Regenerating Scalable Langevin Exact Method

An application to Big Data problems

Divakar Kumar, Murray Pollock, Gareth Roberts

Department of statistics, University of Warwick

D.Kumar@warwick.ac.uk

Main Objective

- **How to simulate from an intractable distribution** $\pi \propto \prod_{i=1}^{N} \pi_i$?.
- **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 $\frac{d\mathbb{Q}}{d\mathbb{W}}(X) \leq M$.
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 sampling 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 $\phi_{2\mu}(X_t)$ can be interpreted as the state-dependent 'killing' rate of a Brownian motion. The density of a *killed Brownian motion* conditioned on its survival is called the quasi-stationary density. The quasi-stationary density of a killed Brownian motion with killing rate $\phi_{2\mu}$ is

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

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?
3. Problem-3: It is difficult to unveil the sample path of a Brownian motion conditioned on its survival until large time t.

Sampling from the QSD of Brownian motion

- ScaLE method uses SMC-based approach to simulate from the quasi-stationary density of a 'killed' Brownian motion.

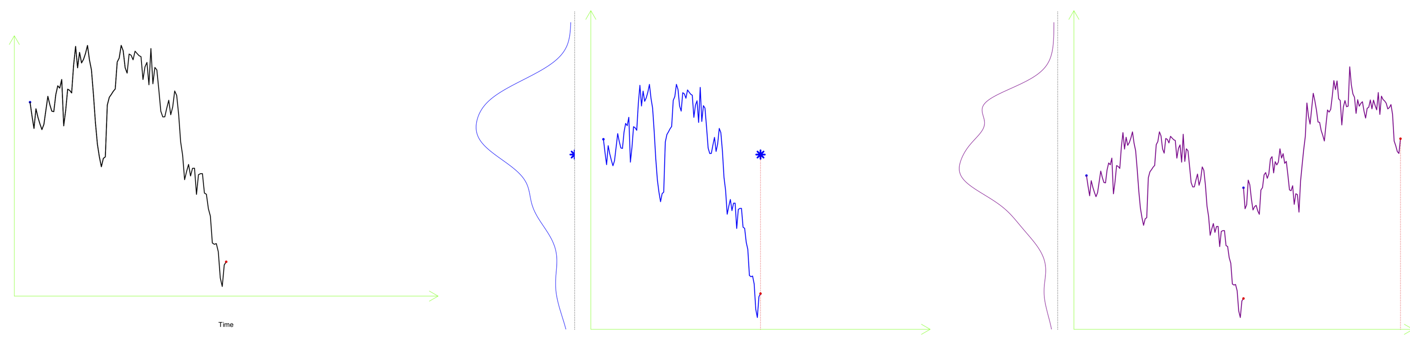
Result - 1: Poisson Thinning

Let $\tau_1, ..., \tau_k$ be the Poisson process with rate M where M is such that $\sup_x \phi(x) \leq M$. Let $X_{\tau_1}, ..., X_{\tau_k}$ be the realised skeleton of a Brownian motion $\{X_t : t \geq 0\}$ 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

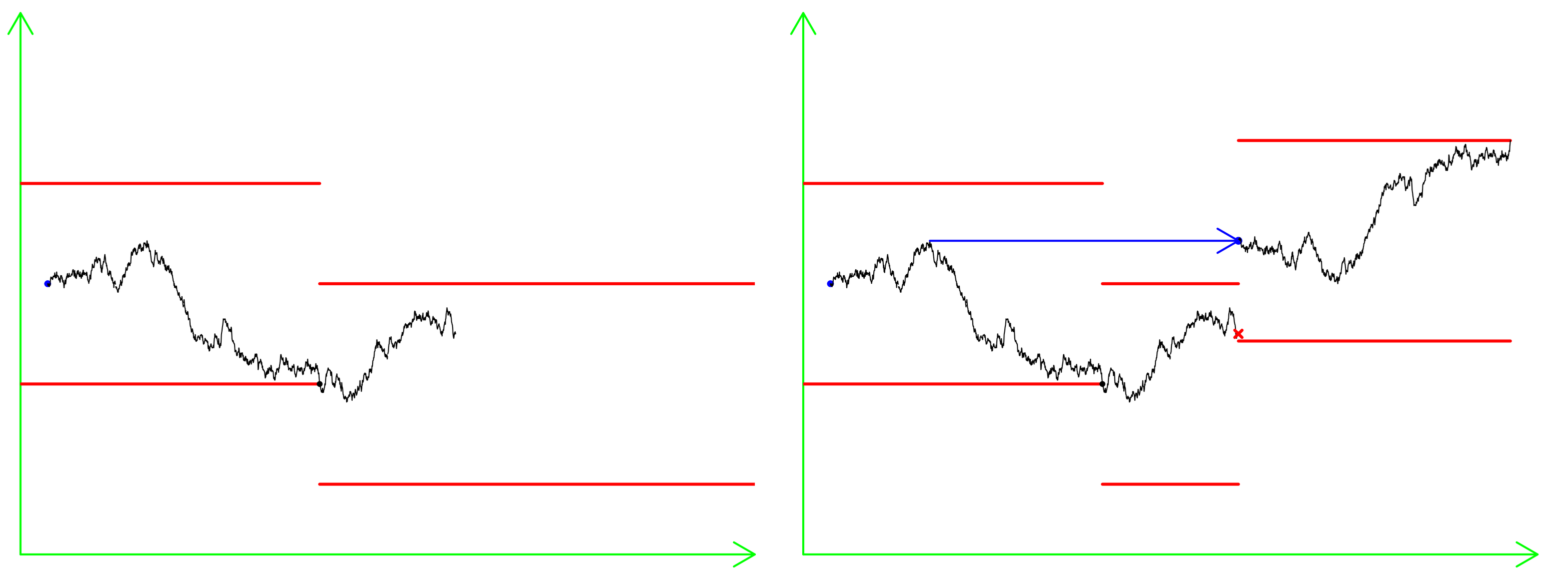
1. Initialize the probability vector $\pi = \pi_0$ on the non-absorbing states of Markov chain.
2. Select a non-absorbing state of the Markov chain x_0 and set $X_0 = x_0$.
3. Simulate the Markov chain normally starting with X_0 until absorption. Update $\pi$by counting the number of visits to each state 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.



