Resampled Scalable Langevin Exact Method

A method to simulate from an intractable distribution

Divakar Kumar, Dr. Murray Pollock, Prof. Gareth Roberts
Department of statistics, University of Warwick
D.Kumar@warwick.ac.uk

Abstract
The *Scalable Langevin Exact (ScaLE)* is a diffusion based approach to simulate from an intractable distribution. The ScaLE is a recent alternative to gradient based Langevin MCMC schemes such as MALA which circumvents the need to use Metropolis type correction . Consequently, due to this circumvention this methodology is highly applicable to 'Big Data' problems. The ScaLE method approximates the intractable distribution of interest with the *quasi-stationary* distribution of a 'killed' *Brownian motion*. ReScaLE method uses a different simulation mechanism to sample from the quasi-stationary density of a 'killed' Brownian motion.

Introduction

Re-sampled Scalable Langevin Exact (ReScaLE) method is a novel method of sampling from an intractable distribution of interest. Contrary to ScaLE, ReScaLE method uses an alternative approach to simulate from the quasi-stationary distribution of a killed Brownian motion whose invariant distribution is given by concerned intractable distribution. The aim of this poster is to furnish an introduction ReScaLE methodology as a tool to sample from an intractable distribution of interest. ReScaLE method uses a recent advancement in the quasi-stationary literature proposed by Glynn & Blanchet.

Main Objective

- **How to simulate from an intractable distribution π ?.**
- **Simulate the stationary distribution of Langevin diffusion**

$$dX_t = \frac{1}{2}\nabla \log \pi(X_t)dt + dB_t, \quad X_0 = x_0, t \in [0, T]. \tag{1}$$

Path-Space Rejection Sampling for diffusion

1. Propose path X from a measure $\mathbb{W}$ for the target measure $\mathbb{Q}$ such that $\frac{d\mathbb{Q}}{d\mathbb{W}}(X) \le M$.
2. Accept the path X with probability

$$P_{\mathbb{W}}(X) := \frac{1}{M}\frac{d\mathbb{Q}}{d\mathbb{W}}(X) \tag{2}$$

Exactly simulating the trajectories of Langevin diffusion

$$p_{0,t}(\cdot, y) = w_{0,t}(\cdot, y)\mathbf{E}_{\mathbb{W}_{|X_t=y}}\left(\frac{d\mathbb{Q}}{d\mathbb{W}}(X)\right) \tag{3}$$

For $\mu(x) = \frac{1}{2}\nabla \log \pi(x)$, the transition density is:

$$p_{0,t}(x_0 = 0, x) \propto \exp\left\{-\frac{x^2}{2t}\right\}\{\pi(x)\}^{\frac{1}{2}}\mathbb{E}_{x_0,x}\left(\exp\left\{-\int_0^t \phi_\mu(X_s)ds\right\}\right) \longrightarrow \pi. \tag{4}$$

where

$$l := \inf_x \frac{\mu^2 + \mu'}{2}(x) \quad \phi_\mu(X_s) := \frac{(\mu(X_s)^2 + \mu'(X_s))}{2} - l \tag{5}$$

Double the drift! - Drop $\pi(x)$

$$p_{0,t}(x_0 = 0, x) \propto \exp\left\{-\frac{x^2}{2t}\right\}\{\pi(x)\}\mathbb{E}_{x_0,x}\left(\exp\left\{-\int_0^t \phi_{2\mu}(X_s)ds\right\}\right) \longrightarrow \pi^2. \tag{6}$$

Killed Brownian Motion

The quasi-stationary density of a killed Brownian motion with killing rate $\phi_{2\mu}$ is

$$q_{0,t}(0, x) := \exp\left\{-\frac{x^2}{2t}\right\}\mathbb{E}_{x_0,x}\left(\exp\left\{-\int_0^t \phi_{2\mu}(X_s)ds\right\}\right). \tag{7}$$

Problems:

1. Problem-1: How to continuously sample trajectory of a Brownian motion?
2. Problem-2: How to simulate the quasi-stationary density of a killed Brownian motion?

Sampling from the QSD of Brownian motion

Result - 1:

Let $\tau_1, ..., \tau_k$ be the Poisson process with rate M where M is such that $\sup_x \phi(x) \le M$. Let $X_{\tau_1}, ..., X_{\tau_k}$ be the realised skeleton of a Brownian motion $\{X_t : t \ge 0\}$ at times $\tau_1, ..., \tau_k$. If process is killed at τ_j with probability $\frac{\phi(X_{\tau_j})}{M}$. Then,

$$\mathbb{P}(\text{Process survived until time}t) = \exp\left\{-\int_0^t \phi(X_s)ds\right\}$$

Result - 2: Glynn & Blanchet's approach of estimating the QSD

1. Initialize the probability vector $\pi = \pi_0$ on the non-absorbing states of Markov chain.
2. Select a non-absorbing state of the Markov chain x_0 and set $X_0 = x_0$.
3. Simulate the Markov chain normally starting with X_0 until absorption. Update $\pi$by counting the number of visits to each state until absorption.
4. Choose an initial position according to normalized vector π and goto step 3.
5. Steps 3. and 4. are repeated many times to get an estimate of quasi-stationary dist.

