Markov Chain Monte Carlo:

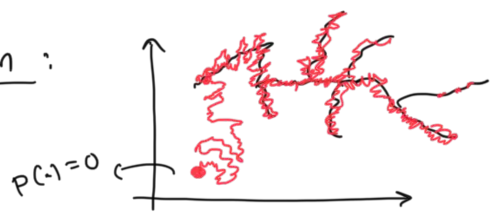
Goal: Sample from $p(x)$, $x \in \mathbb{R}^d$ and compute:

$$\mathbb{E}_{x \sim p(x)}[f(x)] \approx \frac{1}{n} \sum_{i=1}^{n} f(x_i) \;, \quad x_i \overset{\text{i.i.d}}{\sim} p(x)$$

Challenge: $p(x)$, $f(x)$ may be complicated.

$p(x_0) = 0$

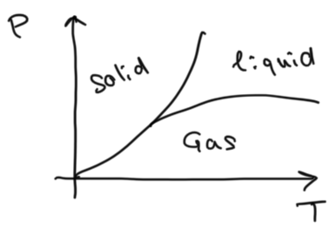
Intuition:



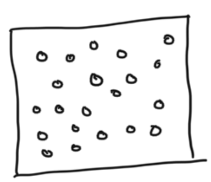
"notional high-dimensional space"

Start at some x_0, find some region of high-probability and then explore the space by moving randomly yet staying close to regions of high-probab

Metropolis 1953:



"Hard disks in a box"



$x \in \mathbb{R}^{2n}$

$\mathbb{E}_{x \sim p(x)}[f(x)]$

$$\boxed{p(x) = \frac{1}{Z_p} e^{-\frac{E(x)}{kT}}}$$

Goal: Start from x_0 and simulate a sequence of state $x_1, \ldots, x_n$ that have high probability under $p(x)$.

Bayesian inference using the Metropolis algorithm:

Given some model with parameters $\theta \in \mathbb{R}^d$ and some data $\mathcal{D}$:

Goal: sample $\theta \sim p(\theta|\mathcal{D})$ $\quad p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)\, p(\theta)}{\ldots}$

Goal: sample ... $p(D)$

$\propto p(D|\theta)\,p(\theta)$

Approach: Construct a Markov Chain with a stationary distribution $\pi(\theta|D) \xrightarrow{i \to +\infty} p(\theta|D)$.

- Proposal distribution / proposal matrix: $\underline{Q_{ab}} = p(x=b \mid x=a)$

  stochastic matrix: $\sum_b Q_{ab} = 1$, $Q_{ab} = (a,b)$, $a,b >$

- $p(\theta|D) = \dfrac{\tilde{p}(\theta|D)}{p(D)}$ $:= p(D|\theta)\,p(\theta)$, $p(D) = \int p(D|\theta)\,p(\theta$

Algorithm:

1. Choose a symmetric proposal matrix Q (Metropolis – Hastings does not require a symmetric Q
2. Initialize a state $\theta_0 \in \mathbb{R}^d$
3. For $i = 0, 1, \dots, n-1$:

   i.) Sample θ from $Q = p(\theta \mid \theta_i)$ (e.g. $\theta = \theta_i + \varepsilon$, $\mathcal{N}($

   ii) Sample $u \sim U(0,1)$

   iii) Accept or reject θ according to the random Metropolis rule:

   $$\text{If} \quad u < \frac{\tilde{p}(\theta|D)}{\tilde{p}(\theta_i|D)} \quad \begin{cases} \text{True}, & \theta_{i+1} = \theta \\ \text{False}, & \theta_{i+1} = \theta_i \end{cases}$$

   $$\hookrightarrow \; u < \frac{p(\theta|D)\,p(\theta)}{p(\theta_i|D)\,p(\theta_i)}$$

4.) Output: $\theta_1, \dots, \theta_n \sim$ M.C.

$$\mathbb{E}_{\theta \sim p(\theta|D)} \left[p(f_\theta(x^*) | \theta | D) \right] \approx \frac{1}{m} \sum_{i=1}^{m} p(f(x^*) | \theta_i, D)$$

$$\theta_i \sim M.C.$$

Theorem: If $(\theta_1, \dots, \theta_n)$ is an irreducible (time-homogen discrete Markov Chain with a stationary distribution $\pi(\theta|D)$, then:

