[AI2613 Lecture 5] FT of Countably Infinite Markov Chains, Some Applications of Markov Chains

June 10, 2023

1 Recurrence and Positive Recurrence

Recall that we say a state i is *recurrent* if $\mathbf{P}_i[T_i < \infty] = 1$. This is equivalent to $\mathbf{E}_i[N_i] = \infty$. Otherwise, we say the state is *transient*. A transient state j will be visited for finite times with probability 1. From Proposition 3 of last lecture, we know that *recurrence* is a class property, that is, given a recurrent state i, all the other states that i can reach in finite steps are also recurrent. We are only concerned with irreducible Markov chains in this lecture. So we may say a Markov chain is recurrent or transient in the future.

In fact $\mathbf{P}_i[T_i < \infty] = 1 \iff \mathbf{P}_i[N_i = \infty] = 1 \iff \mathbf{E}_i[N_i] = \infty$. I will leave the proof of this as an exercise.

Here we follow the notations of the last lecture, that is: $X_0, X_1, \dots, X_t, \dots$ is a sequence of variables that follows the Markov chain P. $T_i \triangleq \inf\{t > 0 : X_t = i\}$, $N_i \triangleq \sum_{t=0}^{\infty} \mathbb{1}[X_t = i]$, $\mathbf{P}_i[\cdot] = \mathbf{Pr}[\cdot | X_0 = i]$ and $\mathbf{E}_i[\cdot] = \mathbf{E}[\cdot | X_0 = i]$.

本节详细描述了如何证明 Inf MC 的收敛性.

更正一下：Inf MC的收敛性就是指是否Recurrent.

Example 1 (Drunk person and drunk bird) *Imagine a random walk on a grid that we pick a direction uniformly at random at each time step. Can we go back to the original point with probability 1? Or equivalently, is this Markov chain recurrent or transient?*

First we consider the one-dimensional grid. Let $X_0 = 0$ and $X_{t+1} = X_t + \Delta$ where Δ is uniformly at random picked from $\{-1, 1\}$. Then,

$$\mathbf{E}_0[N_0] = \mathbf{E}_0[\sum_{t=0}^{\infty} \mathbb{1}[X_t = 0]] = \sum_{t=0}^{\infty} \mathbf{P}_0[X_t = 0] = \sum_{m=0}^{\infty} \mathbf{P}_0[X_{2m} = 0].$$

where the last equality follows from the fact that we can not go back within exactly odd steps. Then let's compute $\mathbf{P}_0[X_{2m} = 0]$ using the Stirling's formula. For $m \geq 1$,

Stirling's formula: $n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n (1 + o(1))$.

$$\mathbf{P}_0[X_{2m} = 0] = \frac{\binom{2m}{m}}{2^{2m}} \approx \frac{\sqrt{4\pi m}\left(\frac{2m}{e}\right)^{2m}}{2\pi m\left(\frac{m}{e}\right)^{2m}} \cdot 2^{-2m} = \frac{1}{\sqrt{\pi m}}.$$

Thus, $\mathbf{E}_0[N_0] = \sum_{m=0}^{\infty} \mathbf{P}_0[X_{2m} = 0] \approx 1 + \sum_{m=1}^{\infty} \frac{1}{\sqrt{\pi m}}$ which is divergent. This indicates that the Markov chain for random walk on one-dimensional grid is recurrent.

For d-dimensional grid, we regard the game as independently pick Δ_i u.a.r. from $\{-1, 1\}$ for $i \in [d]$ at each time step and walk to $X_{t+1} = X_t + (\Delta_1, \Delta_2, \dots, \Delta_d)$. So we have that $\mathbf{P}_i[X_{2m} = \mathbf{0}] = (\mathbf{P}_i[X_{2m}(1) = 0])^d \approx \left(\frac{1}{\sqrt{\pi m}}\right)^d$. We know that $1 + \sum_{m=1}^{\infty}\left(\frac{1}{\sqrt{\pi m}}\right)^d$ is divergent if and only if $d \leq 2$. Thus, only if the dimension of the grid is 1 or 2, the random walk is recurrent.

尽管$d\le 2$时是Recurrent，但是如果计算$E[T_i]$，即使$P_\sim[X_t=0]<\infty$，但是计算步数的期望要乘t，因此$E[T_i]\to\infty$对$\forall$ dim RW 都成立.

