Foundational Properties of Quasistationary Monte Carlo Methods

Andi Wang

University of Oxford

a.wang@stats.ox.ac.uk

Joint work with Martin Kolb, Gareth Roberts and David Steinsaltz

August 3rd 2017

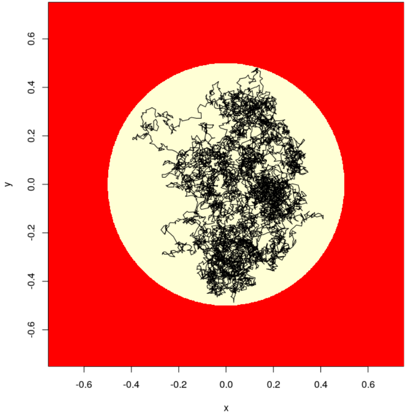
Overview

Quasistationarity: boundary killing

Ant on volcanic island undergoing Brownian motion, killed at τ_∂ when it touches lava.

Quasistationarity: boundary killing

Ant on volcanic island undergoing Brownian motion, killed at τ_∂ when it touches lava.



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

Quasistationarity: interior killing

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

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

Define killing time via

$$\tau_\partial := \inf \left\{ t \geq 0 : \int_0^t \kappa(X_s)\,\mathrm{d}s > \xi \right\}$$

where $\xi \sim \mathrm{Exp}(1)$, independent of X.

Quasistationarity: interior killing

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

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

Define killing time via

$$\tau_\partial := \inf \left\{ t \geq 0 : \int_0^t \kappa(X_s)\, \mathrm{d}s > \xi \right\}$$

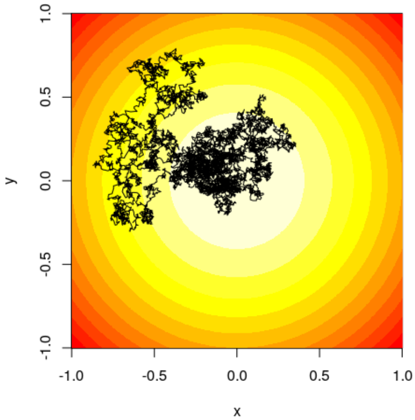
where $\xi \sim \mathrm{Exp}(1)$, independent of X. We will consider $\mathbb{P}_x(X_t \in \cdot \,|\tau_\partial > t)$.

Quasistationarity: interior killing example

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

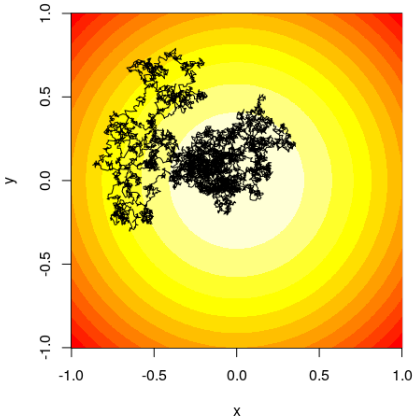
Quasistationarity: interior killing example

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



Quasistationarity: interior killing 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?

Important concepts

Definitions

A density π on $\mathbb{R}^d$ is called a *quasilimiting distribution* if

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

for each Borel-measurable $E \subset \mathbb{R}^d$ as $t \to \infty$, for any starting point $x \in \mathbb{R}^d$.

π is called *quasistationary* if for any $t \geq 0$

$$\mathbb{P}_\pi(X_t \in \cdot \,|\tau_\partial > t) = \pi(\cdot).$$

Important concepts

Definitions

A density π on $\mathbb{R}^d$ is called a *quasilimiting distribution* if

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

for each Borel-measurable $E \subset \mathbb{R}^d$ as $t \to \infty$, for any starting point $x \in \mathbb{R}^d$.

π is called *quasistationary* if for any $t \geq 0$

$$\mathbb{P}_\pi(X_t \in \cdot \,|\tau_\partial > t) = \pi(\cdot).$$

- Laws $\{\mathbb{P}_x(X_t \in \cdot \,|\tau_\partial > t)\}_{t \geq 0}$ are not consistent.

Important concepts

Definitions

A density π on $\mathbb{R}^d$ is called a *quasilimiting distribution* if

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

for each Borel-measurable $E \subset \mathbb{R}^d$ as $t \to \infty$, for any starting point $x \in \mathbb{R}^d$.

