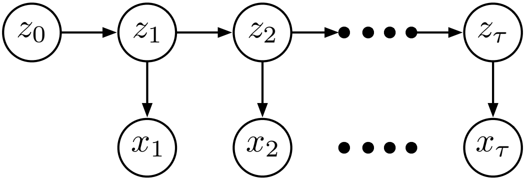
Fast MCMC sampling for Markov jump processes and extensions

Vinayak Rao and Yee Whye Teh

Rao: Department of Statistical Science, Duke University
Teh: Department of Statistics, Oxford University
Work done at: Gatsby Computational Neuroscience Unit, UCL
Supported by Gatsby Charitable Foundation

Hidden Markov Models

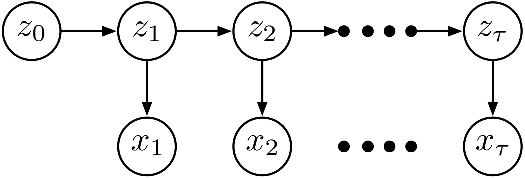


$$p(z_{t+1} = j | z_t = i) = T_{ij} \qquad p(x_t = k | z_t = i) = E_{ik}$$

- speech recognition, time series, dynamical models, natural language processing...
- efficient inference and learning: forward-backward, Baum-Welch.

Continuous-Time Hidden Markov Models

- Natural in models of physical, chemical and other continuous-time processes.



$$p(z_{t+dt} = j | z_t = i) = (1 - dt)\delta(i, j) + A_{ij} dt$$

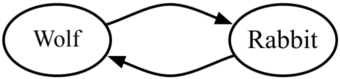
$$\begin{bmatrix} -A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & -A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & -A_{nn} \end{bmatrix}$$

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

Predator-Prey (Lotka-Volterra) Model



$$p(R_{t+dt} = r + 1 | R_t = r) = \alpha r dt$$
$$p(W_{t+dt} = w - 1 | W_t = w) = \beta w dt$$
$$p(W_{t+dt} = w + 1 | W_t = w, R_t = r) = \gamma r w dt$$
$$p(R_{t+dt} = r - 1 | W_t = w, R_t = r) = \delta r w dt$$

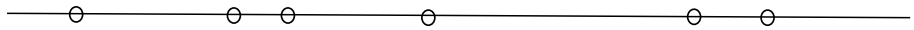
- suppose an ecologist collects data on animal populations at certain time points.
- can she infer the likely trajectories of population sizes?
- can she estimate the parameters $\alpha, \beta, \gamma, \delta$?

Overview

- The simplest example: the Poisson process on the real line.
- Markov jump processes
- Continuous time Bayesian networks.
- These relate back to the basic Poisson process via the idea of *uniformization*.
- We use this connection to develop tractable models and efficient MCMC sampling algorithms.

The Poisson process (on the real line)

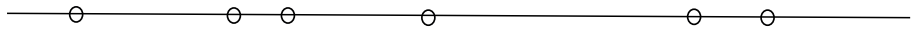
The homogeneous Poisson process with rate λ:



- the probability of an event in a small interval dt is λdt
- time between successive events has distribution $\exp(\lambda)$

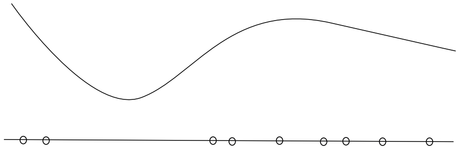
The Poisson process (on the real line)

The homogeneous Poisson process with rate λ:



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

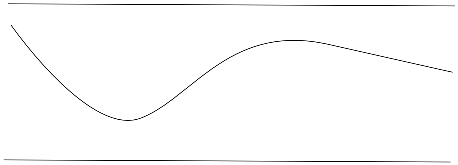
Thinning [Lewis and Shedler, 1979]

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 $\lambda(t)$.

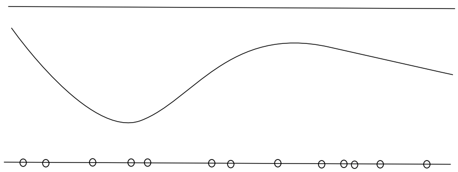
- Choose $\Omega > \lambda(t)\ \forall t$.



