Introduction to Markov chain and mixing time

Song Chendong

July 24, 2024

Example

A certain frog lives in a pond with two pads, east and west. Every morning, the frog moves to the other pad with probabilty p, and remains where he is with probability $1 - p$.

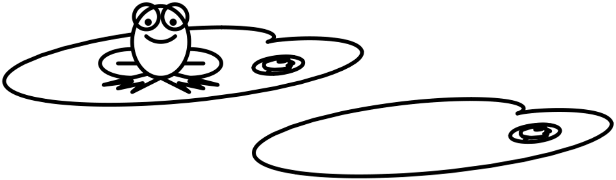

FIGURE 1.1. A randomly jumping frog. Whenever he tosses heads, he jumps to the other lily pad.

- We use the matrix

$$P = \begin{pmatrix} P(e,e) & P(e,w) \\ P(w,e) & P(w,w) \end{pmatrix} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$$

  to describe each move of the frag.
- If a chain has distribution μ at time t, then the chain has distribution μP at time $t+1$. Multiplying a row vector by P on the right takes you from today's distribution to tommorow's distribution.

Markov Chain and Transition Matrix

Definition

A sequence of random variables $(X_0, X_1, ...)$ is a ***Markov chain*** with state space Ω and transition matrix P if for all $x, y \in \Omega$, all $t > 1$, and all events $H_{t-1} = \bigcap_{s=0}^{t-1} \{X_s = x_s\}$ satisfying $\mathbf{P}\left(H_{t-1} \cap \{X_t = x\}\right) > 0$ we have

$$\mathbf{P}\{X_{t+1} = y | H_{t-1} \cap \{X_t = x\}\} = \mathbf{P}\{X_{t+1} = y | X_t = x\} = P(x, y)$$

Markov Chain and Transition Matrix

Definition

A sequence of random variables $(X_0, X_1, ...)$ is a ***Markov chain*** with state space Ω and transition matrix P if for all $x, y \in \Omega$, all $t > 1$, and all events $H_{t-1} = \bigcap_{s=0}^{t-1} \{X_s = x_s\}$ satisfying $\mathbf{P}\left(H_{t-1} \cap \{X_t = x\}\right) > 0$ we have

$$\mathbf{P}\{X_{t+1} = y | H_{t-1} \cap \{X_t = x\}\} = \mathbf{P}\{X_{t+1} = y | X_t = x\} = P(x, y)$$

Remark

We call a non-negative matrix P is ***stochastic***, if

$$\sum_{y \in \Omega} P(x, y) = 1, \forall x \in \Omega$$

A Markov chain is uniquely determined by its starting state and transition matrix.

Irreducibility and Periodicity

Definition

- A chain P is called **irreducible** if for any two states $x, y \in \Omega$, there exists an integer t (depending on x, y) such that $P^t(x, y) > 0$.
- Let $\mathcal{T}(x) := \{t \geq 1 : P^t(x, x) > 0\}$ be the set of times when it is possible for the chain to return to starting position x. The **period of state x** is defined to be the great common divisor of $\mathcal{T}(x)$. The **periodic of chain** is the common period of all states.
- The chain is **aperiodic** if all states have period 1. If a chain is not aperiodic, we call it **periodic**.

Theorem

If P is irreducible and aperiodic, then there is an integer r such that $P^r(x, y) > 0$ for all x, y in Ω

Hitting Time, Recurrence and transience

Definition

- For $A \subset \Omega$, let us define the **hitting time** of A

$$\tau_A = \min\left\{n \geq 0 : X_n \in A\right\},$$

and

$$\tau_A^+ = \min\left\{n \geq 1 : X_n \in A\right\};$$

Especially, if $A = \{x\}$, we often write τ_A as τ_x.

- For a countable Markov chain X_n, a state $x \in \Omega$ is called **recurrent** if $\mathbf{P}_x\left[\tau_x^+ < \infty\right] = 1$; **transient** if $\mathbf{P}_x\left[\tau_x^+ < \infty\right] < 1$.

Hitting time

Theorem

For any state x *and* y *of an irreducible chain,* $\mathbf{E}_x(\tau_y^+) < \infty$

Theorem

For any $x \in \Omega$*, x is recurrent if and only if* $\sum_{t\geq 1} \mathbf{P}_x(X_t = x) = \infty$.

Stationary Distribution

Definition

If a probability distribution π satisfy $\pi = \pi P$, i.e. $\sum_{x \in \Omega} P(x, y)\pi(x) = \pi(y)$, then we call π satisfying a **stationary distribution**.

Theorem

Let P be a transition matrix of an irreducible Markov chain, then there exists a unique probability distribution π.

