Quasistationary Monte Carlo Methods: Foundations and Stochastic Approximation

Author: Andi Wang (`andi.wang@spc.ox.ac.uk`)

Supervisors: Prof. David Steinsaltz (Oxford), Prof. Gareth Roberts (Warwick)

UNIVERSITY OF OXFORD

DEPARTMENT OF STATISTICS

Motivation: why quasistationary Monte Carlo?

Performing Bayesian inference in a 'Big Data' setting is a challenging problem. A modern researcher may have a posterior distribution of the form $\pi(x) \propto \prod_{i=0}^N f_i(x)$ with N incredibly large: in such a setting merely evaluating the posterior density is an inhibitive $O(N)$ calculation. The recent Scalable Langevin Exact (ScaLE) Algorithm [2] proposes a new paradigm for Big-Data Bayesian inference: by constructing a killed diffusion with the appropriate *quasistationary distribution* and utilising *subsampling* techniques, this novel algorithm attains an $O(1)$ runtime in the size of the data and retains *exactness* of the target distribution. In this poster we give theoretical results underpinning quasistationary Monte Carlo methods and describe an alternative stochastic approximation method, known as Regenerating ScaLE (ReScaLE).

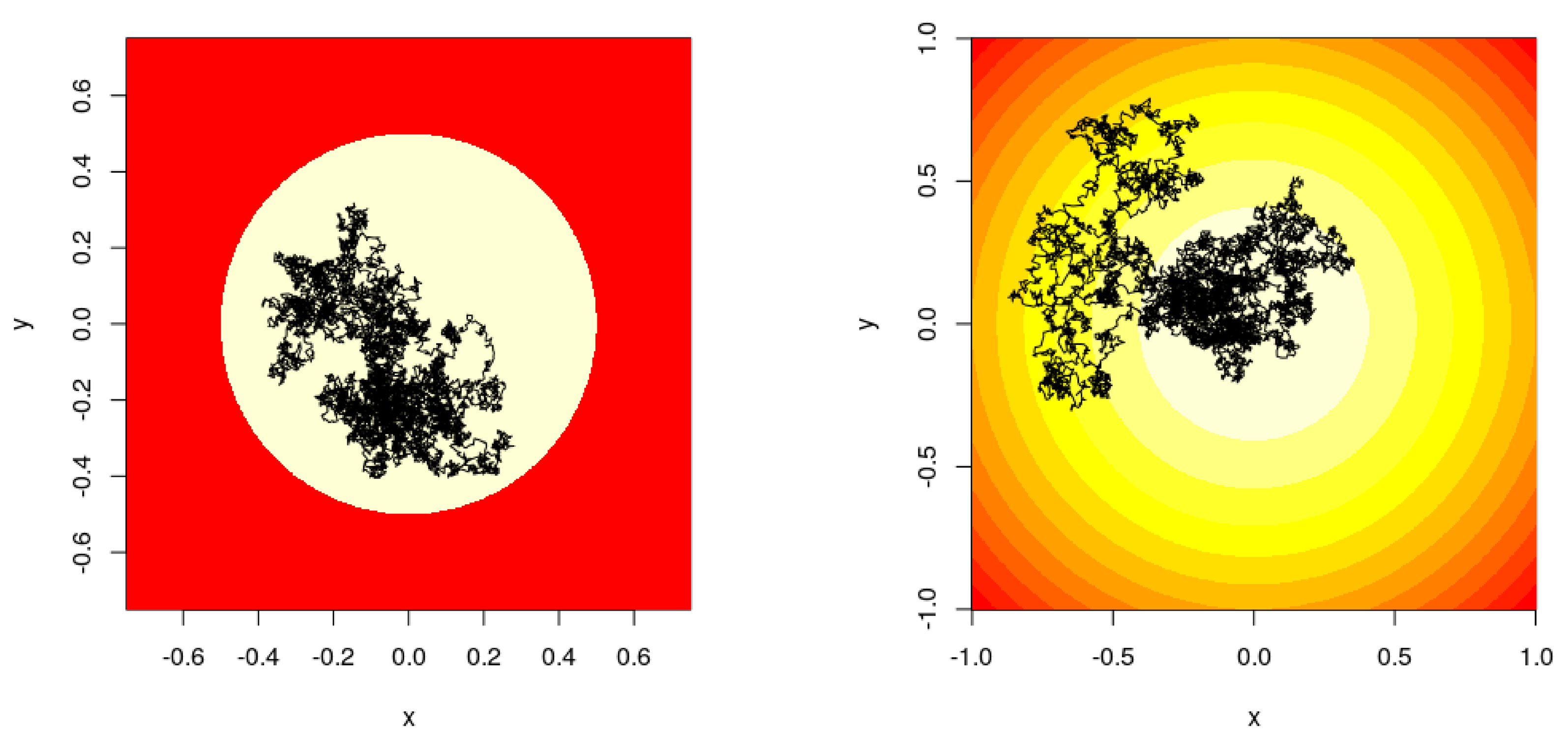
Introduction: what is quasistationary Monte Carlo?

