Boost your favorite Markov Chain Monte Carlo using Kac formula: the Kick-Kac Teleportation algorithm.

R. Douc, A. Durmus, A. Enfroy and J. Olsson

Télécom Sudparis, Institut Polytechnique de Paris
randal.douc@telecom-sudparis.eu

Institut Mines-Télécom

Outline

Outline

The Kac formula for Markov chains

Definitions and Notation

- First return time to the set C: $\sigma_{\mathsf{C}} = \inf\{k \geqslant 1 : X_k \in \mathsf{C}\}$.
- C is π-accessible iff $\mathbb{P}_x(\sigma_{\mathsf{C}} < \infty) > 0$ for π-almost all $x \in \mathsf{X}$.

Theorem (The Kac Formula)

Let P be a Markov kernel on $\mathsf{E} \times \mathcal{E}$ with invariant probability measure π. Then, for all π-accessible sets $\mathsf{C} \in \mathcal{E}$, we have $\pi = \pi^0_{\mathsf{C}} = \pi^1_{\mathsf{C}}$, *where*

$$\pi^0_{\mathsf{C}}(f) = \int_{\mathsf{C}} \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_{\mathsf{C}}-1} f(X_k)\right], \quad (1)$$

$$\pi^1_{\mathsf{C}}(f) = \int_{\mathsf{C}} \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_{\mathsf{C}}} f(X_k)\right]. \quad (2)$$

The Kac formula for Markov chains

Definitions and Notation

- First return time to the set C: $\sigma_{\mathsf{C}} = \inf\{k \geqslant 1 \,:\, X_k \in \mathsf{C}\}$.
- C is π-accessible iff $\mathbb{P}_x(\sigma_{\mathsf{C}} < \infty) > 0$ for π-almost all $x \in \mathsf{X}$.

Theorem (The Kac Formula)

Let P be a Markov kernel on $\mathsf{E} \times \mathcal{E}$ with invariant probability measure π. Then, for all π-accessible sets $\mathsf{C} \in \mathcal{E}$, we have $\boxed{\pi = \pi^0_{\mathsf{C}} = \pi^1_{\mathsf{C}}}$, *where*

$$\pi^0_{\mathsf{C}}(f) = \int_{\mathsf{C}} \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_{\mathsf{C}}-1} f(X_k)\right], \tag{1}$$

$$\pi^1_{\mathsf{C}}(f) = \int_{\mathsf{C}} \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_{\mathsf{C}}} f(X_k)\right]. \tag{2}$$

Proof of the Kac Theorem I

By the last-exit decomposition and the Markov property, for all bounded functions $f \geqslant 0$ and all $n \geqslant 1$,

$$\pi(f) = \mathbb{E}_\pi[f(X_n)] = \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_\mathsf{C} \leqslant n\}] + \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_\mathsf{C} > n\}]$$

$$= \sum_{\ell=1}^{n} \underbrace{\mathbb{E}_\pi\left[f(X_n)\mathbf{1}_\mathsf{C}(X_\ell)\prod_{k=\ell+1}^{n}\mathbf{1}_{\mathsf{C}^c}(X_k)\right]}_{\mathbb{E}_\pi\left[\mathbf{1}_\mathsf{C}(X_\ell)\mathbb{E}_{X_\ell}\left[f(X_{n-\ell})\prod_{k=1}^{n-\ell}\mathbf{1}_{\mathsf{C}^c}(X_k)\right]\right]} + \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_\mathsf{C} > n\}]$$

Noting that π is invariant and setting $k = n - \ell$, we finally get

$$\pi(f) = \sum_{k=0}^{n-1}\int_\mathsf{C}\pi(\mathrm{d}x)\mathbb{E}_x[f(X_k)\mathbf{1}\{\sigma_\mathsf{C} > k\}] + \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_\mathsf{C} > n\}]$$

$$= \int_\mathsf{C}\pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{(n-1)\wedge(\sigma_\mathsf{C}-1)} f(X_k)\right] + \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_\mathsf{C} > n\}]\,. \tag{3}$$

Proof of the Kac Theorem I

By the last-exit decomposition and the Markov property, for all bounded functions $f \geqslant 0$ and all $n \geqslant 1$,

$$\pi(f) = \mathbb{E}_\pi[f(X_n)] = \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_{\mathsf{C}} \leqslant n\}] + \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_{\mathsf{C}} > n\}]$$

$$= \sum_{\ell=1}^{n} \underbrace{\mathbb{E}_\pi\left[f(X_n)\mathbf{1}_{\mathsf{C}}(X_\ell)\prod_{k=\ell+1}^{n}\mathbf{1}_{\mathsf{C}^c}(X_k)\right]}_{\mathbb{E}_\pi\left[\mathbf{1}_{\mathsf{C}}(X_\ell)\mathbb{E}_{X_\ell}\left[f(X_{n-\ell})\prod_{k=1}^{n-\ell}\mathbf{1}_{\mathsf{C}^c}(X_k)\right]\right]} + \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_{\mathsf{C}} > n\}]$$

Noting that π is invariant and setting $k = n - \ell$, we finally get

$$\pi(f) = \sum_{k=0}^{n-1}\int_{\mathsf{C}}\pi(\mathrm{d}x)\mathbb{E}_x[f(X_k)\mathbf{1}\{\sigma_{\mathsf{C}} > k\}] + \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_{\mathsf{C}} > n\}]$$

$$= \int_{\mathsf{C}}\pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{(n-1)\wedge(\sigma_{\mathsf{C}}-1)} f(X_k)\right] + \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_{\mathsf{C}} > n\}]\,. \tag{3}$$

