Sampling rare trajectories via sequential Monte Carlo in reverse time

Jere Koskela,
Dario Spanò, & Paul A Jenkins

MCQMC 2018, Rennes, France, 03.07.2018

Warwick Statistics

Outline

The rare trajectory problem

Sequential Monte Carlo

A generic reverse time proposal distribution

The asynchronous transfer mode network

Containment of a hyperbolic diffusion

The rare trajectory problem

- Canonical Markov chain

$$\left(\Omega := \prod_{i=0}^{\infty} \Omega_i, \mathcal{F} := \bigotimes_{i=0}^{\infty} \mathcal{F}_i, \{X_i\}_{i=0}^{\infty}, \mathbb{P}_\mu\right). \tag{1}$$

- $\mathbb{P}_\mu(x_{0:n}) := \mu(x_0) \prod_{i=1}^{n} p(x_{i-1}, x_i)$.
- Large initial set $I \subset \Omega$ with large $\mathbb{P}_\mu$-probability and $\mu(I) = 1$.
- Target set $T \subset \Omega$.
- Quantity of interest $f : \left(\prod_{i=0}^{\tau_T} \Omega_i\right) \mapsto \mathbb{R}$ which takes non-negligible values in a small, rare subset of T under the law of (1).
- A *reaction coordinate* is crucial to the success of most current methods.

Sequential Monte Carlo

- Target density: $\pi(X_{0:n}) = \mu(X_0) \prod_{i=1}^{n} p(X_{i-1}, X_i)$.
- Proposal densities: $q(X_0)$ and $q(X_{i-1}, X_i)$.
- At time $i = 0$:
  1. Sample $X_0^{(j)} \sim q(\cdot)$ for $j \in [N]$.
  2. Set $w_j \leftarrow \mu(X_0^{(j)})/q(X_0^{(j)})$ for $j \in [N]$.
- For times $i \in [n]$:
  1. Sample $a_j \sim \text{Categorical}(w_{1:N})$ for $j \in [N]$.
  2. Sample $X_i^{(j)} \sim q(X_{i-1}^{(a_j)}, \cdot)$ for $j \in [N]$.
  3. Set $w_j \leftarrow p(X_{i-1}^{(a_j)}, X_i^{(j)})/q(X_{i-1}^{(a_j)}, X_i^{(j)})$.
- $\mathbb{E}_\pi[f(X_{0:n})] \approx \sum_{j=1}^{N} \frac{w_j f(X_{0:n}^{(j)})}{\sum_{j=1}^{N} w_j}$.

Reverse time sequential Monte Carlo

- Target density: $\pi(X_{0:n}) = \mu(X_0) \prod_{i=1}^{n} p(X_{i-1}, X_i)$.
- Proposal densities: $q(X_n)$ and $q(X_i, X_{i-1})$.
- At time $i = n$:
  1. Sample $X_n^{(j)} \sim q(\cdot)$ for $j \in [N]$.
  2. Set $w_j \leftarrow 1/q(X_n^{(j)})$ for $j \in [N]$.
- For times $i \in \{n, \ldots, 1\}$:
  1. Sample $a_j \sim \text{Categorical}(w_{1:N})$ for $j \in [N]$.
  2. Sample $X_{i-1}^{(j)} \sim q(X_i^{(a_j)}, \cdot)$ for $j \in [N]$.
  3. Set $w_j \leftarrow p(X_{i-1}^{(a_j)}, X_i^{(j)})/q(X_i^{(a_j)}, X_{i-1}^{(j)})$.
- Set $w_j \leftarrow w_j \mu(X_0^{(j)})$ for $j \in [N]$.
- $\mathbb{E}_\pi[f(X_{0:n})] \approx \sum_{j=1}^{N} \frac{w_j f(X_{0:n}^{(j)})}{\sum_{j=1}^{N} w_j}$.

Reverse time proposals

- Reverse time dynamics can be characterised via Nagasawa's formula:
$$\tilde{p}(x, y) = \frac{G(\mu, y)}{G(\mu, x)} p(y, x),$$
where
$$G(\mu, x) = \mathbb{E}_\pi \left[\sum_{i=0}^{\tau_T} \mathbb{1}_{\{x\}}(X_i) \right]$$
is the occupation measure.
- Nagasawa's formula generalises detailed balance to chains without stationary distributions.
- $G(\mu, x)$ is typically at least as hard to compute as $\mathbb{E}_\pi[f]$.