Thinning [Lewis and Shedler, 1979]

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

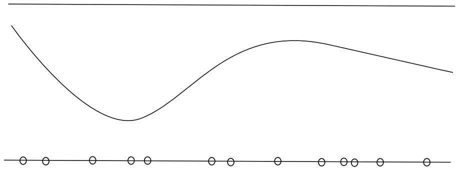
- Choose $\Omega > \lambda(t)\ \forall t$.
- Sample from a Poisson process with rate Ω.



Thinning [Lewis and Shedler, 1979]

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

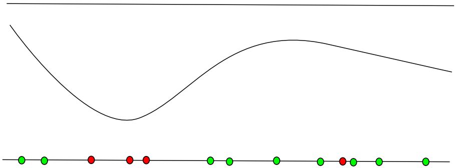
- Choose $\Omega > \lambda(t)\ \forall t$.
- Sample from a Poisson process with rate Ω.
- Keep each point with probability $\frac{\lambda(t)}{\Omega}$, otherwise 'thin'.



Thinning [Lewis and Shedler, 1979]

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

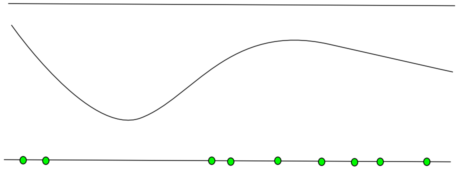
- Choose $\Omega > \lambda(t)\ \forall t$.
- Sample from a Poisson process with rate Ω.
- Keep each point with probability $\frac{\lambda(t)}{\Omega}$, otherwise 'thin'.



Thinning [Lewis and Shedler, 1979]

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

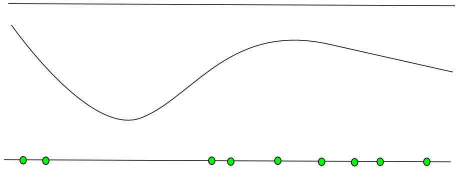
- Choose $\Omega > \lambda(t) \; \forall t$.
- Sample from a Poisson process with rate Ω.
- Keep each point with probability $\frac{\lambda(t)}{\Omega}$, otherwise 'thin'.



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'.

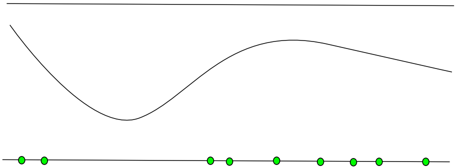


Follows from the *complete randomness* of the Poisson process.

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.

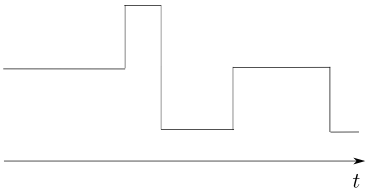
Other continuous time processes like Markov jump processes and renewal processes are *not* completely random: *Uniformization*—thin points by running a *Markov chain*.

Uniformization (at a high level)

- Define Ω larger than the fastest rate at which 'events occur'.
- Draw a set of 'potential jump times' from a Poisson process with rate Ω.
- Construct a discrete-time 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.



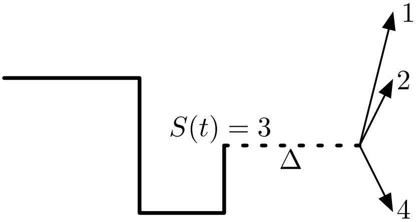
$$\begin{bmatrix} -A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & -A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & -A_{nn} \end{bmatrix}$$

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

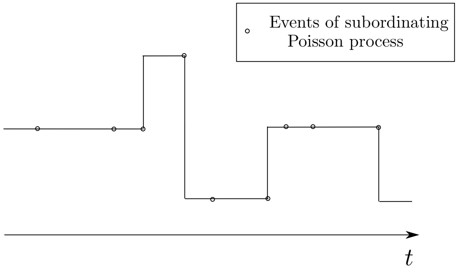
Gillespie's Algorithm



- set $t = 0$.
- draw $S(0) \sim \pi$ from the initial state distribution.
- while $t < \tau$:
  - set $i = S(t)$.
  - draw $\Delta \sim \text{Exp}(A_{ii})$.
  - set $S(t') = i$ for $t < t' < t + \Delta$.
  - set $t = t + \Delta$.
  - draw $S(t) \sim (A_{i1} \cdots A_{i,i-1}, 0, A_{i,i+1} \cdots A_{i,n})/A_{ii}$.