Definition 1 (Positive recurrence) *If a state i is recurrent and $\mathbf{E}_i[T_i] < \infty$, we say it is* positive recurrent. *If the state is recurrent but with $\mathbf{E}_i[T_i] = \infty$, then we say it is* null recurrent.

Example 2 (Drunk person) *We have proved that the Markov chain of drunk person is recurrent. One can show that, even in one-dimension, the chain is null transient (exercise).*

我们将$E_i[T_i]=\infty$的[R] MC称为Null Recurrent MC

将$E_i[T_i]<\infty$的称为Positive MC

1.1 1-D Random Walk

Consider the following one-dimensional random walk:

Let X_t be the position at time step t. Let $T_{i\to j}$ be the first hitting time of state j starting from i, that is, $T_{i\to j} = \min\{t > 0 | X_t = j \wedge X_0 = i\}$. Define event $\mathscr{A}$ = [the first step is to the right]. Then we consider the problem that when will this Markov chain be recurrent. Note that

[R]定义 ←

$$\begin{aligned}
\underbrace{\mathbf{Pr}\left[T_{0\to 0} < \infty\right]} &= \mathbf{Pr}\left[T_{0\to 0} < \infty | \bar{\mathscr{A}}\right]\mathbf{Pr}\left[\bar{\mathscr{A}}\right] + \mathbf{Pr}\left[T_{0\to 0} < \infty | \mathscr{A}\right]\mathbf{Pr}\left[\mathscr{A}\right] \\
&= (1-p)\cdot 1 + p\cdot \mathbf{Pr}\left[T_{1\to 0} < \infty\right], && (1) \\
\mathbf{Pr}\left[T_{1\to 0} < \infty\right] &= \mathbf{Pr}\left[T_{1\to 0} < \infty | \bar{\mathscr{A}}\right]\mathbf{Pr}\left[\bar{\mathscr{A}}\right] + \mathbf{Pr}\left[T_{1\to 0} < \infty | \mathscr{A}\right]\mathbf{Pr}\left[\mathscr{A}\right] \\
&= (1-p)\cdot 1 + p\cdot \mathbf{Pr}\left[T_{2\to 0} < \infty\right], && (2) \\
\mathbf{Pr}\left[T_{2\to 0} < \infty\right] &= \mathbf{Pr}\left[T_{2\to 1} < \infty \wedge T_{1\to 0} < \infty\right] \\
&= \mathbf{Pr}\left[T_{2\to 1} < \infty\right]\cdot \mathbf{Pr}\left[T_{1\to 0} < \infty\right] \\
&= \mathbf{Pr}\left[T_{1\to 0} < \infty\right]^2. && (3)
\end{aligned}$$

Let $y \triangleq \mathbf{Pr}\left[T_{1\to 0} < \infty\right]$ for brevity. Combine Equation (2) and Equation (3), we have $y = 1 - p + py^2$ which then yields $y = 1$ or $y = \frac{1-p}{p}$. By Equation (1), $\mathbf{Pr}\left[T_{0\to 0} < \infty\right] = 1$ or $2 - 2p$.

- When $p < \frac{1}{2}$, $2 - 2p$ is meaningless as a probability. So $\mathbf{Pr}\left[T_{0\to 0} < \infty\right] = 1$ and the Markov chain is recurrent.
- When $p = \frac{1}{2}$, $2 - 2p = 1$. The Markov chain is also recurrent in this situation.
- When $p > \frac{1}{2}$, we verify that $\mathbf{Pr}\left[T_{0\to 0} < \infty\right] < 1$, and therefore $\mathbf{Pr}\left[T_{0\to 0} < \infty\right] = 2 - 2p$. Let $\{\Delta_k\}_{k=0}^{\infty}$ be a sequence of i.i.d. random variables with

$$\Delta_k = \begin{cases} +1, & \text{w.p. } p \\ -1, & \text{w.p. } 1-p \end{cases}.$$

Given a sufficiently large $n \in \mathbb{N}$, we can walk to n from 0 in n steps (i.e. $X_n = n$) with probability $p^n > 0$. Assume that we have arrived at n, consider the probability that we go back to 0 from n in exactly k steps. Apparently, this probability is zero when $k < n$. For every $k \geq n$, we