Proposition 1

Consider a transition from $x_{i-1} = \{z, y\}$ to $x_i = \{z', y\}$. Suppose that the *conditional sampling distribution*

$$\pi(z|y) := \mathbb{P}_\mu(Z_i = z | Y_i = y)$$

is conditionally independent of i. Then

$$\frac{G(\mu, \{z', y\})}{G(\mu, \{z, y\})} = \frac{\pi(z'|y)}{\pi(z|y)}.$$

- $\pi(z|y)$ is of the same dimension as z, i.e. lower than $\{z, y\}$.
- The conditional time homogeneity assumption of Proposition 1 is fairly weak.
- It is possible to use the RHS to define a proposal even when Proposition 1 fails.

Asynchronous transfer mode network

- m on-off sources, each on-source generates packets at rate λ.
- Packets are served FIFO by a common server at rate μ.
- Off-sources turn on at rate α_0, on-sources turn off at rate α_1.
- State: $\{i, j\} \in \mathbb{N}_0 \times [m]$.
- Quantity of interest: probability of an initially empty queue hitting level b before emptying again, with exactly k on-sources at the hitting time.
- Initial law:

$$\mu(\{0, j\}) = \binom{m}{j} \left(\frac{\alpha_0}{\alpha_0 + \alpha_1}\right)^j \left(\frac{\alpha_1}{\alpha_0 + \alpha_1}\right)^{n-j}.$$

Asynchronous transfer mode network

- Initial set $I = \cup_{j=0}^{m} \{(0, j)\}$.
- Terminal set $T = I \cup \left(\cup_{j=0}^{m} \{(b, j)\} \right)$.
- Quantity of interest $f(i, j) := \mathbb{1}_{\{(b,k)\}}(i, j)$.

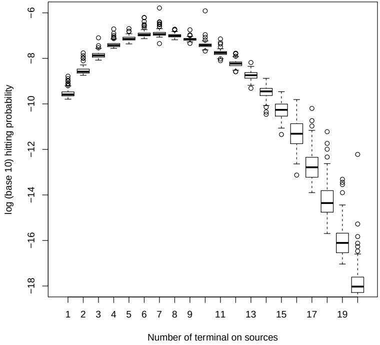
The ATM network proposal

- Single transitions only change either i or j, not both.
- $\Rightarrow$ Only need approximate CSDs $\hat{\pi}_i(i|j)$ and $\hat{\pi}_j(j|i)$
- Subsequent results use

$$\hat{\pi}_i(i|j) \propto \left(\frac{\lambda \max\{j,1\}}{\mu}\right)^i,$$

$$\hat{\pi}_j(j|i) \propto \hat{\pi}_i(i|j)\binom{m}{j}\left(\frac{\alpha_0}{\alpha_0+\alpha_1}\right)^j\left(\frac{\alpha_1}{\alpha_0+\alpha_1}\right)^{m-j}.$$

ATM results I



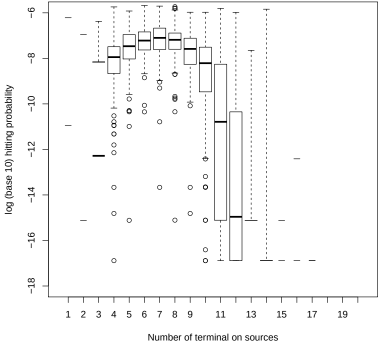


Figure: Simulated hitting probabilities using reverse time SMC (left) and adaptive multilevel splitting (right) with parameters $m = 20$, $b = 10$, $\lambda = 0.5$, $\mu = 10.0$, $\alpha_0 = 1.0$, $\alpha_1 = 3.0$. Run times are comparable at 130 seconds per replicate each.

ATM results II

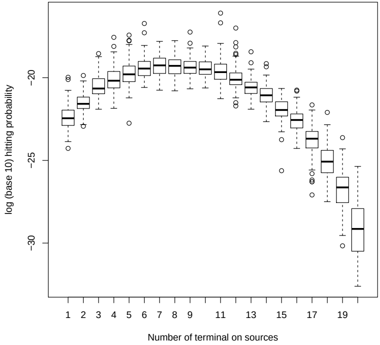


Figure: Simulated hitting probabilities using reverse time SMC with parameters $m = 20$, $b = 30$, $\lambda = 0.5$, $\mu = 10.0$, $\alpha_0 = 1.0$, $\alpha_1 = 3.0$, and 10 000 particles. The run time was 13 minutes per replicate.