Uniformization for MJPs

- 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.
[Jensen, 1953]

Uniformization for MJPs [Jensen, 1953]

Proposition

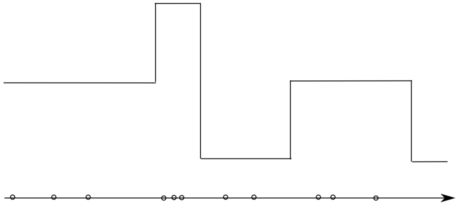
For any $\Omega \geq \max_i |A_{ii}|$*, the (continuous time) sequence of states obtained by the uniformized process is a sample from a MJP with initial distribution* π *and rate matrix* A*.*

Posterior inference

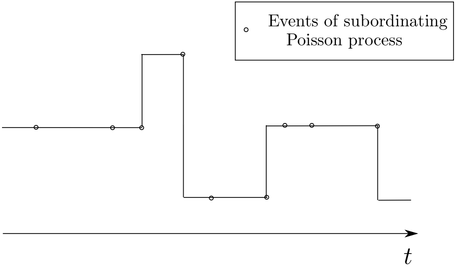
Given noisy observations of an MJP, obtain samples from the posterior.

Observations can include:

- 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 and continuous time Bayesian networks (later).

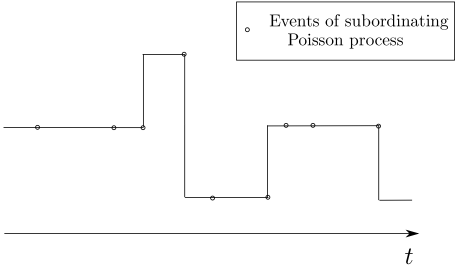


Auxiliary variable Gibbs sampler



Inference via MCMC.

Auxiliary variable Gibbs sampler



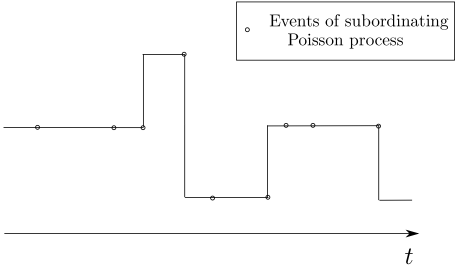
Inference via MCMC.

State space of Gibbs sampler consist of:

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

[Rao and Teh, 2011]

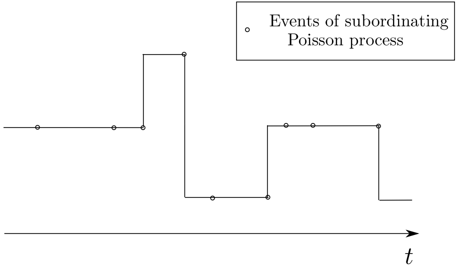
Auxiliary variable Gibbs sampler



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 $1 - \frac{A_{\mathsf{s}(t)\mathsf{s}(t)}}{\Omega}$ (diagonal of the transition matrix $B = (I + \frac{1}{\Omega}A)$),
  - thus, distributed according to a inhomogeneous Poisson process with piecewise constant rate $(\Omega - A_{\mathsf{s}(t)\mathsf{s}(t)})$.

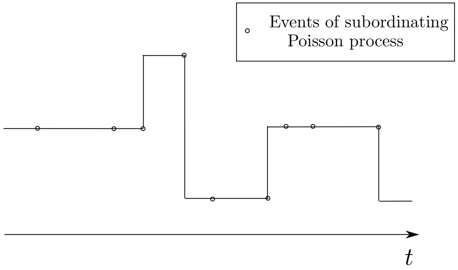
Auxiliary variable Gibbs sampler



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.

