating these as additional observation.
- However, increases coupling between new and old paths.



a) ESS per unit time b) the same, now scaled by interval length.

Scaling the overall time-discretization to rate of the most unstable state results in a fine granularity and long chains. This is inefficient, as the system typically spends less time in such states.

- Long intervals result in larger path excursions, so that larger event rates are witnessed. As our sampler adapts to this, the number of thinned events starts to become comparable, as does performance.
- But, truncating the system over long intervals can introduce biases.
- Running our sampler on the truncated system offers no real benefit.

Effect of an unstable state

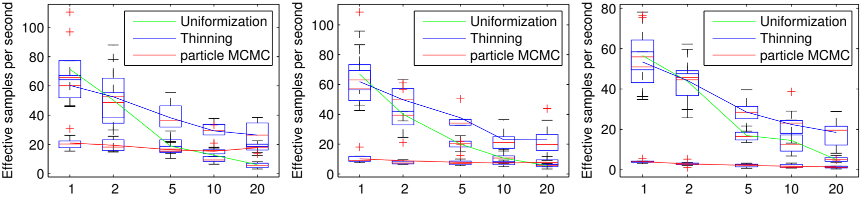


Figure: Comparison of samplers as the leaving rate γ of a state increases. Temperature decreases from left to right

Conclusions

- The idea of uniformization relates more complicated continuous time discrete state processes to the basic Poisson process.
- We demonstrated how this connection can be used to develop tractable models and efficient MCMC inference schemes.
- We have looked/ are still looking into extending the work here to:
  - semi-Markov jump processes,
  - inhomogeneous MJPs, MJPs with infinite state spaces etc,
  - continuous state diffusion processes (SDEs).
  - repulsive point processes in space (with David Dunson).
  - Dirichlet (and PY) diffusion trees (with David Knowles).
- Other applications we wish to study, such as survival analysis, queuing systems etc.

Bibliography I

Fearnhead, P. and Sherlock, C. (2006).
An exact Gibbs sampler for the Markov-modulated Poisson process.
Journal Of The Royal Statistical Society Series B, 68(5):767–784.

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.
Ann Appl Stat, 3(3):1204.

Jensen, A. (1953).
Markoff chains as an aid in the study of Markoff processes.
Skand. Aktuarietiedskr., 36:87–91.

Lewis, P. A. W. and Shedler, G. S. (1979).
Simulation of nonhomogeneous Poisson processes with degree-two exponential polynomial rate function.
Operations Research, 27(5):1026–1040.

Rao, V. and Teh, Y. W. (2011).
Fast MCMC sampling for Markov jump processes and continuous time Bayesian networks.
In *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*.