Simple Random Walk on Graphs

Given a graph $G = (V, E)$, we can define simple random walk on G to be the Markov chain with state space V and transition matrix

$$P(x, y) = \begin{cases} \frac{1}{\deg(x)} & \text{if } y \sim x \\ 0 & \text{otherwise} \end{cases}$$

That is to say, when the chain is at vertex x, it examines all the neighbors of x, picks one uniformly at random, and moves to the chosen vertex.

Example

Consider the graph G shown in Figure 1.4. The transition matrix of simple random walk on G is

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

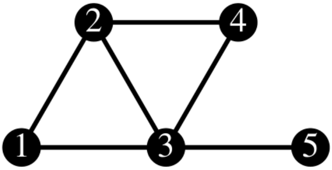

FIGURE 1.4. An example of a graph with vertex set $\{1, 2, 3, 4, 5\}$ and 6 edges.

Example

We conclude that the probability measure

$$\pi(y) = \frac{\deg(y)}{2|E|} \quad \text{for all } y \in \mathcal{X}$$

which is proportional to the degrees, is always a stationary distribution for the walk. For the graph in Figure 1.4,

$$\pi = \left(\frac{2}{12}, \frac{3}{12}, \frac{4}{12}, \frac{2}{12}, \frac{1}{12}\right)$$

Simple Random Walk on $\mathbb{Z}^d$

Consider a walker walking randomly on the grid $\mathbb{Z}^d$, if at a given time the walker is at x , then at the next time moment it will be at one of x's $2d$ neighbours chosen uniformly at random. This process is called **simple random walk in dimension d**.

Theorem

Simple random walk in dimension d is recurrent for $d = 1, 2$ and transient for $d \geq 3$.

"a drunken man always returns home, but a drunken bird will eventually be lost."

Gambler's ruin

Example

Assume that a gambler making fair unit bets on coin flips will abandon the game when her fortune falls to 0 or rise to n. Let X_t be gambler's fortune at time t and τ be the time required to be absorbed at one of 0 or n.

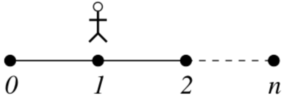

FIGURE 2.1. How long until the walk reaches either 0 or n? What is the probability of each?

Proposition

Assume that $X_0 = k$. Then $\mathbf{P}\{X_\tau = n\} = \dfrac{k}{n}$ and $\mathbf{E}(\tau) = k(n-k)$.

Gambler's Run

Proof.

Let p_k be the possibility that the gambler reaches a fortune of n before ruin starting with k. Clearly $p_0 = 0$ and $p_n = 1$, while

$$p_k = \frac{1}{2}p_{k-1} + \frac{1}{2}p_{k+1}(1 \leq k \leq n-1)$$

Solving this equation yields $p_k = \dfrac{k}{n}$ for $0 \leq k \leq n$. Also, write f_k be the expectation that the steps the gambler takes before ruin starting with k. So $f_0 = 0$ and $f_n = 0$. It's also true that

$$f_k = \frac{1}{2}(f_{k-1}+1) + \frac{1}{2}(f_{k+1}+1) \quad (1 \leq k \leq n-1)$$

Solving this equation by letting $g_k = f_k + k^2$ yields $f_k = k(n-k)$. □

Example

(YCMC2016) Consider the numbers 1, 2, .., 12 written around a ring as they usually are on a clock. A random walker starts at 12 and at each step moves at random to one of its two nearest neighbors (with probability half-half). What is the probability that she will visit all the other numbers before her first returning back to 12?

Coupon Collecting

Example

Consider a collector attempting to collect a complete set of coupons. Assume that each new coupon is chosen uniformly and independently from the set of n possible types, and let τ be the random number of coupons collected when the set first contains every type.

Proposition

- $\mathbf{E}(\tau) = n\sum_{k=1}^{n} \dfrac{1}{k} \sim n \log n.$
- For any $c > 0$. $\mathbf{P}\{\tau > n\log n + cn\} \leq e^{-c}.$

Proof.