Proof of the Kac Theorem I

By the last-exit decomposition and the Markov property, for all bounded functions $f \geqslant 0$ and all $n \geqslant 1$,

$$\pi(f) = \mathbb{E}_\pi[f(X_n)] = \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_\mathsf{C} \leqslant n\}] + \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_\mathsf{C} > n\}]$$

$$= \sum_{\ell=1}^{n} \underbrace{\mathbb{E}_\pi\left[f(X_n)\mathbf{1}_\mathsf{C}(X_\ell)\prod_{k=\ell+1}^{n}\mathbf{1}_{\mathsf{C}^c}(X_k)\right]}_{\mathbb{E}_\pi\left[\mathbf{1}_\mathsf{C}(X_\ell)\mathbb{E}_{X_\ell}\left[f(X_{n-\ell})\prod_{k=1}^{n-\ell}\mathbf{1}_{\mathsf{C}^c}(X_k)\right]\right]} + \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_\mathsf{C} > n\}]$$

Noting that π is invariant and setting $k = n - \ell$, we finally get

$$\pi(f) = \sum_{k=0}^{n-1}\int_\mathsf{C} \pi(\mathrm{d}x)\mathbb{E}_x[f(X_k)\mathbf{1}\{\sigma_\mathsf{C} > k\}] + \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_\mathsf{C} > n\}]$$

$$= \int_\mathsf{C} \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{(n-1)\wedge(\sigma_\mathsf{C}-1)} f(X_k)\right] + \mathbb{E}_\pi[f(X_n)\mathbf{1}\{\sigma_\mathsf{C} > n\}] \,. \tag{3}$$

Checking the π-accessibility.

Lemma

Let P be a Markov kernel on $\mathsf{E} \times \mathcal{E}$ with a unique invariant probability measure π. Then, any set $\mathsf{C} \in \mathcal{E}$ such that $\pi(\mathsf{C}) > 0$ is π-accessible.

P has a unique invariant probability measure

$\Rightarrow$ the associated dynamical system $(\mathsf{E}^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}}, \theta, \mathbb{P}_\pi)$ is ergodic

$\Rightarrow$ the Birkhoff theorem applies.

Then, for all $\mathsf{C} \in \mathcal{E}$ such that $\pi(\mathsf{C}) > 0$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \mathbf{1}_{\mathsf{C}}(X_k) = \pi(\mathsf{C}) > 0 \quad \mathbb{P}_\pi - \text{a.s.}$$

Thus, $\mathbb{P}_\pi(\sigma_{\mathsf{C}} < \infty) = 1$. And therefore $\mathbb{P}_x(\sigma_{\mathsf{C}} < \infty) = 1 > 0$ for π-almost all $x \in \mathsf{X}$.

Checking the π-accessibility.

Lemma

Let P be a Markov kernel on $\mathsf{E} \times \mathcal{E}$ with a unique invariant probability measure π. Then, any set $\mathsf{C} \in \mathcal{E}$ such that $\pi(\mathsf{C}) > 0$ is π-accessible.

P has a unique invariant probability measure

$\Rightarrow$ the associated dynamical system $(\mathsf{E}^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}}, \theta, \mathbb{P}_\pi)$ is ergodic

$\Rightarrow$ the Birkhoff theorem applies.

Then, for all $\mathsf{C} \in \mathcal{E}$ such that $\pi(\mathsf{C}) > 0$,

$$\lim_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} \mathbf{1}_{\mathsf{C}}(X_k) = \pi(\mathsf{C}) > 0 \quad \mathbb{P}_\pi - \text{a.s.}$$

Thus, $\mathbb{P}_\pi(\sigma_{\mathsf{C}} < \infty) = 1$. And therefore $\mathbb{P}_x(\sigma_{\mathsf{C}} < \infty) = 1 > 0$ for π-almost all $x \in \mathsf{X}$.

Checking the π-accessibility.

Lemma

Let P be a Markov kernel on $\mathsf{E} \times \mathcal{E}$ with a unique invariant probability measure π. Then, any set $\mathsf{C} \in \mathcal{E}$ such that $\pi(\mathsf{C}) > 0$ is π-accessible.

P has a unique invariant probability measure

$\Rightarrow$ the associated dynamical system $(\mathsf{E}^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}}, \theta, \mathbb{P}_\pi)$ is ergodic

$\Rightarrow$ the Birkhoff theorem applies.

Then, for all $\mathsf{C} \in \mathcal{E}$ such that $\pi(\mathsf{C}) > 0$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \mathbf{1}_{\mathsf{C}}(X_k) = \pi(\mathsf{C}) > 0 \quad \mathbb{P}_\pi - \text{a.s.}$$

Thus, $\mathbb{P}_\pi(\sigma_{\mathsf{C}} < \infty) = 1$. And therefore $\mathbb{P}_x(\sigma_{\mathsf{C}} < \infty) = 1 > 0$ for π-almost all $x \in \mathsf{X}$.