Comments



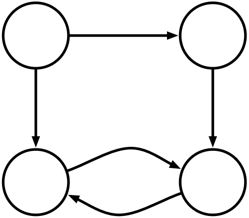
- Complexity: $O(n^2 P)$, where P is the (random) number of points.
- Can take advantage of sparsity in transition rate matrix A.
- Sampler is ergodic for any $\Omega > \max_i |A_{ii}|$.
- Only dependence between successive samples is via the transition times of the trajectory.
- Increasing Ω reduces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

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).

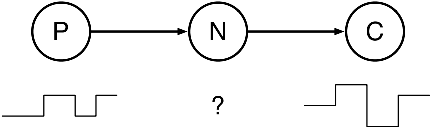
Continuous-time Bayesian networks (CTBNs)



- Compact representations of large state space MJPs with structured rate matrices.
- Applications include ecology, chemistry, network intrusion detection, human computer interaction etc.
- The rate matrix of a node at time is determined by the configuration of its parents at that time.

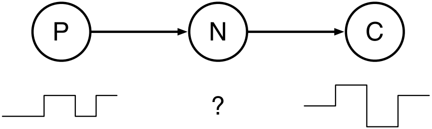
[Nodelman et al., 2002]

Gibbs sampling CTBNs via uniformization



- The trajectories of all nodes are piecewise constant.
- In a segment of constant parent (P) values, the dynamics of N are controlled by a fixed rate matrix A^P.
- Each child (C) trajectory is effectively a *continuous-time* observation when resampling the trajectory of N.

Gibbs sampling CTBNs via uniformization



- Sample potential jump times from a Poisson process with rate $\Omega^P - A^P_{ii}$.
- Between two successive potential jump times, N remains in a constant state.
  - This state must account for the likelihood of children's states.
  - The state must also explain relevant observations.
- With the resulting 'likelihood' function and transition matrix $B = (I + \frac{1}{\Omega} A^P)$, sample new trajectory using forward-filtering backward-sampling.

Existing approaches to inference

[El-Hay et al., 2008] describe a Gibbs sampler involving time discretization, which is expensive and approximate.

[Fan and Shelton, 2008] uses particle filtering which can be inaccurate for long time intervals.

[Nodelman et al., 2002, Nodelman et al., 2005, Opper and Sanguinetti, 2007, Cohn et al., 2010] use deterministic approximations (mean-field and expectation propagation) which are biased and can be inaccurate.

Experiments

- We compare our uniformization-based sampler with a state-of-the-art CTBN Gibbs sampler of [El-Hay et al., 2008]. search on the time interval.
- When comparing running times, we measured times required to produce same effective sample sizes.

Experiments

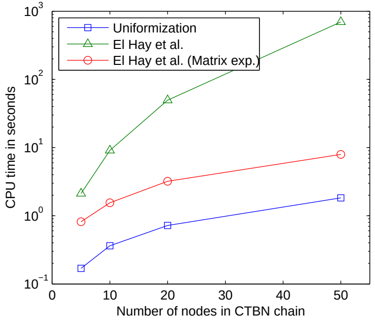


Figure: CPU time vs length of CTBN chain.

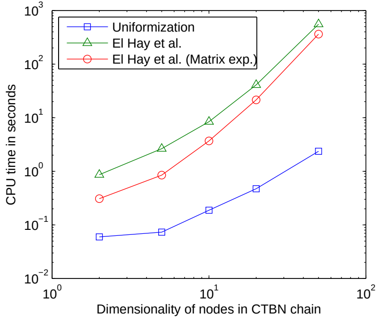


Figure: CPU time vs number of states of CTBN nodes.

The plots above were produced for a CTBN with a chain topology, increasing the number of nodes in the chain (left) and the number of states of each node (right).

Experiments

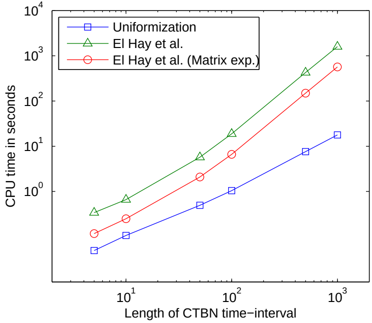


Figure: CPU time vs time interval of CTBN paths.