Big data setting – $\mathcal{O}(1)$ computational cost

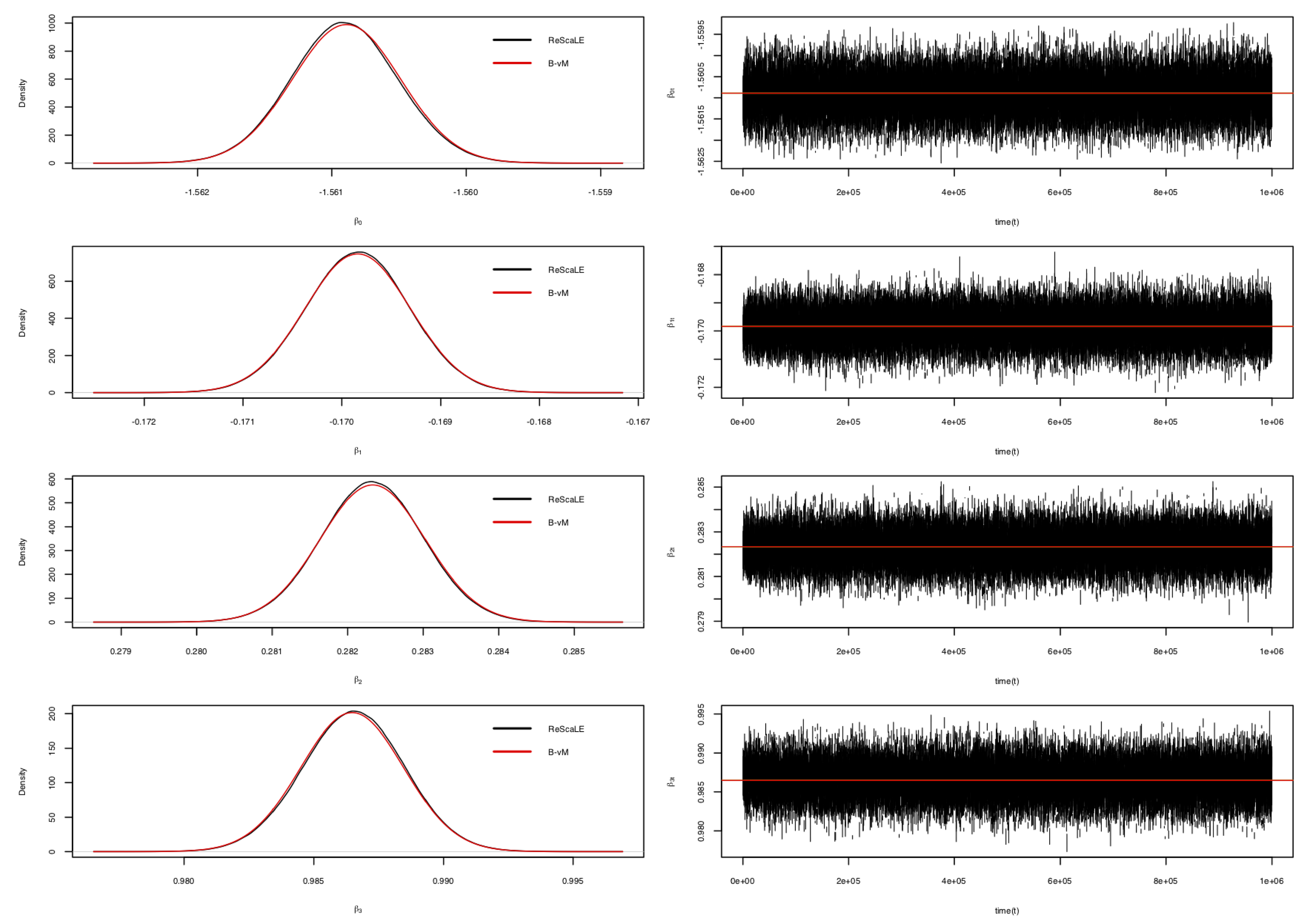
Replace the expensive ϕ function by an unbiased estimators $\hat{\phi}$, which are cheaper to evaluate.