Outline

Using the Kac formula

Recall the formula: $$\pi(f) = \int_{\mathsf{C}} \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_{\mathsf{C}}-1} f(X_k)\right]$$

Algorithm: The memoryless teleportation process

1. **Initialization:** draw Y_0
2. For $k \leftarrow 1$ to n
   a. draw $Y_k^\star \sim P(Y_{k-1}, \cdot)$
   b. If $Y_k^\star \notin \mathsf{C}$, set $Y_k \leftarrow Y_k^\star$
   c. Otherwise, draw $Y_k \sim \pi_{\mathsf{C}}$ where $\pi_{\mathsf{C}}(\cdot) = \frac{\pi(\cdot \cap \mathsf{C})}{\pi(\mathsf{C})}$.

Remarks:

1. The set C is chosen by the user.
2. Choose for example $\mathsf{C} \subset \{x \in \mathsf{E} : \pi(x) \leqslant \epsilon q(x)\}$ where q is a density from which we can draw. If $\pi(x)/(\epsilon q(x))$ is computable for Q-almost all $x \in \mathsf{X}$, then we can draw π_{C} by rejection.

Using the Kac formula

Recall the formula: $$\pi(f) = \int_{\mathsf{C}} \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_{\mathsf{C}}-1} f(X_k)\right]$$

Algorithm: The memoryless teleportation process

1. **Initialization:** draw Y_0
2. For $k \leftarrow 1$ to n
   - a. draw $Y_k^\star \sim P(Y_{k-1}, \cdot)$
   - b. If $Y_k^\star \notin \mathsf{C}$, set $Y_k \leftarrow Y_k^\star$
   - c. Otherwise, draw $Y_k \sim \pi_{\mathsf{C}}$ where $\pi_{\mathsf{C}}(\cdot) = \frac{\pi(\cdot \cap \mathsf{C})}{\pi(\mathsf{C})}$.

Remarks:

1. The set C is chosen by the user.
2. Choose for example $\mathsf{C} \subset \{x \in \mathsf{E} : \pi(x) \leqslant \epsilon q(x)\}$ where q is a density from which we can draw. If $\pi(x)/(\epsilon q(x))$ is computable for Q-almost all $x \in \mathsf{X}$, then we can draw π_{C} by rejection.

Using the Kac formula

Recall the formula: $$\pi(f) = \int_{\mathsf{C}} \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_{\mathsf{C}}-1} f(X_k)\right]$$

Algorithm: The memoryless teleportation process

1. **Initialization:** draw Y_0
2. For $k \leftarrow 1$ to n
   - a. draw $Y_k^\star \sim P(Y_{k-1}, \cdot)$
   - b. If $Y_k^\star \notin \mathsf{C}$, set $Y_k \leftarrow Y_k^\star$
   - c. Otherwise, draw $Y_k \sim \pi_{\mathsf{C}}$ where $\pi_{\mathsf{C}}(\cdot) = \frac{\pi(\cdot \cap \mathsf{C})}{\pi(\mathsf{C})}$.

Remarks:

1. The set C is chosen by the user.
2. Choose for example $\mathsf{C} \subset \{x \in \mathsf{E} : \pi(x) \leqslant \epsilon q(x)\}$ where q is a density from which we can draw. If $\pi(x)/(\epsilon q(x))$ is computable for Q-almost all $x \in \mathsf{X}$, then we can draw π_{C} by rejection.

Using the Kac formula

Recall the formula:

$$\pi(f) = \int_{\mathsf{C}} \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_{\mathsf{C}}-1} f(X_k)\right]$$

Algorithm: The memoryless teleportation process

1. **Initialization:** draw Y_0
2. For $k \leftarrow 1$ to n
   a. draw $Y_k^\star \sim P(Y_{k-1}, \cdot)$
   b. If $Y_k^\star \notin \mathsf{C}$, set $Y_k \leftarrow Y_k^\star$
   c. Otherwise, draw $Y_k \sim \pi_{\mathsf{C}}$ where $\pi_{\mathsf{C}}(\cdot) = \frac{\pi(\cdot \cap \mathsf{C})}{\pi(\mathsf{C})}$.

Remarks:

1. The set C is chosen by the user.
2. Choose for example $\mathsf{C} \subset \{x \in \mathsf{E} : \pi(x) \leqslant \epsilon q(x)\}$ where q is a density from which we can draw. If $\pi(x)/(\epsilon q(x))$ is computable for Q-almost all $x \in \mathsf{X}$, then we can draw π_{C} by rejection.

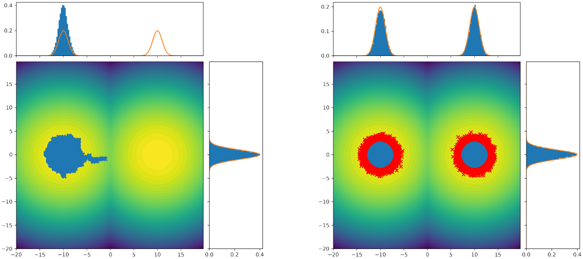

Figure: MALA (left) versus Memoryless teleportation MALA process (right)

Set D $= [-15, 15]^2$ and $(\gamma, \epsilon) = (0.1, 1.3/4\pi)$. Iterations: 10^6.

- **Target:** $\pi = 0.5\mathcal{N}(-a, I_2) + 0.5\mathcal{N}(a, I_2)$ with $a = (10, 0)^T$
- **The Markov kernel P:** MALA with proposal $Y_k = X_k + \gamma\nabla \ln \pi(X_k) + \sqrt{2\gamma}\mathcal{N}(\mathbf{0}, I_2)$
- **The critical set** C $= \{(x, y) \in$ D $: \pi(x, y) \leqslant \epsilon q(x, y)\}$ with q the density of Unif(D).