在上一讲中我们通过分析概率分布得出p<½存在稳态分布而p>½时不存在稳态分布的结论，
然而，并不能由此就确定RW是否Recurrent.
现对该问题作Recurrency证明

我们使用Recurrence定义来证明[R]的条件：
通过分析1-d RW，可得左(1)、(2)、(3)方程，联立解得"Pr[T0→0<∞]=1 or 2-2p"
由此可知p≤½时，Pr=1，是常返的，当然，我们也可猜到p<½是positive [R]，p=½是Null的
之后也会证，而p>½时，留解Pr=2-2p 之后也会证，故不是常返的。

upper bound the probability $\mathbf{Pr}\left[T_{n\to 0} = k\right]$:

$$\begin{aligned}
\mathbf{Pr}\left[T_{n\to 0} = k\right] &\le \mathbf{Pr}\left[\sum_{t=1}^{k} \Delta_t = -n\right] \\
&\le \mathbf{Pr}\left[\sum_{t=1}^{k} \Delta_t - \mathbf{E}\left[\sum_{t=1}^{k} \Delta_t\right] \le -n - \mathbf{E}\left[\sum_{t=1}^{k} \Delta_t\right]\right] \\
&\le \exp\left\{-\frac{2k\left(\frac{n+(2p-1)k}{k}\right)^2}{4}\right\}.
\end{aligned}$$

where the third inequality follows from the Hoeffding's inequality.

该部分解释了 $p>\frac{1}{2}$ 时取解 $p=2-2p$ 的原因，从而证其 transient.

大概思路是说 $p>\frac{1}{2}$ 时，计算够远的点 n 回 0 处的概率可用期望和 concentration ineq bound 住. 由此得 $P[T_{n\to 0}<\infty]$ 的概率 <1，得证 transient.

Then we calculate the probability that we can go back from n to 0. By union bound,

$$\begin{aligned}
\mathbf{Pr}\left[T_{n\to 0} < \infty\right] &= \mathbf{Pr}\left[\bigcup_{k\ge n} \left[T_{n\to 0} = k\right]\right] \\
&\le \sum_{k=n}^{\infty} \mathbf{Pr}\left[T_{n\to 0} = k\right] \\
&\le \exp\{-(2p-1)n\} \sum_{k=n}^{\infty} \exp\left\{-\frac{n^2}{2k} - \frac{(2p-1)^2 k}{2}\right\}.
\end{aligned}$$

Note that

$$\begin{aligned}
\sum_{k=n}^{\infty} \exp\left\{-\frac{n^2}{2k}\right\} \cdot \exp\left\{-\frac{(2p-1)^2 k}{2}\right\} &\le \sum_{k=n}^{\infty} \exp\left\{-\frac{(2p-1)^2 k}{2}\right\} \\
&= \frac{\exp\left\{-\frac{(2p-1)^2}{2} n\right\}}{1 - \exp\left\{-\frac{(2p-1)^2}{2}\right\}}
\end{aligned}$$

Thus,

$$\mathbf{Pr}\left[T_{n\to 0} < \infty\right] \le \frac{\exp\left\{-\frac{(2p-1)^2}{2} n - (2p-1)n\right\}}{1 - \exp\left\{-\frac{(2p-1)^2}{2}\right\}}. \tag{4}$$

We can find a sufficiently large constant n such that $\mathbf{Pr}\left[T_{n\to 0} < \infty\right] < 1$ since the RHS of Equation (4) is exponentially small with regard to n. So for sufficiently large n, the probability that we walk to n and never come back to 0 is larger than $p^n \cdot \mathbf{Pr}\left[T_{n\to 0} = \infty\right] > 0$. Thus, this Markov chain is transient.