Consider the diffusion defined through the stochastic differential equation (SDE)

$$dX_t = \nabla A(X_t)dt + dW_t, \quad X_0 = x_0 \in \mathbb{R}^d \tag{1}$$

where W is a standard d-dimensional Brownian motion, $A : \mathbb{R}^d \to \mathbb{R}$ is some smooth function.

Imagine an ant on a volcanic island, evolving as a diffusion (X_t), say a Brownian motion, which is incinerated at $\tau_\partial :=$ *the moment it touches lava*, Figure 1a. Given it has survived for a long time, where is it likely to be? That is, what can we say about $\mathbb{P}(X_T \in \cdot \,|\tau_\partial > T)$, for large T?



(a) Brownian motion with boundary killing.

(b) Brownian motion killed at rate $\kappa(x) = 10\|x\|^2$.

Figure 1: Examples of *killed diffusions*, conditioned to survive until $T = 1$.

Let's fix a *killing rate* $\kappa : \mathbb{R}^d \to [0, \infty)$, and now the ant spontaneously combusts at

$$\tau_\partial := \inf\left\{t \geq 0 : \int_0^t \kappa(X_s)\,ds > \xi\right\}, \quad \xi \sim \text{Exp}(1) \text{ is independent.}$$

See Figure 1b. Again, let's consider

$$\mathbb{P}(X_T \in A\,|\tau_\partial > T) \quad \text{as } T \to \infty \tag{2}$$

for an arbitrary measurable set $A \subset \mathbb{R}^d$. Does this converge to $\mu(A)$ for some distribution μ? If so μ is referred to as the *quasilimiting* distribution, and it is also *quasistationary*, in the sense that $\mathbb{P}_\mu(X_T \in A\,|\tau_\partial > T) = \mu(A)$, for all $T \geq 0$, measurable $A \subset \mathbb{R}^d$.

Question: given a more general diffusion X and a target density π on $\mathbb{R}^d$, can we *choose* κ such that π is the quasistationary distribution? And if so, how fast is this convergence?

Convergence to Quasistationarity

Assumptions: Suppose we are interested in killing the diffusion X defined by the SDE (1). We make standard technical assumptions so the SDE (1) has a unique non-explosive weak solution. We assume that the target density $\pi : \mathbb{R}^d \to [0, \infty)$ is a positive, smooth, integrable function on $\mathbb{R}^d$. Assume that $\tilde{\kappa}$ is bounded below, and that a compatibility integral condition holds:

$$\tilde{\kappa}(x) := \frac{1}{2}\left(\frac{\Delta\pi}{\pi} - \frac{2\nabla A \cdot \nabla\pi}{\pi} - 2\Delta A\right)(x), \quad x \in \mathbb{R}^d, \qquad \int_{\mathbb{R}^d} \frac{\pi^2(x)}{\exp(2A(x))}\,dx < \infty.$$

Then set $\kappa := \tilde{\kappa} + K$, constant K chosen so that $\kappa(x) \geq 0$ for all $x \in \mathbb{R}^d$.

Theorem [3]. Under the above assumptions we have:

1. X killed at rate κ has quasilimiting distribution π.
2. The conditioned laws (2) converge to quasistationarity π at the same rate (in $\mathcal{L}^2$) at which the Langevin diffusion targeting $\pi^2/\exp(2A)$ converges to stationarity.