Some properties of the memoryless teleportation process

Assumption A1

The Markov kernel P allows an invariant probability measure π such that C is π-accessible.

1. **The kernel** S. The memoryless teleportation process is associated to the Markov kernel S defined by: for all $(y, h) \in \mathsf{E} \times \mathsf{F}_+(\mathsf{E})$,

$$Sh(y) = \int_{\mathsf{C}^c} P(y, \mathrm{d}y')h(y') + P(y, \mathsf{C})\pi_{\mathsf{C}}(h)$$

2. **The** π**-invariance**. Integrating wrt π,

$$\pi Sh = \pi P(\mathbf{1}_{\mathsf{C}^c} h) + \pi P(\mathsf{C})\pi_{\mathsf{C}}(h) = \pi(\mathbf{1}_{\mathsf{C}^c} h) + \pi(\mathbf{1}_{\mathsf{C}} h) = \pi(h)$$

and hence, the Markov kernel S is π-invariant.

Some properties of the memoryless teleportation process

Assumption A1

The Markov kernel P allows an invariant probability measure π such that C is π-accessible.

1. **The kernel S.** The memoryless teleportation process is associated to the Markov kernel S defined by: for all $(y, h) \in \mathsf{E} \times \mathsf{F}_+(\mathsf{E})$,

$$Sh(y) = \int_{\mathsf{C}^c} P(y, \mathrm{d}y')h(y') + P(y, \mathsf{C})\pi_{\mathsf{C}}(h)$$

2. **The π-invariance.** Integrating wrt π,

$$\pi Sh = \pi P(\mathbf{1}_{\mathsf{C}^c} h) + \pi P(\mathsf{C})\pi_{\mathsf{C}}(h) = \pi(\mathbf{1}_{\mathsf{C}^c} h) + \pi(\mathbf{1}_{\mathsf{C}} h) = \pi(h)$$

and hence, the Markov kernel S is π-invariant.

Some properties of the memoryless teleportation process

Assumption A1

The Markov kernel P allows an invariant probability measure π such that C is π-accessible.

1. **The kernel** S. The memoryless teleportation process is associated to the Markov kernel S defined by: for all $(y,h) \in \mathsf{E} \times \mathsf{F}_+(\mathsf{E})$,

$$Sh(y) = \int_{\mathsf{C}^c} P(y, \mathrm{d}y')h(y') + P(y, \mathsf{C})\pi_{\mathsf{C}}(h)$$

2. **The** π**-invariance**. Integrating wrt π,

$$\pi Sh = \pi P(\mathbf{1}_{\mathsf{C}^c} h) + \pi P(\mathsf{C})\pi_{\mathsf{C}}(h) = \pi(\mathbf{1}_{\mathsf{C}^c} h) + \pi(\mathbf{1}_{\mathsf{C}} h) = \pi(h)$$

and hence, the Markov kernel S is π-invariant.

The Markov kernel S is not in general π-reversible

Proposition

Assume A1 and $\mathcal{E}_{\mathsf{C}^c}$ is countably generated. Then S is π-reversible if and only if the following two conditions are satisfied,

- (a) $\int_{\mathsf{E}\times\mathsf{E}} \mathbf{1}_{\mathsf{A}}(x,y)\pi(\mathrm{d}x)P(x,\mathrm{d}y) = \int_{\mathsf{E}\times\mathsf{E}} \mathbf{1}_{\mathsf{A}}(y,x)\pi(\mathrm{d}x)P(x,\mathrm{d}y)$ for all $\mathsf{A} \in \mathcal{E}_{\mathsf{C}^c} \otimes \mathcal{E}_{\mathsf{C}^c}$;
- (b) there exists a measure μ on $(\mathsf{C}^c, \mathcal{E}_{\mathsf{C}^c})$ such that $P(y,\cdot)|_{\mathsf{C}^c} = \mu$ for π-almost all $y \in \mathsf{C}$.

1. The condition (b) is restrictive. ▶ The Markov kernel S is "mostly" non-reversible.
2. Caveat: Sampling exactly from π_{C} may restrict the choice of the critical set C.

The Markov kernel S is not in general π-reversible

Proposition

Assume A1 and $\mathcal{E}_{\mathsf{C}^c}$ is countably generated. Then S is π-reversible if and only if the following two conditions are satisfied,

(a) $\int_{\mathsf{E}\times\mathsf{E}} \mathbf{1}_{\mathsf{A}}(x,y)\pi(\mathrm{d}x)P(x,\mathrm{d}y) = \int_{\mathsf{E}\times\mathsf{E}} \mathbf{1}_{\mathsf{A}}(y,x)\pi(\mathrm{d}x)P(x,\mathrm{d}y)$ for all $\mathsf{A} \in \mathcal{E}_{\mathsf{C}^c} \otimes \mathcal{E}_{\mathsf{C}^c}$;

(b) there exists a measure μ on $(\mathsf{C}^c, \mathcal{E}_{\mathsf{C}^c})$ such that $P(y,\cdot)|_{\mathsf{C}^c} = \mu$ for π-almost all $y \in \mathsf{C}$.

1. The condition (b) is restrictive. ▶ The Markov kernel S is "mostly" non-reversible.
2. Caveat: Sampling exactly from π_{C} may restrict the choice of the critical set C.

Outline

The Markov teleportation process

Assumption A2

The Markov kernel Q on $\mathsf{C} \times \mathcal{C}$ allows the restriction π_{C} as invariant probability measure.