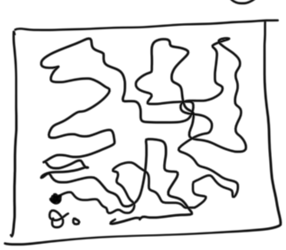
i.) $$\frac{1}{n} \sum_{i=1}^{n} f(\theta_i) \xrightarrow[n \to +\infty]{a.s.} \mathbb{E}_{\theta \sim p(\theta|D)} [f(\theta)]$$ $\forall$ any bound function $f: \Theta \to \mathbb{R}$

If further the chain is aperiodic, then:

$$p(\theta_n | \theta_0, D) \xrightarrow[n \to +\infty]{} p(\theta_n | D)$$

i.e. θ_n is a good sample from the target posterior $p(\theta|D$

Θ



A dynamical system is ergodic if it can reach all possible states in a finite time, regardless o the initial condition θ_0.

What is a Markov Chain?

$$\theta_0 \to \theta_1 \to \dots \to \theta_n$$

$i = 0, 1, \dots, n$

$\Theta := \{\theta_0, \theta_1, \dots, \theta_n\}$

$$\overbrace{p(\theta_i | \theta_0, \theta_1, \dots, \theta_{i-1}) = p(\theta_i | \theta_{i-}}^{\text{Markov property}}$$

$$p(\theta_0, \theta_1, \dots, \theta_n) = p(\theta_0) p(\theta_1 | \theta_0) \dots$$
$$\dots p(\theta_n | \theta_{n-1})$$

If a discrete Markov Chain is irreducible ✓, has a stationary distribution, and is aperiodic ✓, then it is an Ergodic Markov Chain.

Definitions :

1.) A Markov-Chain is time-homogeneous / discrete-time if :

$$p(\theta_{i+1} = b \mid \theta_i = a) = Q_{ab}, \quad \forall i, \forall a,b \in \Theta$$

i.e. the transition probabilities do not depend on time (or the iteration index i)

2.) A pmf π on Θ is a stationary/invariant distribution with respect to Q_{ab} if : $\pi Q = \pi$, i.e. $\sum_{a \in \Theta} \pi_a Q_{ab} = \pi_b, \forall b$

⊛ π is called a left-eigenvector and has eigenvalue 1.

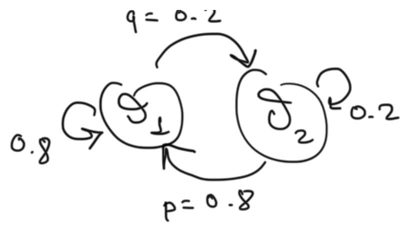
3.) A M.C. is irreducible if $\forall a,b \in \Theta$, $\exists t \geq 0$ s.t.:

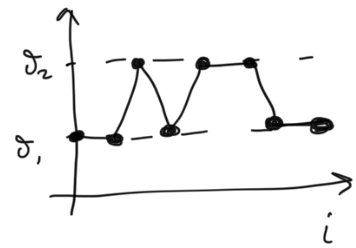
$$p(\theta_t = b \mid \theta_0 = a) > 0, \quad \forall a,b \in \Theta$$

4.) An irreducible M.C. is called aperiodic if :

$$\forall a \in \Theta, \quad \gcd\{t : p(\theta_t = a \mid \theta_0 = a) > 0\} = 1$$

↳ set of times for which if we start a $a \in \Theta$, we get back to a after some time e.g. $\Theta = \{1, 2\}$

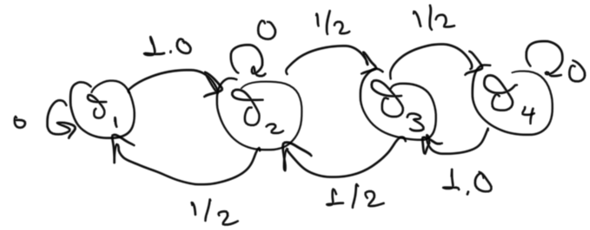




$$Q = \begin{bmatrix} p & q \\ p & q \end{bmatrix}$$

- irreducible ✓
- aperiodic ✓

$\Theta = \{1, 2, 3, 4\}$



$$Q = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

- irreducible ✓
- aperiodic ✗