

Lecture-13: The Monte Carlo Method

1 The Monte Carlo Method

In fields like Statistical Physics and Combinatorial Optimization we often come across the problem of sampling a configuration $X \in \mathcal{X}^N$ from a given distribution $P(X)$. This can be difficult when N is large, because there are too many configurations and/or because the distribution $P(X)$ is specified by Boltzmann formula with computationally infeasible Partition Function. This motivates the study of Monte Carlo Method.

1.1 Properties of Markov Chains

Let $(X_t \in \mathcal{X}^N : t \in \mathbb{N})$ be a Markov chain with state space \mathcal{X}^N , and set of transition rates $w(x \rightarrow y)$ with $x, y \in \mathcal{X}^N$, which satisfy the following conditions.

- (i) **Irreducibility** : For any pair of configurations $x, y \in \mathcal{X}^N$, there exists a path (x_0, x_1, \dots, x_n) of length n , connecting $x_0 = x$ and $x_n = y$ with non-zero probability, i.e. $w(x_i \rightarrow x_{i+1}) > 0$ for all $i \in \{0, \dots, n-1\}$.
- (ii) **Aperiodicity** : For any pair of configurations $x, y \in \mathcal{X}^N \times \mathcal{X}^N$, there exists a positive integer $n(x, y)$ such that for any $n \geq n(x, y)$, there exists a path of length n connecting x and y with non-zero probability. Notice that, for an irreducible chain, aperiodicity is easily enforced by allowing the configuration to remain unchanged with non-zero probability: $w(x \rightarrow x) > 0$.
- (iii) **Stationarity** : There exists a distribution $\pi \in \mathcal{M}(\mathcal{X}^N)$ such that

$$\sum_{x \in \mathcal{X}^N} \pi(x)w(x \rightarrow y) = \pi(y) \quad \forall y \in \mathcal{X}^N.$$

- (iv) **Reversibility** : The transition probability satisfy the detailed balance equation

$$\pi(x)w(x \rightarrow y) = \pi(y)w(y \rightarrow x) \quad \forall x, y \in \mathcal{X}^N.$$

Note : For Finite chains Irreducibility and Aperiodicity implies Stationarity.

Theorem 1.1. Let $X_0, X_1, \dots, X_t, \dots$ be random variables distributed according to the Markov chain with transition probabilities $w(x \rightarrow y)$ and initial condition $X_0 = x_0$. Let the Markov chain satisfy the conditions (i) – (iii). Let $f : \mathcal{X}^N \rightarrow \mathbb{R}$ be any real valued function. Then

1. The probability distribution of X_t converges to the stationary distribution:

$$\lim_{t \rightarrow \infty} P[X_t = x] = \pi(x).$$

2. Time averages converge to averages over the stationary distribution

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t f(X_s) = \sum_x \pi(x)f(x) \quad \text{almost surely.}$$

The **Goal** is to design and simulate a process that converges to the stationary distribution of interest.

2 Metropolis Chains

Goal : Given state space \mathcal{X}^N and stationary distribution π , find a Markov chain with transition probability matrix P such that $\pi P = \pi$.

The approach of Metropolis chains is to first consider an initial base markov chain with certain transition probability matrix and modify the transition probabilities appropriately so that the new chain has the desired stationary distribution.

2.1 Symmetric base chain

Let ψ be a symmetric transition probability matrix of the base chain over \mathcal{X}^N .

Since ψ is symmetric and has vector of ones as an eigenvector, it is easy to see that ψ satisfies reversibility with respect to uniform distribution on \mathcal{X}^N , i.e.

$$\psi \mathbb{1} = \mathbb{1} \implies \frac{\mathbb{1}^T}{|\mathcal{X}^N|} \psi = \frac{\mathbb{1}^T}{|\mathcal{X}^N|} \implies \frac{1}{|\mathcal{X}^N|} \psi(x,y) = \frac{1}{|\mathcal{X}^N|} \psi(y,x) \quad \forall x,y \in \mathcal{X}^N$$