THE KEY IDEA

If the candidate $Y_k^\star$ falls into the region C, instead of sampling according to π_{C}, we pick in the past history of $\{Y_k : k \in \mathbb{N}\}$ the last index where the process is in C and we move according to Q.

Assumption A2

The Markov kernel Q on $\mathsf{C} \times \mathcal{C}$ allows the restriction π_{C} as invariant probability measure.

The key idea

If the candidate $Y_k^\star$ falls into the region C, instead of sampling according to π_{C}, we pick in the past history of $\{Y_k : k \in \mathbb{N}\}$ the last index where the process is in C and we move according to Q.

The extended Markov chain

Algorithm: the Markov teleportation process.

1. **Initialization:** draw (Y_0, Z_0)
2. for $k \leftarrow 1$ to n
   1. draw $Y_k^\star \sim P(Y_{k-1}, \cdot)$
   2. if $Y_k^\star \notin \mathsf{C}$, set $(Y_k, Z_k) \leftarrow (Y_k^\star, Z_{k-1})$
   3. Otherwise draw $Z_k \sim Q(Z_{k-1}, \cdot)$ and set $Y_k \leftarrow Z_k$

1. $\{Y_k : k \in \mathbb{N}\}$ is not a Markov chain in general
2. $\{(Y_k, Z_k) : k \in \mathbb{N}\}$ is a Markov chain with kernel

$$Rh(y,z) = \int_{\mathsf{C}^c} P(y, \mathrm{d}y')h(y', z) + P(y, \mathsf{C}) \int_{\mathsf{C}} Q(z, \mathrm{d}z')h(z', z'), \tag{4}$$

The extended Markov chain

Algorithm: the Markov teleportation process.

1. **Initialization:** draw (Y_0, Z_0)
2. for $k \leftarrow 1$ to n
   1. draw $Y_k^\star \sim P(Y_{k-1}, \cdot)$
   2. if $Y_k^\star \notin \mathsf{C}$, set $(Y_k, Z_k) \leftarrow (Y_k^\star, Z_{k-1})$
   3. Otherwise draw $Z_k \sim Q(Z_{k-1}, \cdot)$ and set $Y_k \leftarrow Z_k$

1. $\{Y_k \,:\, k \in \mathbb{N}\}$ is not a Markov chain in general
2. $\{(Y_k, Z_k) \,:\, k \in \mathbb{N}\}$ is a Markov chain with kernel

$$Rh(y,z) = \int_{\mathsf{C}^c} P(y, \mathrm{d}y')h(y',z) + P(y,\mathsf{C}) \int_{\mathsf{C}} Q(z,\mathrm{d}z')h(z',z'), \tag{4}$$

Algorithm: the Markov teleportation process.

1. **Initialization:** draw (Y_0, Z_0)
2. for $k \leftarrow 1$ to n
   1. draw $Y_k^\star \sim P(Y_{k-1}, \cdot)$
   2. if $Y_k^\star \notin \mathsf{C}$, set $(Y_k, Z_k) \leftarrow (Y_k^\star, Z_{k-1})$
   3. Otherwise draw $Z_k \sim Q(Z_{k-1}, \cdot)$ and set $Y_k \leftarrow Z_k$

1. $\{Y_k \,:\, k \in \mathbb{N}\}$ is not a Markov chain in general
2. $\{(Y_k, Z_k) \,:\, k \in \mathbb{N}\}$ is a Markov chain with kernel

$$Rh(y,z) = \int_{\mathsf{C}^c} P(y, \mathrm{d}y') h(y', z) + P(y, \mathsf{C}) \int_{\mathsf{C}} Q(z, \mathrm{d}z') h(z', z'), \tag{4}$$

Define the measure $\bar{\pi}$ by: for all $h \in \mathsf{F}_+(\mathsf{E} \times \mathsf{C})$

$$\bar{\pi}(h) = \int_{\mathsf{C}} \pi(\mathrm{d}x)\, \mathbb{E}_x^P \left[\sum_{k=0}^{\sigma_{\mathsf{C}}-1} h(X_k, x)\right]$$

Lemma

- (a) Assume A1. Then, the marginal of $\bar{\pi}$ on the first component is π.
- (b) If in addition A2 holds, then $\bar{\pi}$ is an invariant probability measure for the Markov kernel R.

Define the measure $\bar{\pi}$ by: for all $h \in \mathsf{F}_+(\mathsf{E} \times \mathsf{C})$

$$\bar{\pi}(h) = \int_{\mathsf{C}} \pi(\mathrm{d}x)\, \mathbb{E}_x^P \left[\sum_{k=0}^{\sigma_{\mathsf{C}}-1} h(X_k, x) \right]$$

Lemma

- **a** Assume A1. Then, the marginal of $\bar{\pi}$ on the first component is π.
- **b** If in addition A2 holds, then $\bar{\pi}$ is an invariant probability measure for the Markov kernel R.