ReScaLE method - Pseudocode

Algorithm 0.1: ReScaLE Algorithm(μ, x_0)

1. $l \leftarrow \inf_{x\in\mathbb{R}} \frac{\mu^2+\mu'}{2}, \phi \leftarrow \frac{\mu^2+\mu'}{2} - l, M \leftarrow \sup_{x\in\mathbb{R}} \phi(x)$
2. $t_0 \leftarrow 0; X_{t_0} \leftarrow x_0$
3. **do** $\begin{cases} (t_1, t_2, ...) \sim \text{Poisson Process of rate } M \text{ starting at } t_0 \\ (X_{t_1}, X_{t_2}, ...) \sim \text{Brownian Motion started at position } X_{t_0} \\ \text{Kill the process at } X_{t_i} \text{ with probability } \phi(X_{t_i})/M \\ \textbf{exit} \text{ once kill occurs} \end{cases}$
4. **starting time** $\sim U[0, \mathbf{t_{kill}}]$
5. **starting value** $\sim$ Brownian Bridge conditioned on neighbors of **starting time**
6. **GOTO** 2. with $t_0 \leftarrow \mathbf{t_{kill}}; X_{t_0} \leftarrow$ **starting value**

return $((X_{t_1}, X_{t_2}, ...))$

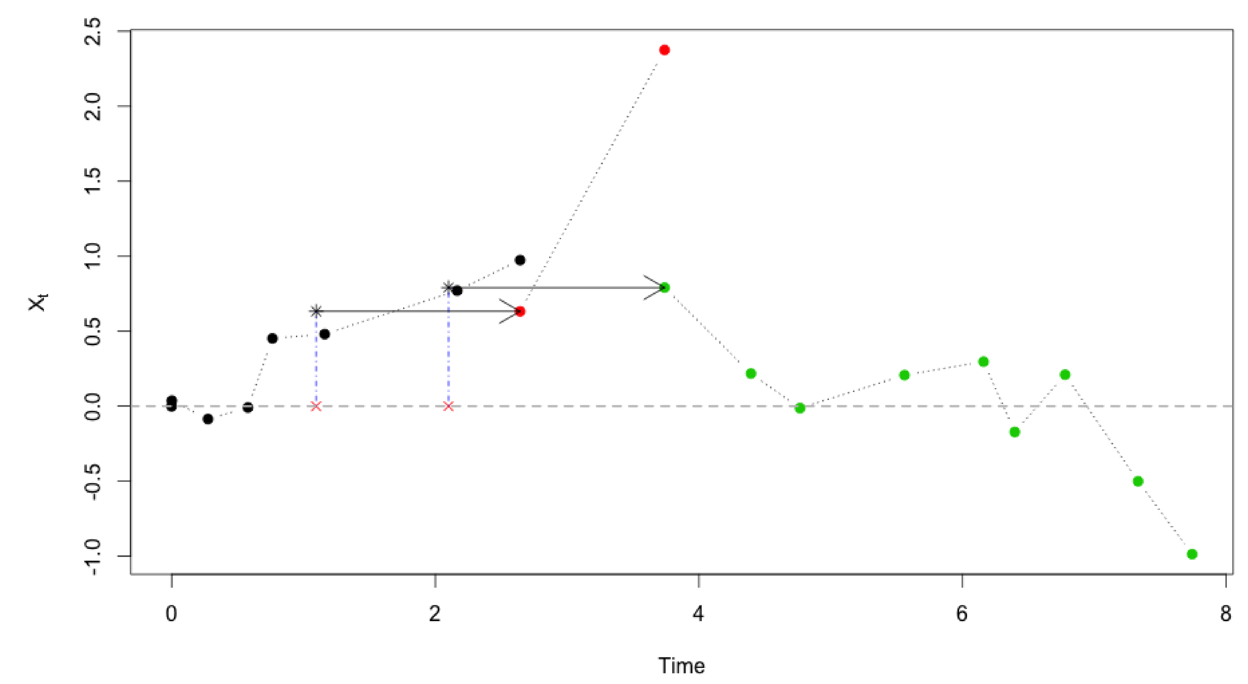
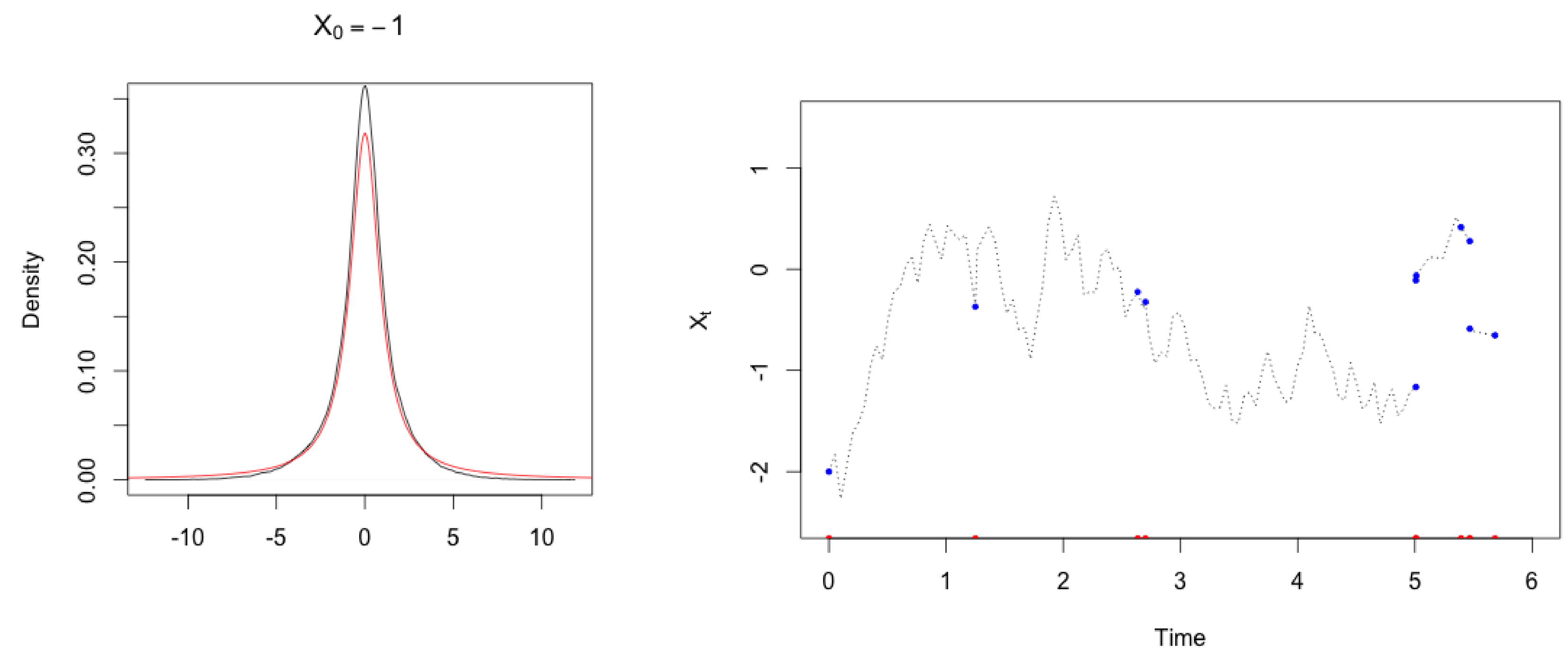


Figure 1: An illustration of ReScaLE methodology

An application of ReScaLE methodology to sample from Cauchy density



Current challenges and further research

- No formal proof exists for the regenerative algorithm by Glynn and Blanchet for CTMC on general state space.
- Current ReScaLE method shows persistence issues.
- How to make the method 'adaptive'?
- How to 'speed-up' the method for faster convergence to quasi-stationary density?

References

[1] A. Beskos, O. Papaspiliopoulos, and G. O. Roberts. Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6):1077–1098, 2006.

[2] Jose Blanchet, Peter Glynn, and Shuheng Zheng. Empirical Analysis of a Stochastic Approximation Approach for Computing Quasi-stationary Distributions. *Evolve - A bridge between Probability*, pages 19–37, 2012.

[3] Gareth O. Roberts, Murray Pollock, Paul Fearnhead, and Adam M Johansen. An Unbiased and Scalable Monte Carlo Method for Bayesian Inference for Big Data. *Upcoming paper*, (2):1–20, 2016.

[4] Gareth O Roberts and Richard L Tweedie. Exponential Convergence of Langevin Distributions and Their Discrete Approximations. *Bernoulli*, 2(4):pp. 341–363, 1996.

Acknowledgements

This is a joint work with Dr. Murray Pollock & Prof. Gareth Roberts under the Oxford-Warwick Statistics Program (OxWaSP) - a joint DPhil program being run between the department of statistics, University of Oxford and University of Warwick, UK.