Simulation via Stochastic Approximation

Suppose X has quasilimiting distribution π. How might we try to draw from π? Rejection sampling. ScaLE: continuous-time sequential Monte Carlo [2]. Regenerating ScaLE (ReScaLE): stochastic approximation, *c.f.* [1] for compact state space, discrete-time case. See Figure 2.

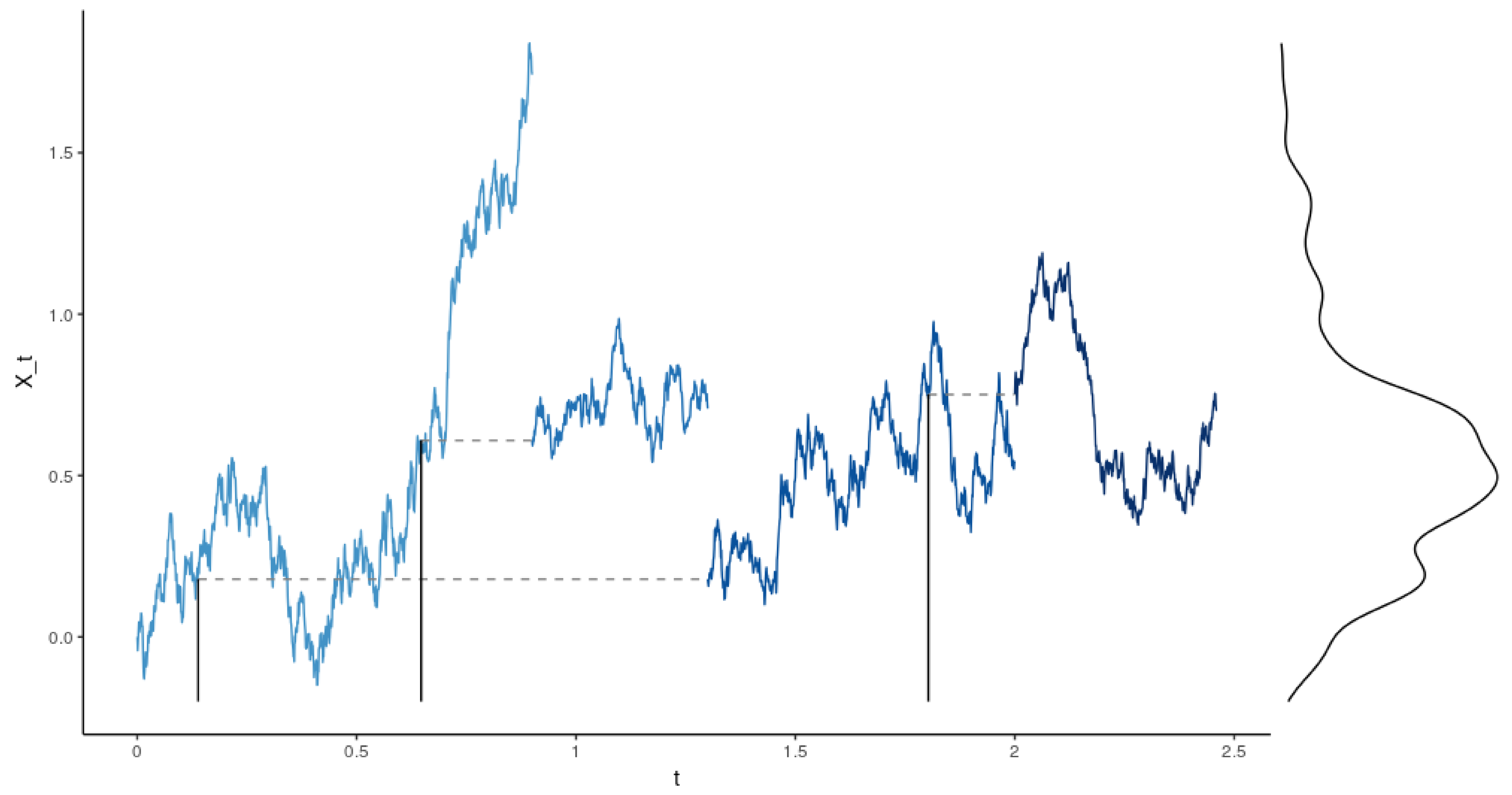


Figure 2: A trajectory of ReScaLE until $T = 2.5$.

