Theoretical Properties of Quasistationary Monte Carlo Methods

Andi Wang

University of Oxford

a.wang@stats.ox.ac.uk

Joint with Divakar Kumar, Gareth Roberts and David Steinsaltz

11 July 2018

Oxford-Warwick Statistics Programme

Introduction to Quasistationarity: Ants

Let $X = (X_t)$ be an ant undergoing a diffusion on $\mathbb{R}^d$. Introduce killing rate

$$\kappa : \mathbb{R}^d \to [0, \infty).$$

At rate $\kappa(X_t)$ the ant is *killed*; call this time τ_∂.

Introduction to Quasistationarity: Ants

Let $X = (X_t)$ be an ant undergoing a diffusion on $\mathbb{R}^d$. Introduce killing rate

$$\kappa : \mathbb{R}^d \to [0, \infty).$$

At rate $\kappa(X_t)$ the ant is *killed*; call this time τ_∂.

We will consider $\mathbb{P}_x(X_t \in \cdot \,|\tau_\partial > t)$ for large t.

Introduction to Quasistationarity: Ants

Let $X = (X_t)$ be an ant undergoing a diffusion on $\mathbb{R}^d$. Introduce killing rate

$$\kappa : \mathbb{R}^d \to [0, \infty).$$

At rate $\kappa(X_t)$ the ant is *killed*; call this time τ_∂.

We will consider $\mathbb{P}_x(X_t \in \cdot \,|\tau_\partial > t)$ for large t.

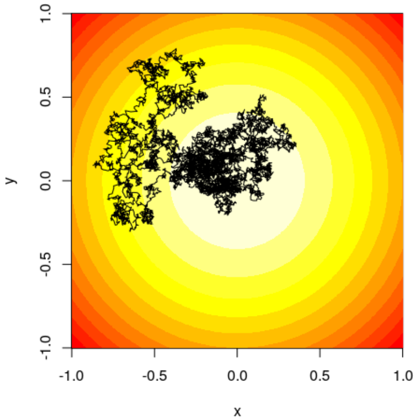
If these converge to π as $t \to \infty$, π is an example of a *quasistationary distribution*.

Example

Take X to be a standard Brownian motion on $\mathbb{R}^2$, $\kappa(y) = \|y\|^2$.

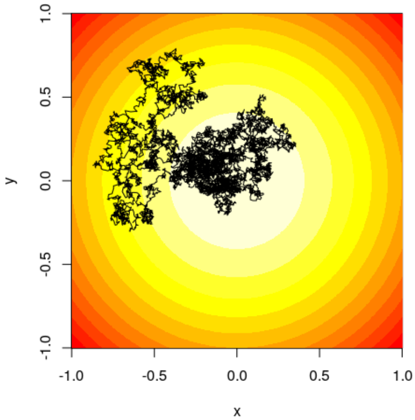
Example

Take X to be a standard Brownian motion on $\mathbb{R}^2$, $\kappa(y) = \|y\|^2$.



Example

Take X to be a standard Brownian motion on $\mathbb{R}^2$, $\kappa(y) = \|y\|^2$.



What can be said about $\mathbb{P}(X_t \in \cdot \,|\tau_\partial > t)$ for large t?

Quasistationary Monte Carlo

Quasistationary Monte Carlo methods aim to sample from a target distribution π, where π is a quasistationary distribution.

[1] Pollock, M., Fearnhead, P., Johansen, A. M., Roberts, G. O. (2016). The Scalable Langevin Exact Algorithm: Bayesian Inference for Big Data. arXiv Preprint: arXiv 1609.03436.

Quasistationary Monte Carlo

Quasistationary Monte Carlo methods aim to sample from a target distribution π, where π is a quasistationary distribution.

The quasistationary framework enables the principled use of *subsampling* techniques to give exact Bayesian inference with a sub-linear cost in the number of observations[1].

[1] Pollock, M., Fearnhead, P., Johansen, A. M., Roberts, G. O. (2016). The Scalable Langevin Exact Algorithm: Bayesian Inference for Big Data. arXiv Preprint: arXiv 1609.03436.

Foundational Results

$$\mathrm{d}X_t = \nabla A(X_t)\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = x \in \mathbb{R}^d.$$

Theorem (Convergence to Quasistationarity)

Under certain assumptions, the diffusion X killed at rate κ has quasilimiting distribution π. That is, for each measurable $E \subset \mathbb{R}^d$ we have as $t \to \infty$,

$$\mathbb{P}_x(X_t \in E | \tau_\partial > t) \to \pi(E).$$

Foundational Results

$$\mathrm{d}X_t = \nabla A(X_t)\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = x \in \mathbb{R}^d.$$

Theorem (Convergence to Quasistationarity)

Under certain assumptions, the diffusion X killed at rate κ has quasilimiting distribution π. That is, for each measurable $E \subset \mathbb{R}^d$ we have as $t \to \infty$,

$$\mathbb{P}_x(X_t \in E | \tau_\partial > t) \to \pi(E).$$

Theorem (Rates of convergence)

Additionally, X converges to quasistationarity π at the same rate as the Langevin diffusion targeting $\pi^2/2A$ converges to stationarity.

