

Monte Carlo: Sampling vs. Counting

Instructor: *Xi Chen*Scribes: *Jun Jiang*

1 Problem

We have $\Omega \subseteq \{0, 1\}^n$ and we want to estimate $|\Omega|$.

Approach: approximately sample an element from Ω uniformly at random.

Example 1. *Example: Counting independent sets in an undirected graph*

$$\Omega : u : \{0, 1\}^n$$

$G = (V, E)$ n vertices

Assumption : \exists a randomized poly algorithm that samples an Independent Set of a given graph G uniformly at random.

Approach: we order the m edges arbitrarily and Let E_i denote the first i edges , $G_i = (V, E_i)$, and $\#IS(G_i)$ be the number of independent sets of G_i

$$\text{We have : } \#IS(G_m) = \frac{\#IS(G_m)}{\#IS(G_{m-1})} \cdot \frac{\#IS(G_{m-1})}{\#IS(G_{m-2})} \cdots \#IS(G_0)$$

Let $T = IS(G_{m-1}) - IS(G_m)$, and $T' = T - \{v\}$

So we have a mapping $T \rightarrow T'$ for every v , and combined with the fact $\#IS(G_m) \geq \#IS(G_{m-1})$. We could make it.

2 Definitions

Definition 2. *Total variation distance*

Two distribution p, q over Ω , the Total Variation Distance is defined following:

$$\|p - q\| = \frac{1}{2} \sum_{x \in \Omega} \|p(x) - q(x)\| = \max_{A \subseteq \Omega} \|p(A) - q(A)\|, \text{ where } p(A) = \sum_{x \in A} p(x)$$

Definition 3. *Markov Chain S*

For a finite state space Ω , x_0, x_1, \dots, x_t is a Markov Chain if

$$Pr[x_t = x | x_{t-1} = y] = p(x, y) \Leftrightarrow Pr[x_t = x | x_0 = x_0, \dots, x_{t-1} = y]$$

Definition 4. *Transition Matrix*

We call P is a Transition Matrix if $p(x, y) \geq 0$ and $\sum_{y \in \Omega} p(x, y) = 1$

Now, it comes to *irreducible* and *aperiodic*

Given P , define a directed graph $G = (\Omega, E)$ edge (x, y) if $p(x, y) > 0$

Definition 5. *irreducible*

We say markov chain is irreducible if the graph is strongly connected

Definition 6. *aperiodic*

We say markov chain is aperiodic if $\gcd\{t : p^t(x, y) > 0\} = 1$. When G is undirected, $\gcd\{t : p^t(x, y) > 0\} > 0$.

Fundamental : If a markov chain is both irreducible and aperiodic, then \exists a unique stationary $\pi : \pi = \pi P$

We use $q^{(t)x}$ to denote the distribution of $\pi(x)$ and x_t if we start with $x_0 = x$

For estimate purpose, we want $\forall x, \lim_{t \rightarrow \infty} \|q_x^{(t)} - \pi\| \rightarrow 0$

We have following notations:

$$\Delta_x(t) = \|q_x^{(t)} - \pi\|$$

$$\Delta^{(t)} = \max_{x \in \Omega} \Delta_x(t)$$

$$\tau_x(\epsilon) = \min\{t : \Delta_x(t) \leq \epsilon\}$$

$$\tau(\epsilon) = \max_{x \in \Omega} \tau_x(\epsilon)$$

Let's look at Random walk on a hypercube on $\{0, 1\}^n$

$x \in \{0, 1\}^n$, we will get the next state randomly pick $j \in [1, n]$ and a bit $b \in [0, 1]$ and set $x_j = b$

\Rightarrow it is irreducible and aperiodic. $n \log n$

Let's look at another example: Graph Coloring

We have a graph $G = (V, E)$ with max degree Δ , try to sample a q coloring uniformly.

$1, \dots, q$ q -coloring

If $q < \Delta$, choose uniformly at random, $v \in V, c \in \{1, \dots, q\}$

If $q > \Delta$ choose the color of v to c if possible

We have some conclusion:

If $q > 4\Delta + 1$

the mixing time $< \tau(\frac{1}{2\epsilon})$

because $\tau(\epsilon) \leq \tau(\frac{1}{2\epsilon}) \cdot \log(\frac{1}{\epsilon})$

We define 2 distributions μ, η on Ω

A joint distribution w over $\Omega \times \Omega$ is a coupling of μ and η

\Leftrightarrow if $\sum_{y \in \Omega} w(x, y) = \mu(x)$ and $\sum_{x \in \Omega} w(x, y) = \eta(y)$

Lemma 7. *Coupling Lemma*

1. if w is a coupling of μ and η

$\Rightarrow Pr_{\{x\}}[x \neq y] \geq \|\mu - \eta\|$

2. \exists a coupling w such that $Pr[x \neq y] = \|\mu - \eta\|$

Proof. With probability $1 - \|\mu - \eta\|$
sample z from A uniform random and set $x = y = z$

With probability $1 - \|\mu - \eta\|$
sample x from C uniform randomly
sample y from B uniform randomly
above 2 sampling is independent on each other

Then, we know $x \sim \mu$ and $y \sim \eta$

So we got $Pr[x \neq y] \geq \|\mu - \eta\|$ from following steps

$$\begin{aligned} Pr[x = y] &= \sum_{z \in \Omega} Pr[x = y = z] \\ &\leq \sum_{z \in \Omega} \min(Pr[x = z] \cdot Pr[y = z]) \\ &= 1 - \|\mu - \eta\| \end{aligned}$$

Now $\Delta_x(t) = \|p_x^{(t)} - \pi\|$

$$\begin{aligned} \Delta_x(t+1) &\leq \Delta_x(t) \\ \|q_x^{(t+1)} - \pi\| &\leq \|q_x^{(t)} - \pi\| \end{aligned}$$

For distributions $q_x^{(t)}$ and π

\exists coupling w of $q_x^{(t)}$ and π such that $Pr_{(x,y) \sim w}[x \neq y] = \|q_x^{(t)} - \pi\|$

$$\|q_x^{(t+1)} - \pi\| \leq Pr_{(x,y) \sim w'}[x \neq y] \leq Pr_{(x,y) \sim w}[x \neq y]$$

where $w' : \text{sample } (x, y) \sim w$

If $x = y$ follow the Markov Chain

$$Pr[x' = y' = z] = Pr(x, z)$$

If $x \neq y$:

x follow MC

y follow MC

above 2 are independent on each other

□

This lecture stops here with one incomplete application example *hypercube*