Markov Chain Monte Carlo Methods

Sherry Sarkar

Big O Meeting

A **Markov chain** is a sequence of random variables $X_0, X_1, ..., X_t$ over some state space $\mathcal{X}$ such that

$$\mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}, ..., X_0 = x_0) = \mathbb{P}(X_t | X_{t-1} = x_{t-1}).$$

This just means that at any time t, the probability of moving to some state $x' \in \mathcal{X}$ is only dependent on the current state.

Most MCs are *time invariant* meaning that we can succinctly represent the MC as a transition matrix P where, at any time t, $P(x, x') = \mathbb{P}(X_{t+1} = x' | X_t = x)$. Note that the rows add up to 1.

Given an initial distribution across the states μ_0, after one step of the MC, we have a new distribution across the states

$$\mu_1 = \mu_0 P.$$

With one more step, we get

$$\mu_2 = \mu_1 P.$$

Thus,

$$\mu_t = \mu_0 P^t$$

We call a MC **irreducible** if for all states $a, b \in \mathcal{X}$, there exists a t such that

$$P^t(a, b) > 0$$

Let $T(x)$ be the set of times $\{t \geq 1 : P^t(x, x) > 0\}$. The **period** of a state x is the greatest common divisor of the set $T(x)$.

Theorem

If a MC is irreducible, then the period of every state is the same.

To gain intuition of what a period is, consider a random walk on a cycle of even length. There are two types of states, even and odd.

Proof:

Say we have two states x and y. Since the MC is irreducible, we know there exits $r, l > 0$ such that $P^r(x, y) > 0$ and $P^l(y, x) > 0$. Say $m = r + l$. Then $m \in T(x)$ and $b \in T(y)$. If we have an element $a \in T(x)$ then a can be represented as $b - m$ where $b \in T(y)$. Thus $T(x) \in T(y) - m$ so the gcd of $T(y)$ divides all elements of $T(x)$. Therefore, gcd $T(y) \leq$ gcd $T(x)$. We can do a parallel argument to get $T(x) \leq$ gcd $T(y)$

A **stationary distribution** of a MC is some probability vector such that

$$\pi = \pi P$$

Theorem

If a MC is irreducible and aperiodic, there exists a unique stationary distribution with all entries greater than 0.

Metropolis Hasting Algorithm - Motivation:

Given that we are at state x,

1. Pick a neighbor y with probability $\frac{1}{\Delta}$ where Δ is the maximum degree of the graph G represented by the MC.
2. Move to y with probability $\min(1, \frac{\pi(y)}{\pi(x)})$
3. With all remaining probability, stay at x.

Theorem

The stationary distribution of this is π assuming the condition of detailed balance.

Three points of interest :

1. Is this even a valid transition matrix?
2. Is π the actual stationary distribution?
3. Is it a unique stationary distribution?

The counting problem : Number of matchings given some graph G

A **randomized approximation scheme** for a counting problem $f : \Sigma \to \mathbb{N}$ is a randomized algorithm that takes an input x, an error tolerance $\epsilon > 0$ and outputs a number N such that the probability that N is within bounds set by this error is greater than 0.5.

An **almost uniform sampler** for a solution set (like the set of matchings for a graph) is a randomized algorithm that takes an input (like a graph) and a sampling tolerance $\delta > 0$ outputs a solution W which is a random variable of the algorithm such that the distance between the distribution of W and a uniform distribution on the solution set is at most δ^3.

Theorem

Let G be a graph with n vertices and m edges, where $m > 1$ to avoid trivialities. If there is an almost uniform sampler for $M(G)$, then there is a randomized approximation scheme for $|M(G)|$.

Some points :

$$|M(G)| = (\alpha_1 \alpha_2 .... \alpha_m)^{-1}$$

where

$$\alpha = \frac{|M(G_{i-1})|}{|M(G_i)|}$$

Note that

$$M(G_{i-1}) \subset M(G_i)$$

and that $M(G_i) - M(G_{i-1})$ can be mapped injectively into $M(G_{i-1})$ by sending M to $M - ei$. Thus,

$$\frac{1}{2} \leq \alpha \leq 1$$

Let Z_i be an indicator function of when $M_i \in M(G_i)$ is in $M(G_{i-1})$.

Markov chain with stationary distribution for an almost uniform sampler: Say you are at state / matching M.

1. With probability $\frac{1}{2}$, set the next state to M.
2. Select $e \in E(G)$ and set $M' = M \oplus e$.
3. If $M' \in M(G)$ then choose M' as your next state. Else, choose next state as M.

1. Why is it irreducible?
2. Why is aperiodic?
3. What is the stationary distribution?

Sources :

1. http://pages.uoregon.edu/dlevin/MARKOV/markovmixing.pdf
2. Mixing. D. Randall. A tutorial on Markov chains in the 44th Symposium on Foundations of Computer Science (FOCS): 4-15, 2003.
3. https://www.math.cmu.edu/ af1p/Teaching/MCC17/Papers/Jerrum