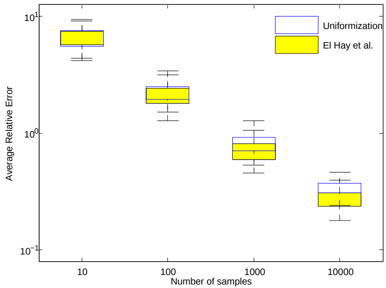
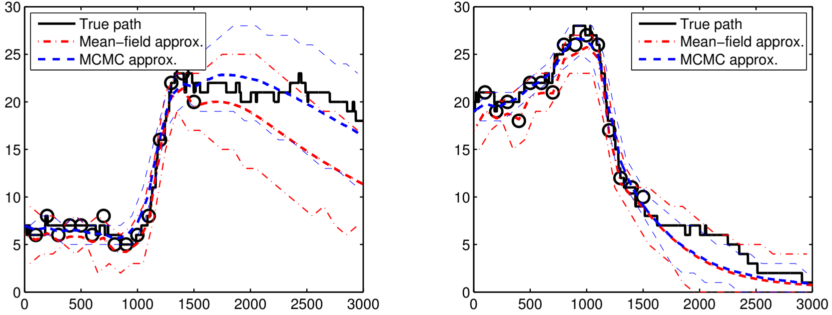


Figure: Average relative error vs number of samples

Produced for the standard 'drug network'.
Left: required CPU time as length of the time interval increases.
Right: (normalized) absolute error in estimated parameters of the network as the (absolute) number of samples increases.

Experiments

Compared against the mean-field approximation of [Opper and Sanguinetti, 2007], for the predator-prey model, a CTBN describing the Lotka-Volterra equations.



Posterior (mean and 90% confidence intervals) over predator paths (observations (circles) only until 1500).

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 extended the work here in a number of directions:
  - renewal processes (Rao and Teh NIPS 2011),
  - semi-Markov jump processes (NIPS 2012),
  - Markov-modulated Poisson processes, inhomogeneous MJPs, MJPs with infinite state spaces etc (Vinayak's thesis).
- Stochastic processes are an important mathematical language for modelling many physical and biological phenomena. There is a need for effective algorithms for inference in these models.

Bibliography I

Adams, R. P., Murray, I., and MacKay, D. J. C. (2009).
Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities.
In *Proceedings of the 26th International Conference on Machine Learning (ICML)*.

Cohn, I., El-Hay, T., Friedman, N., and Kupferman, R. (2010).
Mean field variational approximation for continuous-time bayesian networks.
J. Mach. Learn. Res., 11:2745–2783.

Cunningham, J. P., Yu, B. M., Shenoy, K. V., and Sahani, M. (2008).
Inferring neural firing rates from spike trains using Gaussian processes.
In *Advances in Neural Information Processing Systems 20*.

El-Hay, T., Friedman, N., and Kupferman, R. (2008).
Gibbs sampling in factorized continuous-time Markov processes.
In *UAI*, pages 169–178.

Fan, Y. and Shelton, C. R. (2008).
Sampling for approximate inference in continuous time Bayesian networks.
In *Tenth International Symposium on Artificial Intelligence and Mathematics*.

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.

Bibliography II

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.

Murray, I., Adams, R. P., and MacKay, D. J. (2010).
Elliptical slice sampling.
JMLR: W&CP, 9.

Nodelman, U., Koller, D., and Shelton, C. (2005).
Expectation propagation for continuous time Bayesian networks.
In *Proceedings of the Twenty-first Conference on Uncertainty in AI (UAI)*, pages 431–440, Edinburgh, Scotland, UK.

Nodelman, U., Shelton, C., and Koller, D. (2002).
Continuous time Bayesian networks.
In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 378–387.

Ogata, Y. (1981).
On Lewis' simulation method for point processes.
IEEE Transactions on Information Theory, 27(1):23–31.

Opper, M. and Sanguinetti, G. (2007).
Variational inference for Markov jump processes.
In *NIPS*.

Bibliography III

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*.