The hyperbolic diffusion

- The hyperbolic diffusion solves

$$dX_t = \frac{-X_t}{\sqrt{1+X_t^2}} dt + dW_t.$$

- Stationary density $\pi(x) \propto e^{-\sqrt{1+x^2}}$.
- Intractable transition density.
- Euler discretisation:

$$\begin{aligned} &P_\Delta((m,x),(n,y)) \\ &= \frac{\delta_{m+\Delta}(n)}{\sqrt{2\pi\Delta}} \exp\left(-\frac{1}{2\Delta}\left[y - x\left\{1 - \frac{\Delta}{\sqrt{1+x^2}}\right\}\right]^2\right). \end{aligned}$$

The hyperbolic diffusion

- Interested in a stationary trajectory initially in (l_0, u_0) at time 0, hitting (l_t, u_t) at time t, and remaining contained in the linearly interpolated strip:

$$
\begin{aligned}
I &= \{0\} \times (l_0, u_0), \\
T &= \bigcup_{s \in (0,t]} \left(\{s\} \times \left\{ \frac{l_t - l_0}{t} s + l_0, \frac{u_t - u_0}{t} s + u_0 \right\} \right), \\
f(X_{0:\tau_T}) &= \mathbb{1}_{\{t\}}(\tau_T), \\
\mu &= \pi|_{(l_0, u_0)}, \\
G(\mu, x) &= \pi(x).
\end{aligned}
$$

Hyperbolic diffusion results

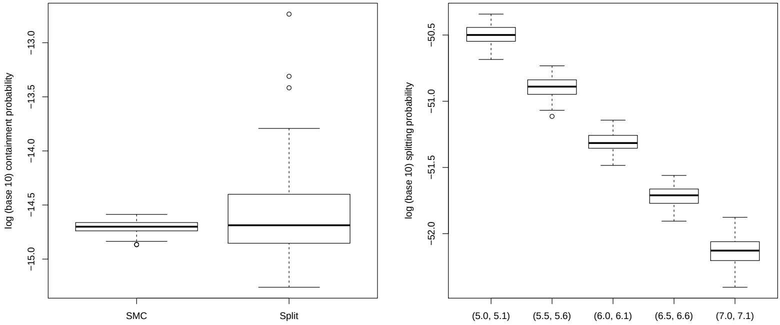


Figure: (Left) 100 replicates of simulated containment probabilities with $(l_0, u_0) = (-1, 1)$, $t = 2$, and $(l_t, u_t) = (5, 5.1)$. (Right) Containment probabilities with various terminal conditions.

Summary

- Aim: Sample trajectories from a large, common initial set I until the first hitting time of a target set T.
- We are interested in $\mathbb{E}_{\pi}[f]$, where f charges the whole trajectory but takes non-negligible values in a small, rare subset of T.
- Forwards-in-time proposals often make opposite assumptions: I is a singleton and the effective support of f may be large.
- The ingredients $\hat{\pi}(\cdot|\cdot)$ and $p(x, y)$ only require local information, not the conditional law given the rare event.
- Multilevel splitting and forwards-in-time SMC need global reaction coordinates, or approximations of the conditional target dynamics given the rare event.
- The CSD $\hat{\pi}(\cdot|\cdot)$ is lower dimensional than the process of interest, which helps with design.

References

V Anantharam, P Heidelberg, & P Tsoucas.
Analysis of rare events in continuous time Markov chains via time reversal and fluid approximation.
IBM Research Report, #RC 16280, 1990.

F Cérou, P Del Moral, T Furon, & A Guyader.
Sequential Monte Carlo for rare event estimation.
Stat. Comput., 22:795–808, 2012.

M R Frater, R A Kennedy, & B D O Anderson.
Reverse-time modelling, optimal control and large deviations.
Syst. Control Lett., 12:351–356, 1989.

M R Frater, R R Bitmean, R A Kennedy, & B D O Anderson.
Fast simulation of rare events using reverse-time models.
Comput. Networks ISDN, 20:315–321, 1990.

C Jarzynski.
Rare events and the convergence of exponentially averaged work values.
Phys. Rev. E, 73:046105, 2006.

J Koskela, D Spanò, & P A Jenkins.
Inference and rare event simulation for stopped Markov processes via reverse-time sequential Monte Carlo.
Stat. Comput., 28(1):131–144, 2018.

A Shwartz & A Weiss
Induced rare events: analysis via large deviations and time reversal.
Adv. Appl. Probab., 25(3):667–689, 1993.