Now we verify that the Markov chain is positive recurrent when $p < \frac{1}{2}$ and null recurrent when $p = \frac{1}{2}$. Note that

接下来在证明 $p<\frac{1}{2}$ 为 PR 和 $p=\frac{1}{2}$ 为 NR，方法是使用类似证 $Pr[T_{n\to 0}<\infty]$ 的技巧，证 $T_{0\to 0}$ 的值，可证在 $p=\frac{1}{2}$ 时，$E[T]=\infty$，$p<\frac{1}{2}$ 时 $E[T]=\frac{1-p}{1-2p}<\infty$

得证.

$$T_{0\to 0} = \mathbb{1}[\bar{\mathcal{A}}] \cdot 1 + \mathbb{1}[\mathcal{A}](1 + T_{1\to 0}) \tag{5}$$

$$T_{1\to 0} = \mathbb{1}[\bar{\mathcal{A}}] \cdot 1 + \mathbb{1}[\mathcal{A}](1 + T_{2\to 0}) \tag{6}$$

$$T_{2\to 0} = T_{2\to 1} + T_{1\to 0}. \tag{7}$$

Note that $\mathbf{E}[T_{2\to 1}] = \mathbf{E}[T_{1\to 0}]$. Taking the expectation of Equation (6) and combining with Equation (7), we have

$$\mathbf{E}[T_{1\to 0}] = 1 - p + p(1 + 2\mathbf{E}[T_{1\to 0}]),$$

which yields $\mathbf{E}[T_{1\to 0}] = \frac{1}{1-2p}$. Take the expectation of Equation (5), we get $\mathbf{E}[T_{0\to 0}] = \frac{1-p}{1-2p}$. Thus:

- When $p = \frac{1}{2}$, $\mathbf{E}[T_{0\to 0}] = \infty$ and the Markov chain is null recurrent.
- When $p < \frac{1}{2}$, $\mathbf{E}[T_{0\to 0}] < \infty$ and the Markov chain is positive recurrent.

2 *Some Applications*

接下来举两个MC的应用.
首先是一个遗传问题,用MC的建模
目的是评估后代的个数分布与灭绝概率(断代概率)的关系

2.1 *Galton-Watson Process*

The model was formulated by F. Galton in the study of the survival and extinction of family names. In the nineteenth century, there was concern amongst the Victorians that aristocratic surnames were becoming extinct. In 1873, Galton originally posed the question regarding the probability of such an event, and later H. W. Watson replied with a solution.

Using more modern terms, the process can be defined formally as follows:

Definition 2 (Galton-Watson Process) *Suppose that all the individuals reproduce independently of each other and have the same offspring distribution. More precisely, let G_t denote the number of individuals of t-th generation:*

- *We start from the zero generation. For convenience, let $G_0 = 1$.*
- *Each individual of generation t gives birth to a random number of children of generation $t+1$. That is, $\forall t \ge 0$ and $i \in [G_t]$, let $X_{t,i}$ denote the number of children of the i-th individual in the t-th generation. Then $\{X_{t,i}\}$ is an array of i.i.d. random variables with $\mathbf{Pr}[X_{t,i} = k] = p_k$.*
- *All individuals of generation $t+1$ are children of individuals of generation t:*

$$G_{t+1} = \sum_{i=1}^{G_t} X_{t,i}$$

对该问题进行建模,发现
每代的人数$\{G_t\}$是一个MC.

It is clear that the process $\{G_t\}_{t\ge 0}$ is a Markov chain.

Denote by ρ the probability of extinction, namely

$$\rho \triangleq \mathbf{Pr}[\text{extinction}] = \mathbf{Pr}[\cup_{t\ge 1}\{G_t = 0\}].$$

Then the question is to determine the value of ρ. First we consider two trivial situations:

- When $p_0 = 0$, it is clear that there will be offspring and $\rho = 0$.

首先是初代无后代和所有人只产生
一个后代的情况,前者trivially灭绝
后者通过MC方法可证 $E[G_t] \xrightarrow[t\to\infty]{} 0$
$\therefore Pr[G_t \ge 1] < E[G_t] \to 0$
故此时均寄了

- When $p_0 > 0$ and $p_0 + p_1 = 1$, we can verify that $\rho = 1$. We know that

$$\mathbf{E}\left[G_{t+1} | G_t\right] = p_1 \cdot G_t.$$

Compute the expectation of both sides, we have $\mathbf{E}\left[G_{t+1}\right] = p_1 \mathbf{E}\left[G_t\right]$. This yields that when $t \to \infty$, $\mathbf{Pr}\left[G_t \geq 1\right] \leq \mathbf{E}\left[G_t\right] = p_1^t \mathbf{E}\left[G_0\right] \to 0$.