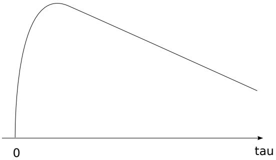
Rasmussen, C. E. and Williams, C. K. I. (2006).
Gaussian Processes for Machine Learning.
MIT Press.

Shanthikumar, J. G. (1986).
Uniformization and hybrid simulation/analytic models of renewal processes.
Oper. Res., 34:573–580.

Renewal processes

- Renewal processes: point processes on the real line ('time').
- Inter-event times drawn i.i.d. from some *renewal density*.
- Homogeneous Poisson process: exponential renewal density.
- Can capture burstiness or refractoriness.

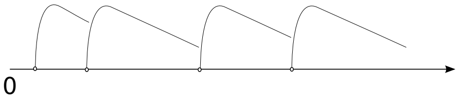
Renewal processes



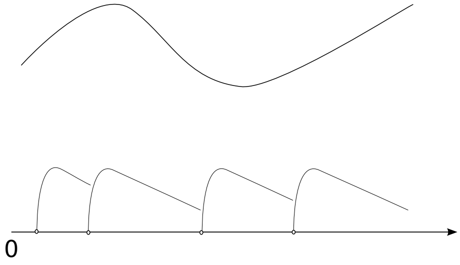
- Associated with the renewal density g is a *hazard function* h.
- For an infinitesimal Δ, $h(\tau)\Delta$ is the probability of the inter-event interval being in $[\tau, \tau + \Delta]$ conditioned on it being at least τ:

$$h(\tau) = \frac{g(\tau)}{1 - \int_0^\tau g(u)du}$$

Modulated renewal processes



Modulated renewal processes



- Modulate the hazard function by some time-varying intensity function $\lambda(t)$:

$$h(\tau, t) \equiv m(h(\tau), \lambda(t))$$

- $m(\cdot, \cdot)$ is some *interaction function*.
- We use multiplicative interactions, $h(\tau, t) = h(\tau)\lambda(t)$.

Direct sampling from prior

The modulated renewal density is:

$$g(\tau|t_{prev}) = \lambda(t_{prev} + \tau)h(\tau) \exp\left(-\int_0^\tau \lambda(t_{prev} + u)h(u)du\right)$$

where t_{prev} is the previous event time.

Naïvely, need to numerically evaluate integrals to generate samples.

- can be time consuming and introduce approximation errors.

Sampling via uniformization

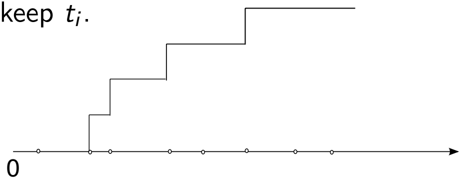
- Assume the intensity function $\lambda(t)$ and the hazard function $h(\tau)$ are bounded: $\exists\, \Omega \geq \max_{t,\tau} h(\tau)\lambda(t)$

Sampling via uniformization

- Assume the intensity function $\lambda(t)$ and the hazard function $h(\tau)$ are bounded: $\exists\, \Omega \geq \max_{t,\tau} h(\tau)\lambda(t)$
- Sample $T = \{t_0 = 0, t_1, t_2, \ldots\}$ from a rate Ω Poisson process.

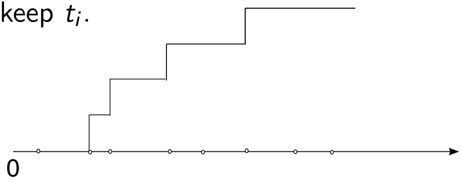
Sampling via uniformization

- Assume the intensity function $\lambda(t)$ and the hazard function $h(\tau)$ are bounded: $\exists\,\Omega \geq \max_{t,\tau} h(\tau)\lambda(t)$
- Sample $T = \{t_0 = 0, t_1, t_2, \ldots\}$ from a rate Ω Poisson process.
- Run $\{Y_0 = 0, Y_1, Y_2, \ldots\}$, an integer-valued Markov chain E
  - $Y_i = Y_{i-1} \rightarrow$ reject t_i,
  - $Y_i = i \rightarrow$ keep t_i.