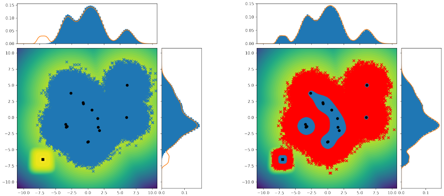

Figure: MALA (left) versus teleportation RW+MALA process (right)

Set $\gamma = 0.8$. Iterations: 10^6

- **Target:** Mixture of Gaussian densities and of a density $\propto \mathrm{e}^{-\|x-x_0\|^4}$
- **The Markov kernel** P**:** MALA with proposal $Y_k = X_k + \gamma \nabla \ln \pi(X_k) + \sqrt{2\gamma}\mathcal{N}(\mathbf{0}, I_2)$
- **The Markov kernel** Q**:** RW with proposal $Y_k = X_k + \sqrt{2\gamma}\mathcal{N}(\mathbf{0}, I_2)$
- **The critical set** $\mathsf{C} = \{(x,y) \in \mathsf{D} : -\log(14.8\pi(x)) > 2\}$

The ergodicity of the associated dynamical system

Proposition

Assume A1 and A2. In addition suppose that

1. π is the unique invariant probability measure for P
2. π_{C} is the unique invariant probability measure for Q

Then, R admits a unique invariant probability measure $\bar{\pi}$.

Consequence: R has a unique invariant probability measure
$\Rightarrow$ the associated dynamical system

$$((\mathsf{E} \times \mathsf{C})^{\mathbb{N}}, (\mathcal{E} \times \mathcal{C})^{\otimes \mathbb{N}}, \Theta, \mathbb{P}^R_{\bar{\pi}})$$

is ergodic
$\Rightarrow$ the Birkhoff theorem applies.

Then, for all functions f such that $\bar{\pi}(|f|) < \infty$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(Y_k, Z_k) = \bar{\pi}(f), \quad \mathbb{P}^R_{\bar{\pi}} - a.s.$$

The ergodicity of the associated dynamical system

Proposition

Assume A1 and A2. In addition suppose that

1. π is the unique invariant probability measure for P
2. π_{C} is the unique invariant probability measure for Q

Then, R admits a unique invariant probability measure $\bar{\pi}$.

Consequence: R has a unique invariant probability measure
$\Rightarrow$ the associated dynamical system

$$((\mathsf{E} \times \mathsf{C})^{\mathbb{N}}, (\mathcal{E} \times \mathcal{C})^{\otimes \mathbb{N}}, \Theta, \mathbb{P}_{\bar{\pi}}^{R})$$

is ergodic
$\Rightarrow$ the Birkhoff theorem applies.

Then, for all functions f such that $\bar{\pi}(|f|) < \infty$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(Y_k, Z_k) = \bar{\pi}(f) \,, \quad \mathbb{P}_{\bar{\pi}}^{R} - a.s.$$

The ergodicity of the associated dynamical system

Proposition

Assume A1 and A2. In addition suppose that

1. π is the unique invariant probability measure for P
2. π_{C} is the unique invariant probability measure for Q

Then, R admits a unique invariant probability measure $\bar{\pi}$.

Consequence: R has a unique invariant probability measure

$\Rightarrow$ the associated dynamical system

$$((\mathsf{E} \times \mathsf{C})^{\mathbb{N}}, (\mathcal{E} \times \mathcal{C})^{\otimes \mathbb{N}}, \Theta, \mathbb{P}^R_{\bar{\pi}})$$

is ergodic

$\Rightarrow$ the Birkhoff theorem applies.

Then, for all functions f such that $\bar{\pi}(|f|) < \infty$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(Y_k, Z_k) = \bar{\pi}(f) , \quad \mathbb{P}^R_{\bar{\pi}} - a.s.$$

Theorem

Assume A1 and A2. In addition suppose that R has a *unique invariant* probability measure $\bar{\pi}$. Then, the two conditions (a) and (b) defined by

a. for any $x \in \mathsf{E}$, $\mathbb{P}_x^P(\sigma_\mathsf{C} < \infty) = 1$;
b. for any $x \in \mathsf{C}$ and any bounded measurable function $h : \mathsf{C} \to \mathbb{R}$,

$$\lim_{\ell \to \infty} \ell^{-1} \sum_{k=0}^{\ell-1} h(X_k) = \pi_\mathsf{C}(h) \,, \qquad \mathbb{P}_x^Q - \text{a.s.}$$

are equivalent to the following property: for any $f : \mathsf{E} \times \mathsf{C} \to \mathbb{R}$ such that $\bar{\pi}(|f|) < \infty$, for any $(y, z) \in \mathsf{E} \times \mathsf{C}$, we get that

$$\lim_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} f(Y_k, Z_k) = \bar{\pi}(f) \,, \qquad \mathbb{P}_{(y,z)}^R - \text{a.s.} \tag{5}$$

Theorem

Assume A1 and A2. In addition suppose that R has a *unique invariant* probability measure $\bar{\pi}$. Then, the two conditions (a) and (b) defined by

a. for any $x \in \mathsf{E}$, $\mathbb{P}_x^P(\sigma_{\mathsf{C}} < \infty) = 1$;
b. for any $x \in \mathsf{C}$ and any bounded measurable function $h : \mathsf{C} \to \mathbb{R}$,

$$\lim_{\ell \to \infty} \ell^{-1} \sum_{k=0}^{\ell-1} h(X_k) = \pi_{\mathsf{C}}(h) , \qquad \mathbb{P}_x^Q - \text{a.s.}$$

are equivalent to the following property: for any $f : \mathsf{E} \times \mathsf{C} \to \mathbb{R}$ such that $\bar{\pi}(|f|) < \infty$, for any $(y, z) \in \mathsf{E} \times \mathsf{C}$, we get that

$$\lim_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} f(Y_k, Z_k) = \bar{\pi}(f) , \qquad \mathbb{P}_{(y,z)}^R - \text{a.s.} \tag{5}$$