π is called *quasistationary* if for any $t \geq 0$

$$\mathbb{P}_\pi(X_t \in \cdot \,|\tau_\partial > t) = \pi(\cdot).$$

- Laws $\{\mathbb{P}_x(X_t \in \cdot \,|\tau_\partial > t)\}_{t \geq 0}$ are not consistent.
- Even under irreducibility there may be many quasistationary functions.

Quasistationary Monte Carlo

Given a diffusion X on $\mathbb{R}^d$, defined through

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

where $A : \mathbb{R}^d \to \mathbb{R}$ is smooth, and a (smooth, positive) target density π on $\mathbb{R}^d$,

Quasistationary Monte Carlo

Given a diffusion X on $\mathbb{R}^d$, defined through

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

where $A : \mathbb{R}^d \to \mathbb{R}$ is smooth, and a (smooth, positive) target density π on $\mathbb{R}^d$, can we define a killing rate $\kappa : \mathbb{R}^d \to [0, \infty)$ so that π is the quasilimiting distribution?

Quasistationary Monte Carlo

Given a diffusion X on $\mathbb{R}^d$, defined through

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

where $A : \mathbb{R}^d \to \mathbb{R}$ is smooth, and a (smooth, positive) target density π on $\mathbb{R}^d$, can we define a killing rate $\kappa : \mathbb{R}^d \to [0, \infty)$ so that π is the quasilimiting distribution?
E.g.

$$\pi(x) \propto \prod_{i=1}^{N} f_i(x)$$

is a Bayesian posterior distribution.

Quasistationary Monte Carlo

Given a diffusion X on $\mathbb{R}^d$, defined through

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

where $A : \mathbb{R}^d \to \mathbb{R}$ is smooth, and a (smooth, positive) target density π on $\mathbb{R}^d$, can we define a killing rate $\kappa : \mathbb{R}^d \to [0, \infty)$ so that π is the quasilimiting distribution?
E.g.

$$\pi(x) \propto \prod_{i=1}^{N} f_i(x)$$

is a Bayesian posterior distribution. Why? Scalable Langevin Exact Algorithm: scales well with N.

See Pollock et al. (2016). The Scalable Langevin Exact Algorithm: Bayesian Inference for Big Data. *arXiv 1609.03436*.

Main Results

- We have proven natural sufficient conditions under which the quasilimiting distribution of X is π.
- We have quantified the rate of convergence to quasistationarity by relating the killed diffusion to an appropriate Langevin diffusion.

When does convergence to QS occur? I

Define $\tilde{\kappa} : \mathbb{R}^d \to \mathbb{R}$ by

$$\tilde{\kappa}(y) := \frac{1}{2}\left(\frac{\Delta\pi}{\pi} - \frac{2\nabla A \cdot \nabla\pi}{\pi} - 2\Delta A\right)(y), \quad y \in \mathbb{R}^d.$$

When does convergence to QS occur? I

Define $\tilde{\kappa} : \mathbb{R}^d \to \mathbb{R}$ by

$$\tilde{\kappa}(y) := \frac{1}{2}\left(\frac{\Delta\pi}{\pi} - \frac{2\nabla A \cdot \nabla \pi}{\pi} - 2\Delta A\right)(y), \quad y \in \mathbb{R}^d.$$

Assumption 1

$\tilde{\kappa}$ is bounded below and not identically zero.

For $K := -\inf_{y\in\mathbb{R}^d} \tilde{\kappa}(y)$, the correct killing rate will be

$$\kappa := \tilde{\kappa} + K.$$

When does convergence to QS occur? II

$$\mathrm{d}X_t = \nabla A(X_t)\,\mathrm{d}t + \mathrm{d}W_t$$

If $\exp(2A)$ is integrable (and under some regularity conditions), it is (proportional to) the invariant density of the unkilled diffusion X.

When does convergence to QS occur? II

$$\mathrm{d}X_t = \nabla A(X_t)\,\mathrm{d}t + \mathrm{d}W_t$$

If $\exp(2A)$ is integrable (and under some regularity conditions), it is (proportional to) the invariant density of the unkilled diffusion X.

Equivalently, writing $U = \log \pi$

$$\tilde{\kappa}(y) = \frac{1}{2}\big(\Delta(U - 2A) + \nabla U \cdot (\nabla U - 2\nabla A)\big).$$

Then Assumption 1 says this discrepancy can't be arbitrarily negative.