An illustration of the ReScaLE method

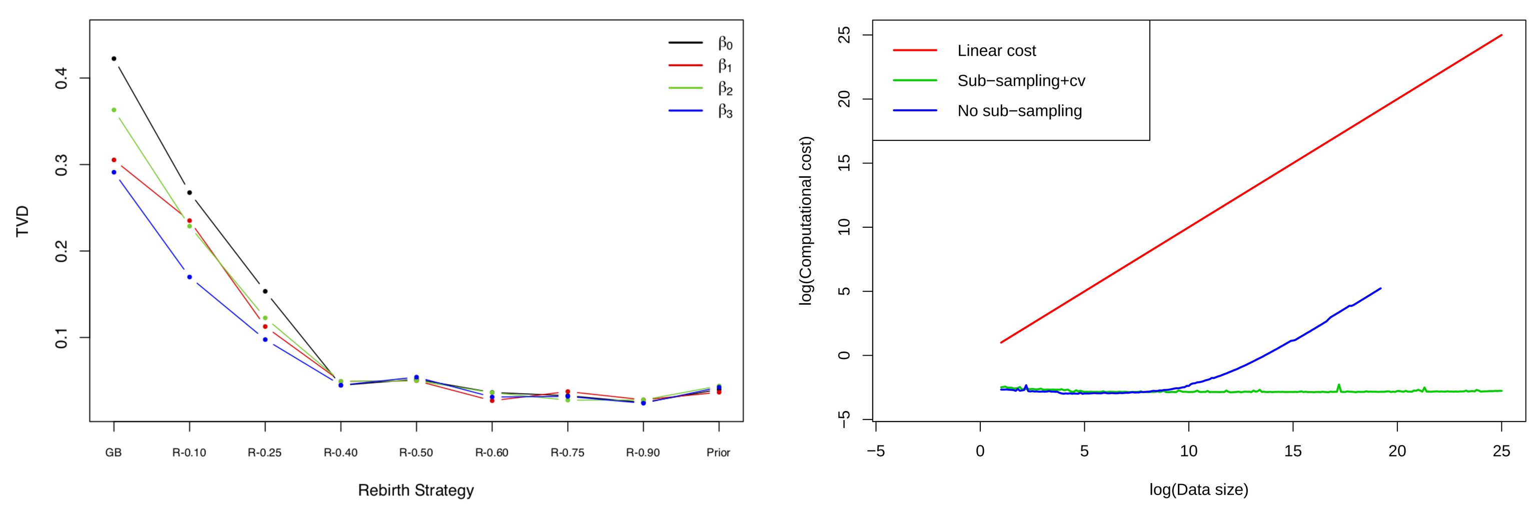


A Big Data Example: The US Domestic Airline data

- **Data:** is of $>$ 12 GB in size with more than 120 million observations.
- **Model:** Logistic model to predict whether a given flight is `delayed` or not, given that whether it is a `weekend` flight , whether flight operates during `night`, and its `distance` from the origin.



Different rebirth strategies & $\mathcal{O}(1)$ computational cost



Current challenges and further research

- No formal proof exists for the regenerative algorithm by Glynn and Blanchet for CTMC on general state space.
- How to make the method 'adaptive' and '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] J. Blanchet, P. Glynn, and S. Zheng. Analysis of a stochastic approximation algorithm for computing quasi-stationary distributions. *Adv. in Appl. Probab.*, 48(3):792–811, 09 2016.

[3] M. Pollock, P. Fearnhead, A. M. Johansen, and G. O. Roberts. The Scalable Langevin Exact Algorithm: Bayesian Inference for Big Data. *ArXiv e-prints*, September 2016.

Acknowledgment

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.