First idea: decompose $\sum_{k=0}^{n-1} f(Y_k, Z_k)$ into cycles outside $\mathsf{C} \times \mathsf{C}$ and using triangular arrays,

$$m^{-1} \sum_{\ell=0}^{m-1} \left(\sum_{i=\sigma^{\ell}_{\mathsf{C}\times\mathsf{C}}}^{\sigma^{\ell+1}_{\mathsf{C}\times\mathsf{C}}-1} f(Y_i, Z_i) - \mathbb{E}^{R}_{(Y_{\sigma^{\ell}_{\mathsf{C}\times\mathsf{C}}}, Z_{\sigma^{\ell}_{\mathsf{C}\times\mathsf{C}}})} \left[\sum_{i=0}^{\sigma_{\mathsf{C}\times\mathsf{C}}-1} f(Y_i, Z_i) \right] \right)$$

converges to 0 almost surely as m goes to infinity. ▶ sufficient conditions for the LLN only. Instead, we propose to use:

Proposition

The two following conditions are equivalent.

1. For any $\xi \in \mathsf{M}_1(\mathcal{E})$ and $g : \mathsf{E} \to \mathbb{R}$ measurable, $\pi(|g|) < \infty$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} g(X_k) = \pi(g) , \qquad \mathbb{P}^{P}_{\xi} - \text{a.s.} ,$$

2. Any **bounded harmonic** function h for P (ie $Ph = h$) is **constant**.

First idea: decompose $\sum_{k=0}^{n-1} f(Y_k, Z_k)$ into cycles outside $\mathsf{C} \times \mathsf{C}$ and using triangular arrays,

$$m^{-1} \sum_{\ell=0}^{m-1} \left(\sum_{i=\sigma^{\ell}_{\mathsf{C}\times\mathsf{C}}}^{\sigma^{\ell+1}_{\mathsf{C}\times\mathsf{C}}-1} f(Y_i, Z_i) - \mathbb{E}^{R}_{(Y_{\sigma^{\ell}_{\mathsf{C}\times\mathsf{C}}}, Z_{\sigma^{\ell}_{\mathsf{C}\times\mathsf{C}}})} \left[\sum_{i=0}^{\sigma_{\mathsf{C}\times\mathsf{C}}-1} f(Y_i, Z_i) \right] \right)$$

converges to 0 almost surely as m goes to infinity. ▶ sufficient conditions for the LLN only. Instead, we propose to use:

Proposition

The two following conditions are equivalent.

i. For any $\xi \in \mathsf{M}_1(\mathcal{E})$ and $g : \mathsf{E} \to \mathbb{R}$ measurable, $\pi(|g|) < \infty$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} g(X_k) = \pi(g) \,, \qquad \mathbb{P}^{P}_{\xi} - \text{a.s.} \,,$$

ii. Any **bounded harmonic** function h for P (ie $Ph = h$) is **constant**.

Assumptions for getting the geometric ergodicity

Assumption G1

There exist $(\lambda, b) \in (0,1) \times (0, \infty)$ and a measurable function $V_P : \mathsf{E} \to [1, \infty)$ such that

$$PV_P \leqslant \lambda V_P + b\mathbf{1}_{\mathsf{C}} \quad \text{and} \quad \sup_{y \in \mathsf{C}} V_P(y) < \infty \,.$$

Assumption A1 is a typical geometric drift condition.

We say that $\mathsf{D} \in \mathcal{E}_\mathsf{C}$ is a $(1, \epsilon\nu)$-small set if for any $x \in \mathsf{D}$ and $\mathsf{A} \in \mathcal{E}_\mathsf{C}$, $Q(x, \mathsf{A}) \geqslant \epsilon\nu(\mathsf{A})$.

Assumption G2

1. There exists an accessible $(1, \epsilon\nu)$-small set $\mathsf{D} \in \mathcal{C}$ for Q such that $\nu(\mathsf{D}) > 0$.
2. $\delta = \inf_{z \in \mathsf{D}} P(z, \mathsf{C}) > 0$.
3. There exist constants $(\lambda, b) \in (0,1) \times (0, \infty)$ and a measurable function $V_Q : \mathsf{C} \to [1, \infty)$ such that

$$QV_Q \leqslant \lambda V_Q + b\mathbf{1}_\mathsf{D} \quad \text{and} \quad \sup_{z \in \mathsf{D}} V_Q(z) < \infty.$$

Theorem

Assume A1, A2, G1 and G2. Then, there exist constants $C > 0$ and $\varrho \in (0,1)$ such that for any probability measure μ on $(\mathsf{E} \times \mathsf{C}, \mathcal{E} \otimes \mathcal{C})$ and any $n \in \mathbb{N}$,

$$\|\mu R^n - \bar{\pi}\|_{\mathrm{TV}} \leqslant C\, \varrho^n\, \mu(V_P \otimes V_Q), \tag{6}$$

where $V_P \otimes V_Q$ is the function $V_P \otimes V_Q(y,z) = V_P(y)V_Q(z)$ for all $(y,z) \in \mathsf{E} \times \mathsf{C}$.

Outline

The Markov teleportation process combines two different kernels:

- P with invariant probability measure π
- Q with invariant probability measure π_{C} where π_{C} is the restriction of π to a given set C.

Idea: Replace the auxiliary distribution π_{C} by a more general $\tilde{\pi}$ which is a probability measure on $(\mathsf{E}, \mathcal{E})$ such that for some M,

$$\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\pi}(x) \leqslant M \tag{7}$$

for π-almost all $x \in \mathsf{E}$.