Initialise $X_0 \sim \mu_0$ and fix $r > 0$. Our diffusion X is killed at rate $t \mapsto \kappa(X_t)$. At each death event T_n, the process is instantaneously reborn according to $X_{T_n} \sim \mu_{T_n}$ independently, where

$$\mu_t = \frac{r\mu_0 + \int_0^t \delta_{X_{s-}}\,ds}{r + t}, \qquad \text{for each } t \geq 0.$$

Conjecture: $\mu_{T_n} \to \pi$ (weakly) almost surely under fairly general settings. We have established this in the restricted case of compact state space with bounded killing. We follow the approach of [1].

Make a time-change: $h(t) = \exp(t) - r, \quad \zeta_t := \mu_{h(t)}$. Then it turns out that ζ_t satisfies a (weak) ODE:

$$\dot{\zeta}_t = \big(-\zeta_t + \Pi(\zeta_t)\big) + \big(\delta_{X_{h(t)-}} - \Pi(\zeta_t)\big).$$

[That is, for any continuous, bounded test function f, $t \mapsto \zeta_t(f)$ satisfies the appropriate ODE on $\mathbb{R}$.]
Here $\Pi(\mu)$ denotes the invariant distribution of the fixed rebirth process X^μ with rebirth distribution μ: that is, the diffusion killed at rate $t \mapsto \kappa(X^\mu_t)$ and instantaneously reborn according to μ.
We argue that the second bracket is "small", and analyse the resulting deterministic ODE.

Proposition. We can define a continuous flow $\mathbb{R}_+ \times \mathcal{P}(\mathbb{R}^d) \to P(\mathbb{R}^d)$, $(t, \mu) \mapsto \Phi_t(\mu)$ such that $t \mapsto \Phi_t(\mu)$ solves the weak, measure-valued ODE $\dot{\nu}_t = -\nu_t + \Pi(\nu_t), \nu_0 = \mu$. Furthermore we have that $\Phi_t(\mu) \to \pi$ as $t \to \infty$ for any compactly support μ.

We say that $t \mapsto \zeta_t$ is an *asymptotic pseudo-trajectory for* Φ if (ζ_t) is tight and for all $T > 0$,

$$\lim_{t\to\infty} \sup_{s\in[0,T]} d(\zeta_{t+s}, \Phi_s(\zeta_t)) = 0.$$

[d is a metric metrising weak convergence of probability measures.]
This holds if (ζ_t) is tight and the signed measures $\epsilon_t(s) := \int_t^{t+s}(\delta_{X(h(u)-)} - \Pi(\zeta_u))\,du$ satisfies for each continuous bounded f and $T > 0$,

$$\lim_{t\to\infty} \sup_{s\in[0,T]} |\epsilon_t(s)f| = 0.$$

Establishing tightness involves showing $\int_0^t \kappa(X_s)\,ds = O(t)$ almost surely; have urn-like dynamics. Controlling $\epsilon_t(s)$ involves showing that $\int_0^t \mathbb{E}\kappa(X_s)\,ds = O(t)$. Can utilise renewal-like structure.

References

[1] Michel Benaïm, Bertrand Cloez, and Fabien Panloup. Stochastic Approximation of Quasi-stationary Distributions on Compact State Spaces and Applications. *arXiv preprint*, 2016.

[2] Murray Pollock, Paul Fearnhead, Adam M. Johansen, and Gareth O. Roberts. The Scalable Langevin Exact Algorithm: Bayesian Inference for Big Data. *arXiv preprint*, 2016.

[3] Andi Q. Wang, Martin Kolb, Gareth O. Roberts, and David Steinsaltz. Theoretical Properties of Quasistationary Monte Carlo Methods. *arXiv preprint*, 2017.

I am grateful to EPSRC for the financial support and my supervisors for their valuable insights.