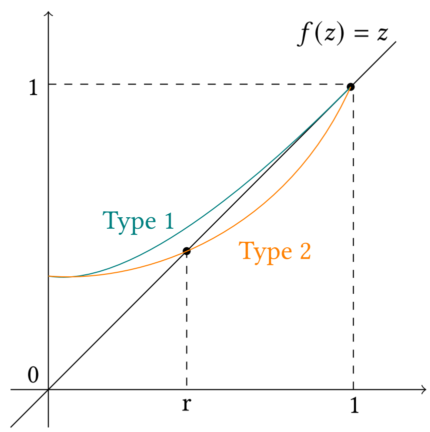
Then we assume that $p_0 > 0$ and $p_0 + p_1 < 1$. By the independence of each individual and the Markov property, we can calculate ρ as follows:

接下来考虑一般情况，
观察到灭绝概率有如左等式；
因此讨论并解该等式：

$$\begin{aligned}
\rho &= \sum_{k=0}^{\infty} \mathbf{Pr}\left[\text{extinction} \wedge G_1 = k\right] \\
&= \sum_{k=0}^{\infty} \mathbf{Pr}\left[\text{extinction} | G_1 = k\right] p_k \\
&= \sum_{k=0}^{\infty} \rho^k p_k.
\end{aligned} \tag{8}$$

将每个代等同于独立的
的第一代看待.

Let $\psi(z) \triangleq \sum_{k=0}^{\infty} p_k z^k$. Then Equation (8) yields that ρ is a fixed point of ψ, i.e., $\psi(\rho) = \rho$. By direct calculation we know ψ is an increasing and convex function on $[0, 1]$ with $\psi(0) = p_0$ and $\psi(1) = 1$. Then there can be two types of ψ depending on whether $\psi'(1)$ is larger than 1 as the following figure shows.



When $\psi'(1) = \sum_{k=1}^{\infty} k p_k = \mathbf{E}\left[X_{t-i}\right] \leq 1$, $z = 1$ is the only fixed point of ψ which corresponds to the Type 1 in the figure. That is to say, when $\mathbf{E}\left[X_{t-i}\right] \leq 1$, we have $\rho = 1$.

When $\mathbf{E}\left[X_{t-i}\right] > 1$ (Type 2), although there are two fixed points of ψ: r and 1, we claim that $\rho = r$ rather than 1 by showing that $\rho \leq r$. Let $q_t \triangleq \mathbf{Pr}\left[G_t = 0\right]$. Then $q_t \leq q_{t+1} < 1$ since $G_t = 0$ can always yields $G_{t+1} = 0$. We induct on t to show that $q_t \leq r$:

- When $t = 0$, it is obvious that $q_0 = 0 < r$.
- Assume that $q_t \leq r$. Since $q_{t+1} = \sum_{k=0}^{\infty} p_k q_t^k = \psi(q_t)$ and ψ is an increasing function, $q_{t+1} = \psi(q_t) \leq \psi(r) = r$.

We know that $\rho = \lim_{t\to\infty} q_t$ and $q_t \leq r$ for all $t \geq 0$. Thus $\rho \leq r$. However, we have shown that ρ is a fixed point of ψ. So $\rho = r$ when $\mathbf{E}[X_{t-i}] > 1$. In conclusion, $\rho = 1$ iff $\mathbf{E}[X_{t-i}] \leq 1$. QED

2.2 2-SAT

SAT is the problem of determining whether a CNF formula has satisfying assignments. k-SAT is the special cases of SAT that the clauses of the CNF formula consist of exact k literals. For example,

$$\phi = (x \vee y) \wedge (y \vee \bar{z}) \wedge (\bar{x} \vee z)$$

is a 2-CNF formula and $x = y = z = 1$ is one of its satisfying assignments. SAT is **NP**-complete and we have k-SAT $\in$ **NP** for $k \geq 3$. One can use an algorithm for finding strongly connected components to solve 2- SAT problem in linear time. Nevertheless, we introduce a simple randomized algorithm that can also solve this problem in polynomial-time with high probability.

We will extend the algorithm to solving 3-SAT in the homework!