Sampling via uniformization

- Assume the intensity function $\lambda(t)$ and the hazard function $h(\tau)$ are bounded: $\exists\, \Omega \geq \max_{t,\tau} h(\tau)\lambda(t)$
- Sample $T = \{t_0 = 0, t_1, t_2, \ldots\}$ from a rate Ω Poisson process.
- Run $\{Y_0 = 0, Y_1, Y_2, \ldots\}$, an integer-valued Markov chain E
  - $Y_i = Y_{i-1} \rightarrow$ reject t_i,
  - $Y_i = i \rightarrow$ keep t_i.



- For $i > j \geq 0$, define

$$p(Y_i = i | Y_{i-1} = j) = \frac{h(t_i - t_j)\lambda(t_i)}{\Omega}$$

- Define $X = \{t_i \in T \text{ s.t. } Y_i = i\}$.

Sampling via uniformization

Proposition

For any $\Omega \geq \max_{t,\tau} h(\tau)\lambda(t)$, X is a sample from a modulated renewal process with hazard $h(\cdot)$ and modulating intensity $\lambda(\cdot)$.

Generalizes [Shanthikumar, 1986] for the stationary case. See also [Ogata, 1981].

Reduction to thinning of Poisson processes

For a Poisson process, the hazard function is a constant:

$$h(\tau) = h$$

Then, the transition probabilities of the Markov chain becomes:

$$p(Y_i = i | Y_{i-1} = j) = \frac{h\lambda(t_j)}{\Omega}$$

This reduces to independent thinning [Adams et al., 2009].

Model specification

- We place a Gaussian Process prior on the intensity function $\lambda(t)$, transformed via a sigmoidal link function.
- The generative process is:

$$
\begin{aligned}
l(\cdot) &\sim \mathcal{GP}(\mu, K) \\
\lambda(\cdot) &= \hat{\lambda}\sigma(l(\cdot)) \\
X &\sim \mathscr{R}(\lambda(\cdot), h(\cdot))
\end{aligned}
$$

We use a gamma family for the hazard function:

$$h(\tau) = \frac{x^{\gamma-1}e^{-x}}{\int_x^\infty u^{\gamma-1}e^{-u}du}$$

where γ is the shape parameter.

- We place hyperpriors on $\hat{\lambda}, \gamma$ and the GP hyperparameters

Inference

Given a set of event times X, obtain samples from the modulating function $\lambda(\cdot)$ (and hyperparameters).

As before, directly sampling from the GP posterior is impossible.

Introduce the rejected events as auxiliary variables and proceed by alternately sampling the rejected events given X and the intensity function, and then the intensity function given X and rejected events.

Inference (details)

Assume the modulating function $\lambda(t)$ is known for all t.

In the interval (X_{i-1}, X_i), events from a rate Ω Poisson process were rejected with probability:

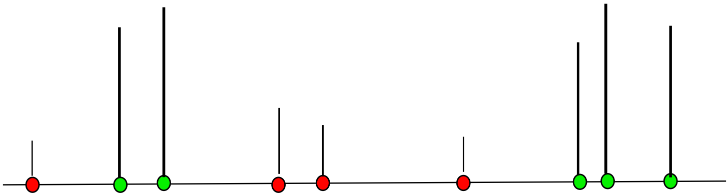
$$1 - \frac{\lambda(t)h(t - X_{i-1})}{\Omega}$$

Under the posterior, these rejected events are distributed as an inhomogeneous Poisson process with rate:

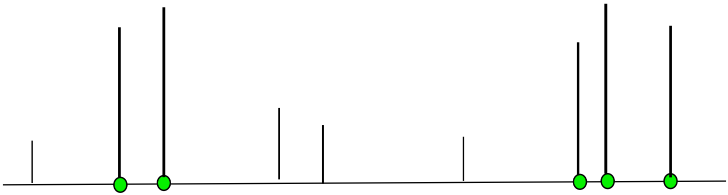
$$\Omega - \lambda(t)h(t - X_{i-1})$$

Catch: we know $\lambda(t)$ only at a discrete set of times. Use uniformization (thinning in fact). We resample the GP on the events and the rejected points using elliptical slice sampling [Murray et al., 2010].