Note that $\psi(x, \cdot)$ is the distribution over next state, starting from state x . Define $a(x,y)$ as the acceptance probability of transition from x to y . The new chain evolves as follows : when at state x , a candidate move is generated from distribution $\psi(x, \cdot)$. If the proposed new state is y , then the move is "accepted" with probability $a(x,y)$ and with remaining probability chain remains at x . Thus the transition probabilities of the new chain becomes,

$$P(x,y) = \begin{cases} \psi(x,y)a(x,y) & \text{if } y \neq x \\ 1 - \sum_{z:z \neq x} \psi(x,z)a(x,z) & \text{if } y = x \end{cases}$$

The transition matrix P has stationary distribution π if reversibility condition holds, i.e.

$$\pi(x)\psi(x,y)a(x,y) = \pi(y)\psi(y,x)a(y,x) \quad \forall x \neq y$$

Define $b(x,y) = \pi(x)a(x,y)$. Thus the condition becomes

$$b(x,y) = b(y,x) \quad \forall x \neq y$$

By definition we see that,

$$\begin{aligned} b(x,y) &\leq \pi(x) \\ b(y,x) &\leq \pi(y) \end{aligned}$$

Thus,

$$\begin{aligned} b(x,y) = b(y,x) &\leq \min\{\pi(x), \pi(y)\} \\ a(x,y) &\leq \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} \end{aligned}$$

Choose

$$a(x,y) = 1 \wedge \left(\frac{\pi(y)}{\pi(x)}\right) := \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\}$$

This choice of acceptance probability is called Metropolis Hastings. The **Metropolis Chain** for probability distribution π and symmetric transition probability matrix ψ is defined as

$$P(x,y) = \begin{cases} \psi(x,y) \left[1 \wedge \frac{\pi(y)}{\pi(x)}\right] & \text{if } y \neq x \\ 1 - \sum_{z:z \neq x} \psi(x,z) \left[1 \wedge \frac{\pi(z)}{\pi(x)}\right] & \text{if } y = x \end{cases}$$

Remark : If the desired stationary distribution π is a Boltzmann distribution, this method has an advantage that we don't need to compute partition function explicitly which is computationally infeasible in many cases.

Example 2.1 (Metropolis algorithm for N -particle system). Consider a system of N Ising spins $\sigma = (\sigma_1 \dots \sigma_N)$ with energy function $E(\sigma)$ and inverse temperature β . We are interested in sampling the Boltzmann distribution $\mu_\beta = \frac{e^{-\beta E(\sigma)}}{Z_\beta}$. The **Metropolis algorithm** with random updatings is defined as follows. Call $\sigma^{(i)}$ the configuration which coincides with σ but for the site i i.e.

$$\sigma^{(i)} : \begin{cases} \sigma_j^{(i)} = \sigma_j & i \neq j \\ \sigma_i^{(i)} = -\sigma_i \end{cases}$$

and let $\Delta E_i(\sigma) \equiv E(\sigma^{(i)}) - E(\sigma)$. At each step, an integer $i \in [N]$ is chosen randomly with flat probability distribution and the spin σ_i is flipped with probability

$$w_i(\sigma) = \exp\{-\beta \max[\Delta E_i(\sigma), 0]\} = \min\left\{1, \frac{\mu_\beta(\sigma^{(i)})}{\mu_\beta(\sigma)}\right\}$$

We can write transition probability matrix as

$$w(\sigma \rightarrow \tau) = \begin{cases} \frac{1}{N} w_i(\sigma) & \tau = \sigma^{(i)} \quad \forall i = 1, 2, \dots, N \\ 1 - \frac{1}{N} \sum_{i=1}^N w_i(\sigma) & \tau = \sigma \\ 0 & \text{otherwise} \end{cases}$$

Consider the detailed balance equation with respect to μ_β

$$\mu_\beta(\sigma) w_i(\sigma) = \min\{\mu_\beta(\sigma), \mu_\beta(\sigma^{(i)})\} = \mu_\beta(\sigma^{(i)}) w_i(\sigma^{(i)})$$

Thus the chain is Irreducible and Reversible.