我们提出了一种基于 RW 的算法，使最后 RW 出的结果高概率为正确解

Let ϕ be a 2-CNF formula and $V = \{v_1, v_2, ..., v_n\}$ be its set of variables The algorithm runs as follows:

⇐ 算法的步骤如左.

我们的目标是证明这一看来奇怪的算法的高概率正确性.

- Pick an arbitrary assignment $\sigma_0 : V \longrightarrow \{\text{true,false}\}$.
- For $t = 0, 1, 2, \ldots, 100n^2$:

  If σ_t satisfies ϕ, output σ_t;

  Else, pick an arbitrary unsatisfying clause, say $c = x \vee y$. Choose from $\{x, y\}$ uniformly at random and flip the assignment of the chosen literal. Let σ_{t+1} be the flipped assignment.
- Output "ϕ is not satisfiable".

Claim 3 *This algorithm outputs the correct answer with probability at least* $1 - \frac{1}{100}$.

Proof. It is clear that if a 2-SAT instance has no solution then our algorithm will always give the correct answer. So we consider the probability that our algorithm outputs no solution conditioned on that the instance indeed has a satisfying assignment.

正确性验证的基本思路在于：本质上该算法将一个 clause 改对的概率是 ≥50% 的，记一个 clause 都不对为 0，对了为 1，则 $\{X_t\}$ 类似一个 1-d RW ($p \geq \frac{1}{2}$).

Our algorithm produces $100n^2 + 1$ assignments $\sigma_0, \sigma_1, \ldots, \sigma_{100n^2}$. We claim that with probability at least $1 - \frac{1}{100}$, some of σ_k for $k \in \{0, \ldots, 100n^2 + 1\}$ is a satisfying assignment. The argument here, at first glance, is a bit weird. We fix an *arbitrary* $\sigma : V \longrightarrow \{\texttt{true}, \texttt{false}\}$ satisfying assignment. We in fact prove the following: For large enough k, conditioned on the event that none of $\sigma_0, \sigma_1, \ldots, \sigma_k$ is a satisfying assignment, $\sigma_{k+1} = \sigma$ holds with high probability.

接下来我们就要证，1-d $p = \frac{1}{2}$ RW 中大概率到 n 数量级范围的点需花 n^2 数量级步①，有这个界限后再证明 $p \geq 50\%$ 1-d RW 比 $p = 50\%$ 时的高概率，概率更高②，从而得证在 $O(n^2)$ 时，就几乎解对 n 个的 2-SAT.

Let $\{X_t\}_{t=0}^{100n^2}$ be a random variable sequence that

$$X_t \triangleq |\{v \in V : \sigma_t(v) = \sigma(v)\}|.$$

Note that $\{X_t\}_{t=0}^{100n^2}$ is not a Markov chain since it only contains partial information of σ_t and we cannot determine the distribution of X_{t+1} given X_t.

First we verify that $\mathbf{Pr}[X_{t+1} = X_t + 1 \mid \sigma_t] \geq \frac{1}{2}$ [1] and $\mathbf{Pr}[X_{t+1} = X_t - 1 \mid \sigma_t] \leq \frac{1}{2}$. WLOG assume we chose the clause $c = x \vee y$ in round t. Since c is not satisfied by σ_t, we have $\sigma_t(x) = \sigma_t(y) = \mathtt{false}$. Similarly, $x \vee y$ is satisfying under σ, so there are three possible assignments of $\sigma(x)$ and $\sigma(y)$:

- If $\sigma(x) = \mathtt{true}$ and $\sigma(y) = \mathtt{false}$, $\mathbf{Pr}[X_{t+1} = X_t + 1 \mid \sigma_t] = \mathbf{Pr}[\text{flip } x] = \frac{1}{2}$ and $\mathbf{Pr}[X_{t+1} = X_t - 1 \mid \sigma_t] = \mathbf{Pr}[\text{flip } y] = \frac{1}{2}$.
- If $\sigma(x) = \mathtt{false}$ and $\sigma(y) = \mathtt{true}$, we have $\mathbf{Pr}[X_{t+1} = X_t + 1 \mid \sigma_t] = \mathbf{Pr}[X_{t+1} = X_t - 1 \mid \sigma_t] = \frac{1}{2}$ similarly.
- If $\sigma(x) = \mathtt{true}$ and $\sigma(y) = \mathtt{true}$, $\mathbf{Pr}[X_{t+1} = X_t + 1 \mid \sigma_t] = \mathbf{Pr}[\text{flip } x \text{ or } y] = 1$.

