Markov Chain Monte Carlo

→ Target π (probability density)

$\text{Supp}(\pi) \subseteq \mathbb{R}^d$

→ Input: Unnormalized density

- $\frac{\pi(x)}{\pi(y)}$

- $\pi(x) = \frac{e^{-V(x)}}{Z}$

$\nabla \log \pi = \nabla V$ (Score)

→ Output:

$X_1, \ldots, X_n \sim \pi$

$\text{Unif}(X_1, \ldots, X_n) \xrightarrow{n \to \infty} \pi$

Markov chain Monte Carlo:

Construct a Markov Chain whose
→ invariant / stationary measure is the target π.

→ $\pi = e^{-V}/Z$ Have: $e^{-V(x)}$

$X_i \in R^d$

A

→ Markov Chain: $X_1, \dots, X_n \cdots$

$$P(X_{t+1} \mid X_t, \dots, X_1) = P(X_{t+1} \mid X_t)$$

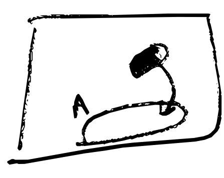
Transition

→ $K(x, A) = P(X_{t+1} \in A \mid X_t = x)$

→ (X_n) is Harris recurrent if $\exists$ a measure π s.t. $\forall \underline{A}$ with $\underline{\pi(A) > 0} \Rightarrow X_n$ visits A infinitely often

$\pi: \mathbb{R}^d \to \mathbb{R}^+$ $\pi(A) = \mathbb{P}(X \in A)$

→ $$\pi(A) = \int_{\mathcal{X}} \underbrace{K(x, A)}_{} \; \underbrace{\pi(dx)}_{}$$



$$\sum_{X_t} \pi(X_t) \cdot P(X_{t+1} = y \mid X_t) = \underline{\pi}(y)$$

→ Ergodic theorem:

$f \in L^1(\pi)$ then

$$\underbrace{\frac{1}{n} \sum_{i=1}^{n} f(X_i)}_{\text{Monte Carlo average}} \xrightarrow{n \to \infty} \int f(x) \pi(dx) = \underset{X \sim \pi}{\underline{E}} f(X)$$

→ CLT for time-inhomogeneous Markov chain with invariant measure π

- Let $g \in L^1(\pi)$ with $Var(g(X)) < \infty$

- $$\hat{g}_N = \frac{1}{N} \sum_{n=0}^{N-1} g(X_n)$$

- $$\sigma^2 := \lim_{N \to \infty} Var(\sqrt{N}\, \hat{g}_N) = Var(g(X)) + 2 \sum_{n=1}^{\infty} cov(g(X_1), g(X_{1+k}))$$

- $$\left(\underline{\hat{g}_N} - \underset{X \sim \pi}{E} \underline{g(X)}\right) \sqrt{N} \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

- Convergence to stationary distribution

For Harris recurrent, aperiodic Markov chain with invariant measure π,

$$\lim_{n \to \infty} \Big\| \underbrace{\int K^n(x, \cdot)\, \mu(dx)}_{} - \pi \Big\|_{TV} = 0$$

$$\| \mu_1 - \mu_2 \|_{TV} = \sup_A | \mu_1(A) - \mu_2(A) |$$

$$\mu_n(A) = \int K^n(x, A)\, \mu(dx)$$

(Prob dist
of samples at time n
that were distributed
according to μ at time 0)

TV: Total variation

→ Mixing time:

$$d_{TV}(\mu_n, \pi) \sim O(e^{-\lambda n})$$

Metropolis-Hastings Algorithm

→ Buffon's needle

Input: unnormalized π, proposal distribution $q(\underset{\text{proposed state}}{x} \mid \underset{\text{current state}}{y})$

Want: samples from π

→ Sample as $X_0 \sim \mu$

→ Propose a new state by sampling from $q(\cdot \mid \underline{X_0})$

$$\tilde{X}_1 \sim q(\cdot \mid X_0)$$

→ $\alpha(\tilde{X}_1, X_0) =$

$$\min\left\{ 1, \frac{\pi(\tilde{X}_1)}{\pi(X_0)} \cdot \frac{q(X_0 \mid \tilde{X}_1)}{q(\tilde{X}_1 \mid X_0)} \right\}$$

Set $X_1 = \tilde{X}_1$ with prob $\alpha(\tilde{X}_1, X_0)$

$X_1 = X_0$ with prob $1-\alpha$.

To show: π is the stationary distribution of this Markov Chain

→ Langevin dynamics
Metropolis Adjusted }

→ Score generative models