When does convergence to QS occur? III

Assumption 2

$$\int_{\mathbb{R}^d} \frac{\pi^2(y)}{\exp(2A(y))}\, \mathrm{d}y < \infty.$$

When does convergence to QS occur? III

Assumption 2

$$\int_{\mathbb{R}^d} \frac{\pi^2(y)}{\exp(2A(y))}\,\mathrm{d}y < \infty.$$

Two ways to interpret this assumption:

- Importance sampling.

When does convergence to QS occur? III

Assumption 2

$$\int_{\mathbb{R}^d} \frac{\pi^2(y)}{\exp(2A(y))} \, \mathrm{d}y < \infty.$$

Two ways to interpret this assumption:

- Importance sampling. *Cf.* rejection sampling condition: there exists some M such that

$$\frac{\pi(y)}{\exp(2A(y))} < M \quad \forall y \in \mathbb{R}^d.$$

  Need $\exp(2A)$ to have heavier tails than π.

When does convergence to QS occur? III

Assumption 2

$$\int_{\mathbb{R}^d} \frac{\pi^2(y)}{\exp(2A(y))}\, \mathrm{d}y < \infty.$$

Two ways to interpret this assumption:

- Importance sampling. *Cf.* rejection sampling condition: there exists some M such that

$$\frac{\pi(y)}{\exp(2A(y))} < M \quad \forall y \in \mathbb{R}^d.$$

  Need $\exp(2A)$ to have heavier tails than π.

- Spectral theory: define $\varphi := \pi / \exp(2A)$, $\mathcal{L}^2(\Gamma)$ given by $\Gamma(\mathrm{d}y) = \exp(2A(y))\, \mathrm{d}y$. Then we require $\varphi \in \mathcal{L}^2(\Gamma)$.

When does convergence to QS occur? IV

Theorem (Convergence to Quasistationarity)

Under Assumptions 1 and 2 (and basic regularity conditions), the diffusion X killed at rate κ has quasilimiting distribution π. That is, for each measurable $E \subset \mathbb{R}^d$ we have

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

as $t \to \infty$.

Key ingredients in proof

Let $-L^\kappa$ denote the infinitesimal generator of the killed diffusion. L^κ can be realised as a positive, self-adjoint (unbounded) operator on $\mathcal{L}^2(\Gamma)$ (Assumption 1 required).

[1] Tuominen and Tweedie (1979). Exponential Decay and Ergodicity of General Markov Processes and Their Discrete Skeletons. *Advances in Applied Probability* **11** 784-803.

Key ingredients in proof

Let $-L^\kappa$ denote the infinitesimal generator of the killed diffusion. L^κ can be realised as a positive, self-adjoint (unbounded) operator on $\mathcal{L}^2(\Gamma)$ (Assumption 1 required). We have

$$L^\kappa \varphi = K\varphi$$

and in fact $K = \lambda_0^\kappa$, the bottom of the spectrum of L^κ (Assumption 2).

[1] Tuominen and Tweedie (1979). Exponential Decay and Ergodicity of General Markov Processes and Their Discrete Skeletons. *Advances in Applied Probability* **11** 784-803.

Key ingredients in proof

Let $-L^\kappa$ denote the infinitesimal generator of the killed diffusion. L^κ can be realised as a positive, self-adjoint (unbounded) operator on $\mathcal{L}^2(\Gamma)$ (Assumption 1 required). We have

$$L^\kappa \varphi = K\varphi$$

and in fact $K = \lambda_0^\kappa$, the bottom of the spectrum of L^κ (Assumption 2).

Tweedie's R-theory[1]: Assumption 2 enables us to show our killed diffusion is λ-positive. This gives us the desired convergence.

[1] Tuominen and Tweedie (1979). Exponential Decay and Ergodicity of General Markov Processes and Their Discrete Skeletons. *Advances in Applied Probability* **11** 784-803.

Rates of Convergence to QS I

How fast does $\mathbb{P}_x(X_t \in \cdot \,|\tau_\partial > t)$ converge to π?

Rates of Convergence to QS I

How fast does $\mathbb{P}_x(X_t \in \cdot \,|\tau_\partial > t)$ converge to π?

Consider the Langevin diffusion Z described by

$$\mathrm{d}Z_t = \frac{1}{2}\nabla \log\left(\frac{\pi^2}{\exp(2A)}\right)(Z_t)\,\mathrm{d}t + \mathrm{d}W_t$$

and let $-L^Z$ be the generator of this process, realised as a positive, self-adjoint operator on the appropriate $\mathcal{L}^2$ space.