这里证为什么该建模类似 $p \geq 50\%$ 的 1-d RW.

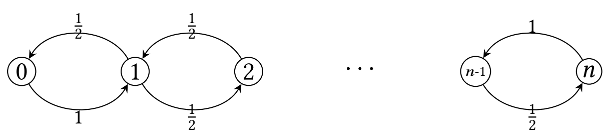
Thus we have $\mathbf{Pr}[X_{t+1} = X_t + 1 \mid \sigma_t] \geq \frac{1}{2}$ on condition that none of $\sigma_0, \sigma_1, \ldots, \sigma_t$ is a satisfying assignment.

[1] Let Y be a random variable. Then function $\mathbf{Pr}[\cdot \mid Y] : \mathrm{Ran}(Y) \to \mathbb{R}$ is defined by $\mathbf{Pr}[\cdot \mid Y] = \mathbf{E}[\mathbb{1}[\cdot] \mid Y]$. Note that $\mathbf{Pr}[\cdot \mid Y]$ is a random variable. Here we slightly abuse the notations and denote the event "$\forall a \in \mathrm{Ran}(Y), \mathbf{Pr}[\cdot \mid Y = a] \geq \frac{1}{2}$" as $\mathbf{Pr}[\cdot \mid Y] \geq \frac{1}{2}$.

Consider the 1-D random walk $\{Y_t\}_{t \geq 0}$ on $[n] \cup \{0\}$ that $Y_0 = X_0$ and for $Y_t \notin \{0, 1\}$

$$Y_{t+1} = \begin{cases} Y_t + 1, & \text{w.p. } \frac{1}{2} \\ Y_t - 1, & \text{w.p. } \frac{1}{2} \end{cases}.$$

If $Y_t = 0$, $Y_{t+1} = Y_t + 1$ w.p. 1 and if $Y_t = n$, then $Y_{t+1} = Y_t - 1$ w.p. 1.



这里使用coupling证明了(*)式，
coupling再认识：为证明两个独立分布的关系，不改变二者特性地构造一种耦合，由耦合的变化进而发现二者的关系.
这里说，我将 X_t、Y_t 的分布耦合起来，发现此时 X_t、Y_t 总有 $Y_t \geq X_t$，因为耦合没有动原来分布的特性，故 $Y_t \geq X_t$ 有普遍性，故 $\Pr[X_t = n] \geq \Pr[Y_t = n]$

Then we have[2]

(*)

$$\begin{aligned} \mathbf{Pr}[\text{the algorithm is correct}] &\geq \mathbf{Pr}\left[\exists t \in [0, 100n^2] \; s.t. X_t = n\right] \\ &\geq \mathbf{Pr}\left[\exists t \in [0, 100n^2] \; s.t. Y_t = n\right]. \end{aligned} \tag{9}$$

[2] The second inequality can be verified by constructing a coupling which satisfies $Y_t \geq X_t$ for all $t \geq 0$. The existence of such coupling is guaranteed by $\mathbf{Pr}[X_{t+1} = X_t + 1 \mid \sigma_t] \geq \mathbf{Pr}[Y_{t+1} = Y_t + 1]$. Specifically, if there is one false and one true in $\{\sigma(x), \sigma(y)\}$, then Y_{t+1} moves the same as X_{t+1}. If $\sigma(x) = \sigma(y) =$true, then Y_{t+1} moves $+1$ or -1 uniformly at random.

Assume that $Y_0 = X_0 = i$. Let $T_{i \to n}$ be the first hitting time of n from i. Then

$$T_{i \to n} = \sum_{k=i}^{n-1} T_{k \to k+1}.$$

For $i > 0$, we have

Recall $\mathscr{A} = [\text{the first step is to the right}]$.

$$\begin{aligned} T_{i \to i+1} &= \mathbb{1}[\mathscr{A}] + \mathbb{1}[\bar{\mathscr{A}}](1 + T_{i-1 \to i+1}) \\ &= \mathbb{1}[\mathscr{A}] + \mathbb{1}[\bar{\mathscr{A}}](1 + T_{i-1 \to i} + T_{i \to i+1}) \end{aligned}$$