Let τ_k be the total number of coupons accumulated when the collection first contains k coupons. Note that $\tau_k - \tau_{k-1}$ has geometric distribution $G(\frac{n-k+1}{n})$, so its expectation is $\frac{n}{n-k+1}$. Therefore,

$$\mathbf{E}(\tau) = \sum_{k=1}^{n} \mathbf{E}(\tau_k - \tau_{k-1}) = n \sum_{k=1}^{n} \frac{1}{n-k+1} = n \sum_{k=1}^{n} \frac{1}{k}.$$

Also let A_i be the event that i-th coupon does not appear. Observe that

$$\mathbf{P}\{\tau > n \log n + cn\} \leq \sum_{i=1}^{n} \mathbf{P}(A_i) \leq n(1 - \frac{1}{n})^{n \log n + cn} \leq e^{-c}.$$

$\square$

Lottery of Genshin

Example

Rule: In the [1,73] draw, the probability of lottery is 0.6%, in the [74,90] draw the probability of lottery increases by 6% if the last draw doesn't lottery.

genshin.cpp

```cpp
#include<iostream>
using namespace std;
int main()
{
    double p[100];
    double e,pr,var;
    for (int i=1; i<=90;i++)
    if (i<=73) p[i]=0.006;
    else p[i]=p[i-1]+0.06;
    e=0;pr=1;
    for (int i=1;i<=90;i++)
    {
        e=e+i*pr*p[i];
        var=var+pr*i*i*p[i];
        pr=pr*(1-p[i]);
    }
    cout<<e<<endl;
    cout<<var-e*e<<endl;
}
```

C:\Users\Starswalker\Docume

```
62.2973
591.086

--------------------------------
Process exited after 0.01746 seconds with return value 0
请按任意键继续. . .
```

Random Walk on Hypercube

Example

The **n-dimensional hypercube** is a graph whose vertices are the binary $n-$tuple $\{0,1\}^n$. Two vertices are connected by the edge when they differ in exactly one coordinate.

The **simple random walk** on the hypercube moves from a vertex $(x^1, x^2, \cdots, x^n)$ by choosing a coordinate $j \in \{1, 2, \cdots n\}$ uniformly randomly and setting the new state as

$(x^1, \cdots, x^{j-1}, 1 - x^j, x^{j+1}, \cdots, x^n)$.

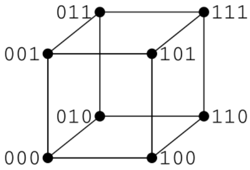

FIGURE 2.2. The three-dimensional hypercube.

Example

Suppose n balls are distributed among two urns, I and II. At each move, a ball is selected uniformly and transferred from its current urn to the other urn. If X_t is the number of balls in urn I at time t, then the transition matrix for X_t is

$$P(j,k) = \begin{cases} \frac{n-j}{n} & \text{if } k = j+1 \\ \frac{j}{n} & \text{if } k = j-1 \\ 0 & \text{otherwise.} \end{cases}$$

Claim

The stationary distribution for this chain is $B(n, \frac{1}{2})$.

Random Walk on Hypercube and Ehrenfest Urn Model

Define the **Hamming weight** $W(x)$ of a vector $x := (x^1, x^2, \cdots, x^n) \in \{0,1\}^n$ to be its number of coordinates with value 1:

$$W(x) = \sum_{j=1}^{n} x^j$$

Then let X_t be the simple random walk on the n-dimensional hypercube, then $W_t = W(X_t)$ is a Markov chain with transisiton probabilities given in Ehrenfest Urn Model.

Convergence Theorem

Theorem

Suppose that P is irreducible and aperiodic, with stationary distribution π. Then there exist constants $\alpha \in (0,1)$ and $C > 0$ such that

$$\max_{x\in\Omega} ||P^t(x,\cdot) - \pi||_{TV} \leq C\alpha^t.$$

where TV is the total variation distance of two distribution,

$$||\mu - \nu||_{TV} = \max_{A\subset\Omega} |\mu(A) - \nu(A)| = \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \nu(x)|$$

Mixing Time

Definition

Let $d(t) := \max_{x\in\Omega} ||P^t(x,\cdot) - \pi||_{TV}$,
the **mixing time** of a Markov chain is defined as
$t_{mix}(\epsilon) := \min_t\{d(t) \leq \epsilon\}$, and $t_{mix} = t_{mix}(\frac{1}{4})$.

There are many ways to estimate the upper bound and the lower bound of the mixing time. Some strategies are listed as follows:

- **Coupling** $||P^t(x,\cdot) - P^t(y,\cdot)||_{TV} \leq \mathbf{P}\{\tau_{couple} > t\}$.
- **Strong stationary time** $d(t) \leq \max_{x\in\Omega} \mathbf{P}_x\{\tau > t\}$ if τ is a strong stationary time.
- **Bottleneck ratio and diameter of the graph** $t_{mix}(\epsilon) \geq L/2, t_{mix} \geq \frac{1}{4\Phi_\star}$.
- **Relaxion time** $t_{mix}(\epsilon) \leq \log(\frac{1}{\epsilon\pi_{min}})t_{rel}$.

Reference

Levin D A, Peres Y. Markov chains and mixing times[M]. American Mathematical Soc., 2017.