- **Adding a second component** Define $\bar{\mathsf{E}} = \mathsf{E} \times [0,1]$ and $\bar{\mathcal{E}} = \mathcal{E} \otimes \mathcal{B}([0,1])$. Let $\bar{P}$ be the Markov kernel on $\bar{\mathsf{E}} \times \bar{\mathcal{E}}$ defined by: for all $\bar{x} = (x,u) \in \bar{\mathsf{E}}$ and $A \in \bar{\mathcal{E}}$,

$$\bar{P}(\bar{x}, \mathsf{A}) = \int P(x, \mathrm{d}x') \mathbf{1}_{[0,1]}(u') \mathbf{1}_{\mathsf{A}}(x', u') \mathrm{d}u' \,. \tag{8}$$

Notation

- Denote by $\bar{\mathbb{P}}_\mu$ the associated probability measure induced on $(\bar{\mathsf{E}}^{\mathbb{N}}, \bar{\mathcal{E}}^{\otimes \mathbb{N}})$ by the Markov kernel $\bar{P}$ and initial distribution μ.
- Whenever $\bar{\mathbb{E}}_{x,u}[\varphi]$ does not depend on $u \in [0,1]$, we simply write $\bar{\mathbb{E}}_{x,*}[\varphi]$.

Define

$$\bar{\mathsf{C}} = \left\{(x,u) \in \bar{\mathsf{E}} \,:\, u \leqslant \alpha(x)\right\} \quad \text{where} \quad \alpha(x) = \frac{1}{M}\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\pi}(x)$$

$$\sigma_{\bar{\mathsf{C}}} = \inf\left\{k \geqslant 1 \,:\, (X_k, U_k) \in \bar{\mathsf{C}}\right\}$$

Proposition

Let P be a Markov kernel on $\mathsf{E} \times \mathcal{E}$ with invariant probability measure π. Let $\alpha : \mathsf{E} \to [0,1]$ be a measurable function such that $\{\alpha > 0\}$ is π-accessible for P. Then,

$$\pi = \pi_\alpha^0 = \pi_\alpha^1 \tag{9}$$

where for all nonnegative or bounded functions f on $(\mathsf{E}, \mathcal{E})$,

$$\pi_\alpha^0(f) = \int_{\mathsf{E}} \pi(\mathrm{d}x)\alpha(x)\bar{\mathbb{E}}_{x,*}\left[\sum_{k=0}^{\sigma_{\bar{\mathsf{C}}}-1} f(X_k)\right] \; ,$$

$$\pi_\alpha^1(f) = \int_{\mathsf{E}} \pi(\mathrm{d}x)\alpha(x)\bar{\mathbb{E}}_{x,*}\left[\sum_{k=1}^{\sigma_{\bar{\mathsf{C}}}} f(X_k)\right]$$

Extended teleportation process

If P is π-invariant and Q is $\tilde{\pi}$-invariant.

Algorithm 3

1. **Initialization** Draw (Y_0, Z_0).
2. for $k \leftarrow 1$ to n
   - a. Draw $(Y_k^\star, U_k) \sim P(Y_{k-1}, \cdot) \otimes \mathbf{Unif}([0,1])$
   - b. If $U_k \geqslant \alpha(Y_k^\star)$, set $(Y_k, Z_k) \leftarrow (Y_k^\star, Z_{k-1})$
   - c. Otherwise draw $Z_k \sim Q(Z_{k-1}, \cdot)$ and set $Y_k \leftarrow Z_k$.

An alternative to Lines a-b in Algorithm 3 would be to replace them by:

a": Draw $Y_k^\star \sim P(Y_{k-1}, \cdot)$ and conditionally to $Y_k^\star$, draw $B_k \sim \mathrm{Bern}(\alpha(Y_k^\star))$

b": **if** $B_k = 0$ **then** , set $(Y_k, Z_k) \leftarrow (Y_k^\star, Z_{k-1})$

Any MH can be seen as a
particular extended teleportation process by defining

- the probability measure $\tilde{\pi}$ on $(\mathsf{E}, \mathcal{E})$ by

$$\tilde{\pi}(h) = \frac{\int_{\mathsf{E}} \pi(\mathrm{d}x)\alpha(x)h(x)}{\int_{\mathsf{E}} \pi(\mathrm{d}x)\alpha(x)}\,, \quad h \in \mathsf{F}_+(\mathsf{E})$$

  where $\alpha(x) = \int Q(x, \mathrm{d}y)\alpha^{MH}(x, y)$

- the π-invariant Markov kernel P is defined by: for all $x \in \mathsf{E}$, $P(x, \cdot) = \delta_x$
- the $\tilde{\pi}$-invariant Markov kernel Q_α is defined by:

$$Q_\alpha(x, A) = \frac{\int_A Q(x, \mathrm{d}y)\alpha^{MH}(x, y)}{\int_{\mathsf{E}} Q(x, \mathrm{d}z)\alpha^{MH}(x, z)}\,, \quad (x, A) \in \mathsf{E} \times \mathcal{E}$$

Outline

About the teleportation algorithm:

- it allows to combine smoothly two Markov kernels targetting different distributions.
- Embedded sets (C_i).
- Non markovian ways of targetting the auxiliary distribution.
- Practical ways of choosing C or more generally α and $\tilde{\pi}$.