Rates of Convergence to QS I

How fast does $\mathbb{P}_x(X_t \in \cdot \,|\tau_\partial > t)$ converge to π?

Consider the Langevin diffusion Z described by

$$\mathrm{d}Z_t = \frac{1}{2}\nabla \log\left(\frac{\pi^2}{\exp(2A)}\right)(Z_t)\,\mathrm{d}t + \mathrm{d}W_t$$

and let $-L^Z$ be the generator of this process, realised as a positive, self-adjoint operator on the appropriate $\mathcal{L}^2$ space.

Theorem

Up to an additive constant, the spectra of L^Z and L^κ coincide.

Rates of Convergence to QS II

So heuristically $\mathbb{P}_x(X_t \in \cdot | \tau_\partial > t)$ converges to π at the same rate as Z converging to $\pi^2/\exp(2A)$.

Rates of Convergence to QS II

So heuristically $\mathbb{P}_x(X_t \in \cdot \,|\tau_\partial > t)$ converges to π at the same rate as Z converging to $\pi^2/\exp(2A)$.

A sufficient condition for a spectral gap is that

$$\lim_{\|x\|\to\infty} \tilde{\kappa}(x) > 0.$$

Rates of Convergence to QS II

So heuristically $\mathbb{P}_x(X_t \in \cdot | \tau_\partial > t)$ converges to π at the same rate as Z converging to $\pi^2/\exp(2A)$.

A sufficient condition for a spectral gap is that

$$\lim_{\|x\|\to\infty} \tilde{\kappa}(x) > 0.$$

It turns out that the Langevin diffusion Z is precisely the Q-process, defined through

$$\mathbb{Q}_x(A) := \lim_{T\to\infty} \mathbb{P}_x(A | T < \tau_\partial)$$

for $A \in \sigma(X_s : s \leq t)$ for some $t \geq 0$. This is the law of the process conditioned never to be killed.

Killed OU process targeting Gaussian

We will kill the diffusion

$$\mathrm{d}X_t = \frac{1}{2\tau^2}(\nu - X_t)\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = x.$$

This has an $\mathcal{N}(\nu, \tau^2)$ invariant distribution.
Our target density is

$$\pi(y) \propto \exp\left\{ -\frac{1}{2\sigma^2}(y-\mu)^2 \right\}.$$

Calculations give

$$\tilde{\kappa}(y) = \frac{1}{2}\left(\frac{(y-\mu)^2}{\sigma^4} - \frac{1}{\sigma^2} + \frac{(\nu - y)(y-\mu)}{\tau^2\sigma^2} + \frac{1}{\tau^2} \right).$$

This is bounded below when

$$\tau^2 > \sigma^2.$$

OU Example continued

Assumption 2 also holds when $\tau^2 > \sigma^2$; we so have convergence to the quasistationary distribution π.

[2] Metafune et al (2002). Spectrum of Ornstein–Uhlenbeck Operators in Lp Spaces with Respect to Invariant Measures. *Journal of Functional Analysis* **196** 40-60.

OU Example continued

Assumption 2 also holds when $\tau^2 > \sigma^2$; we so have convergence to the quasistationary distribution π.

Can be shown[2] that

$$\Sigma(L^Z) = \left\{ \lambda_n^Z = \frac{n(2\tau^2 - \sigma^2)}{2\sigma^2\tau^2} : n = 0, 1, 2, \ldots \right\}.$$

So by our Theorem the spectral gap of our killed process is

$$\lambda_1^Z - \lambda_0^Z = \frac{2\tau^2 - \sigma^2}{2\sigma^2\tau^2} = \frac{1}{\sigma^2} - \frac{1}{2\tau^2}.$$

[2] Metafune et al (2002). Spectrum of Ornstein–Uhlenbeck Operators in Lp Spaces with Respect to Invariant Measures. *Journal of Functional Analysis* **196** 40-60.

Conclusion

Quasistationary Monte Carlo methods such as the ScaLE algorithm are an interesting new development, both from a mathematical and applied perspective.

We have proven some fundamental results in this area, bringing together some tools from applied probability and abstract operator theory.

Thanks for listening!

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

Figure: "*Happy volcano scares ant*"[3]

[3] https://drawception.com/panel/drawing/iMOG6336/happy-volcano-scares-ant/