Inference cartoon

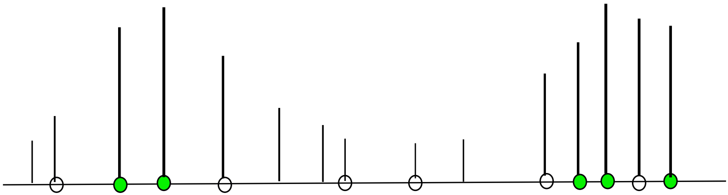


Inference cartoon

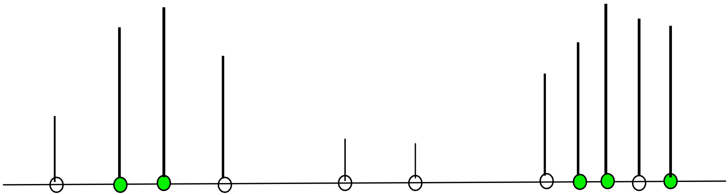


Inference cartoon

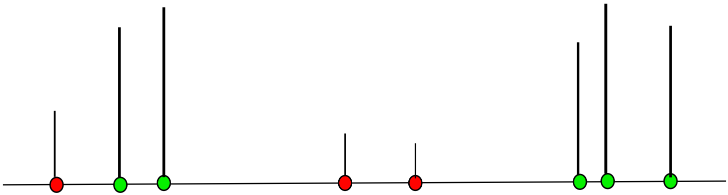
Inference cartoon



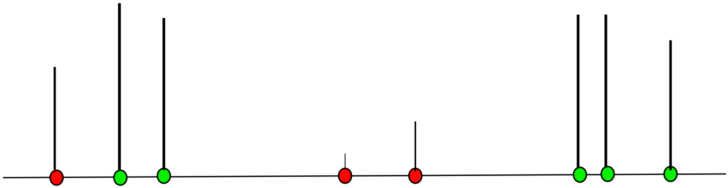
Inference cartoon



Inference cartoon



Inference cartoon



Computational considerations

- Complexity: $O(N^3)$, where $N = |X| + 2|E|$, $|X|$ is the number of observations and $|E|$ is the number of rejected points.
- For large X, we must resort to approximate inference for Gaussian processes [Rasmussen and Williams, 2006].

Experiments

We compare our uniformization based blocked Gibbs sampler with the sampler of [Adams et al., 2009].

	Synthetic dataset 1		
	Mean ESS	Minimum ESS	Time(sec)
Gibbs	93.45 ± 6.91	50.94 ± 5.21	77.85
MH	56.37 ± 10.30	19.34 ± 11.55	345.44
	Coalmine dataset		
	Mean ESS	Minimum ESS	Time(sec)
Gibbs	53.54 ± 8.15	24.87 ± 7.38	282.72
MH	47.83 ± 9.18	18.91 ± 6.45	1703

Table: Sampler comparisons. Numbers are per 1000 samples.

Besides mixing faster our sampler:

- is simpler and more natural to the problem,
- does not require any external tuning.

Algorithm 1 Blocked Gibbs sampler for GP-modulated renewal process on the interval $[0, T]$

Input: Set of event times X, set of thinned times $\tilde{X}_{prev}$ and l instantiated at $X \cup \tilde{X}_{prev}$.
Output: A new set of thinned times $\tilde{X}_{new}$ and a new instantiation $l_{X \cup \tilde{X}_{new}}$ of the $\mathcal{GP}$ on $X \cup \tilde{X}_{new}$.

1: Sample $A \subset [0, T]$ from a Poisson process with rate Ω.
2: Sample $l_A | l_{X \cup \tilde{X}_{prev}}$.
3: Thin A, keeping element $a \in A \cap [X_{i-1}, X_i]$ with probability $\left(1 - \frac{\hat{\lambda}\sigma(l(a))h(a - X_{i-1})}{\Omega}\right)$.
4: Let $\tilde{X}_{new}$ be the resulting set and $l_{\tilde{X}_{new}}$ be the restriction of l_A to this set. Discard $\tilde{X}_{prev}$ and $l_{\tilde{X}_{prev}}$.
5: Resample $l_{X \cup \tilde{X}_{new}}$ using, for example, elliptical slice sampling.