2.2 General base chain

The Metropolis chain can also be defined when the initial transition matrix is not symmetric. For a general (irreducible) transition matrix ψ and an arbitrary probability distribution π on \mathcal{X}^N , the Metropolized chain is executed as follows. When at state x , generate a state y from $\psi(x, \cdot)$. Move to y with probability

$$\frac{\pi(y)\psi(y, x)}{\pi(x)\psi(x, y)} \wedge 1$$

and remain at x with the complementary probability. The transition matrix P for this chain is

$$P(x, y) = \begin{cases} \psi(x, y) \left[1 \wedge \frac{\pi(y)\psi(y, x)}{\pi(x)\psi(x, y)} \right] & \text{if } y \neq x \\ 1 - \sum_{z: z \neq x} \psi(x, z) \left[1 \wedge \frac{\pi(z)\psi(z, x)}{\pi(x)\psi(x, z)} \right] & \text{if } y = x \end{cases}$$

3 Glauber Dynamics

Glauber Dynamics is similar to Metropolis chains. The chain evolves as follows: From state x , we choose $v \in [N]$ uniformly at random and move to new state according to $\pi^{x, v}$ defined as

$$\pi^{x, v}(y) = \pi(y | \Omega(x, v)) = \begin{cases} \frac{\pi(y)}{\pi(\Omega(x, v))} & \text{if } y \in \Omega(x, v) \\ 0 & \text{if } y \notin \Omega(x, v) \end{cases}$$

where $\Omega(x, v) = \{y \in \mathcal{X}^N : y(w) = x(w) \forall w \neq v\}$.

Example 3.1 (Glauber Dynamics for N-particle system). For this system $\Omega(\sigma, i) = \{\sigma, \sigma^{(i)}\}$. Thus,

$$w_i(\sigma) = \pi^{\sigma, i}(\sigma^{(i)}) = \frac{\mu_\beta(\sigma^{(i)})}{\mu_\beta(\sigma^{(i)}) + \mu_\beta(\sigma)} = \frac{\exp(-\beta E(\sigma^{(i)}))}{\exp(-\beta E(\sigma^{(i)})) + \exp(-\beta E(\sigma))}$$

$$w_i(\sigma) = \frac{1}{1 + \exp(\beta \Delta E_i(\sigma))} = \frac{1}{2} \left[1 - \frac{\exp(\beta \Delta E_i(\sigma)) - 1}{\exp(\beta \Delta E_i(\sigma)) + 1} \right]$$

$$w_i(\sigma) = \frac{1}{2} \left[1 - \tanh \left(\frac{\beta \Delta E_i(\sigma)}{2} \right) \right]$$

We can write transition probability matrix as

$$w(\sigma \rightarrow \tau) = \begin{cases} \frac{1}{N} w_i(\sigma) & \tau = \sigma^{(i)} \quad \forall i = 1, 2, \dots, N \\ 1 - \frac{1}{N} \sum_{i=1}^N w_i(\sigma) & \tau = \sigma \\ 0 & \text{otherwise} \end{cases}$$

Consider the detailed balance equation with respect to μ_β

$$\mu_\beta(\sigma) w_i(\sigma) = \frac{\mu_\beta(\sigma) \mu_\beta(\sigma^{(i)})}{\mu_\beta(\sigma^{(i)}) + \mu_\beta(\sigma)} = \mu_\beta(\sigma^{(i)}) w_i(\sigma^{(i)})$$

Thus the chain is Irreducible and Reversible.

Definition 3.2. Given a configuration space \mathcal{X}^V for a graph $G = (V, E)$, we can define **Glauber dynamics** or **Gibbs sampler** to be a the following reversible Markov chain which has stationary distribution π , and transition probabilities

$$P(x, y) = \begin{cases} \frac{1}{N} \frac{\pi(y)}{\pi(s(x, y))}, & y \in s(x), \\ 0, & y \notin s(x). \end{cases}$$

Here we define $V(x, y) = \{v \in V : x \neq y\}$, and the set of possible transitions $s(x) = \{y \in \mathcal{X}^V : |V(x, y)| = 1\}$, and the set of possible transitions at vertex $V(x, y)$ as $s(x, y) = \{z \in \mathcal{X}^V : z_w = x_w, w \notin V(x, y)\}$.

Exercise 3.3. Show that the Glauber dynamics is a reversible Markov chain with the stationary distribution π .