Simulating from QSDs

Suppose we have a killed diffusion X with quasilimiting distribution π. So $\mathbb{P}(X_t \in \cdot | \tau_\partial > t) \to \pi$. How might we simulate try to simulate π?

Simulating from QSDs

Suppose we have a killed diffusion X with quasilimiting distribution π. So $\mathbb{P}(X_t \in \cdot | \tau_\partial > t) \to \pi$. How might we simulate try to simulate π?

1. Rejection sampling. Infeasible.

Simulating from QSDs

Suppose we have a killed diffusion X with quasilimiting distribution π. So $\mathbb{P}(X_t \in \cdot | \tau_\partial > t) \to \pi$. How might we simulate try to simulate π?

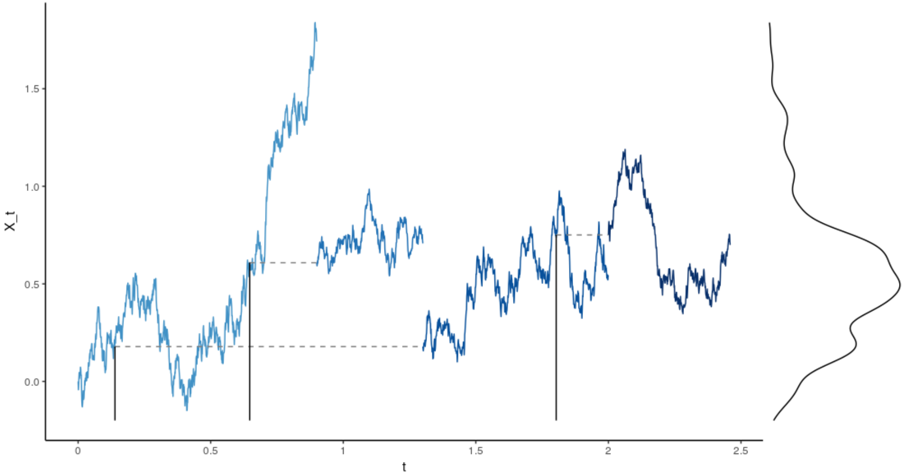
1. Rejection sampling. Infeasible.
2. Continuous-time sequential Monte Carlo. Feasible but involved.

Simulating from QSDs

Suppose we have a killed diffusion X with quasilimiting distribution π. So $\mathbb{P}(X_t \in \cdot | \tau_\partial > t) \to \pi$. How might we simulate try to simulate π?

1. Rejection sampling. Infeasible.
2. Continuous-time sequential Monte Carlo. Feasible but involved.
3. ReScaLE: a stochastic approximation approach.

An Example ReScaLE Trajectory



Convergence Result

Does this algorithm converge to the quasistationary distribution π?

Convergence Result

Does this algorithm converge to the quasistationary distribution π?

Theorem (Convergence in compact setting)

When the state space is compact, we have that (after time-changing) (μ_t) is an *asymptotic pseudo-trajectory* for a deterministic semiflow Φ almost surely.

It follows that μ_t converges to π almost surely.

Convergence Result

Does this algorithm converge to the quasistationary distribution π?

Theorem (Convergence in compact setting)

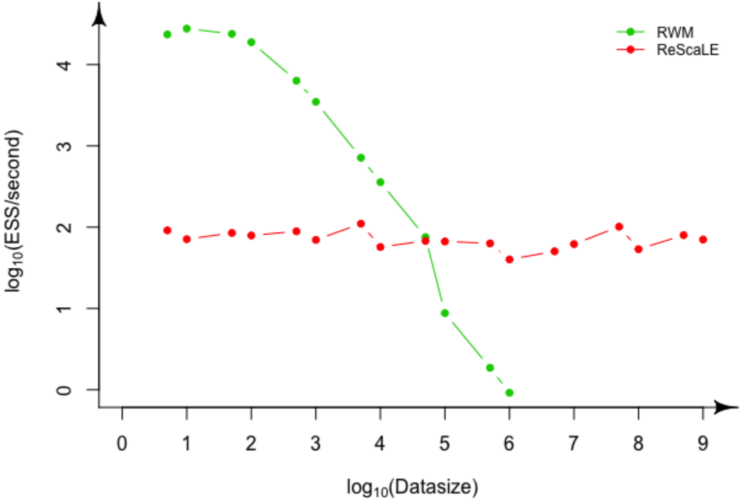
When the state space is compact, we have that (after time-changing) (μ_t) is an *asymptotic pseudo-trajectory* for a deterministic semiflow Φ almost surely.

It follows that μ_t converges to π almost surely.

Conjecture (General setting)

We should have that the Proposition holds much more generally: non-compact state space, unbounded killing rate.

Logistic regression example (courtesy of D. Kumar)



Thanks for listening!

If you are interested to learn more, come see my poster!

Wang, A.Q., Kolb, M., Roberts, G.O. and Steinsaltz, D. (2017) Theoretical Properties of Quasistationary Monte Carlo Methods. *arXiv* 1707.08036. In revision, *Annals of Applied Probability*.

Wang, A.Q., Roberts, G.O. and Steinsaltz, D. Stochastic Approximation of Quasistationary Distributions of Killed Diffusions on Compact Spaces. *In preparation.*