Taking the expectation of both sides, we have $\mathbf{E}[T_{i \to i+1}] = 2 + \mathbf{E}[T_{i-1 \to i}]$. Note that $T_{0 \to 1} = 1$, then

$$\mathbf{E}[T_{i \to n}] = \sum_{k=i}^{n-1} \mathbf{E}[T_{k \to k+1}] = \sum_{k=i}^{n-1} 2k + 1 = n^2 - i^2 \leq n^2.$$

这里使用经典的RW上的递推分析.
不同于之前例1在无穷下分析时 $E(T_{2\to1}) = E(T_{1\to0})$，这里有穷无上结论，但可以一路递推至 $T_{0\to1}$，而 $T_{0\to1} = E(T_{0\to1}) = 1$（我们的MC算法就是如此设计的），故可算出 $\forall i, n$, $E[T_{i\to n}] = n^2 - i^2$

Then we apply the Markov's inequality to give a lower bound for $\mathbf{Pr}\left[\exists t \in [0, 100n^2]\ s.t. Y_t = n\right]$:

$$1 - \mathbf{Pr}\left[\exists t \in [0, 100n^2]\ s.t. Y_t = n\right] = \mathbf{Pr}\left[T_{Y_0 \to n} > 100n^2\right] \leq \frac{\mathbf{E}\left[T_{Y_0 \to n}\right]}{100n^2} \leq \frac{1}{100}.$$

最后使用 Markov ineq：计算偏离期望100倍时，即步数为$100n^2$时，赋值clause对n个的概率 $\geq 1-\frac{1}{100}$ 得证！

By Equation (9), we know that $\mathbf{Pr}$ [the algorithm is correct] is lower bounded by $1 - \frac{1}{100}$. □

3 Fundamental Theorem

该部分是对MC的最终总结和推广。

In this section, we develop the fundamental theorem of Markov chains for chains with possibly infinite states. First we introduce some abbreviations to simplify the expression:

- Aperiodicity:[A],
- Irreducibility:[I],
- Recurrence:[R],
- Positive Recurrence: [PR],
- Has a stationary distribution:[S],
- Has a unique stationary distribution:[U],
- Convergence:[C],
- Finiteness:[F].

之前Finite下的公式升级为：

The finite FTMC can be written as: [F]+[A]+[I]⇒[S]+[U]+[C]. For infinite Markov chains, the theorem need to be modified as: [PR]+[A]+[I]⇒[S]+[U]+[C].

每个条件对应一个结论。

Before the proof of the theorem, we need to prepare some mathematical tools.

3.1 Laws of Large Numbers

这是证明，主要思想是证给定条件下稳态分布中 $\pi_i = \frac{1}{E_i[T_i]}$

从i出发回到i的步数的期望

运用的工具为强大数定律

$\Pr[\lim_{n\to\infty} X_n \to \mu] = 1$

(弱)：

$\lim_{n\to\infty} [\Pr[X_n \to \mu]] = 1$

证明略，因为太长太长了。

$X_1, X_2, \ldots$ is an infinite sequence of independent and identically distributed Lebesgue integrable random variables with expected value $\mathbf{E}[X_1] = \mathbf{E}[X_2] = \cdots = \mu < \infty$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample average. Then we have the following two laws of large numbers.

Theorem 4 (Weak law of large numbers or Khinchin's law) *The sample average converge in probability towards the expected value:*

$$\bar{X}_n \xrightarrow{p} \mu \quad \text{when } n \to \infty.$$

That is, for any positive value ε,

$$\lim_{n\to\infty} \mathbf{Pr}\left[\left|\bar{X}_n - \mu\right| < \epsilon\right] = 1.$$

Theorem 5 (Strong law of large numbers or Kolmogorov's law) *The sample average converges almost surely or with probability* 1 *to the expected value:*

$$\bar{X}_n \xrightarrow{\text{a.s.}} \mu \quad \text{when } n \to \infty.$$

That is,

$$\mathbf{Pr}\left[\lim_{n\to\infty} \bar{X}_n \to \mu\right] = 1.$$

Let $(\Omega, \mathcal{F}, P)$ be the probability space. Here $\bar{X}_n \to \mu$ means $\exists M \in \mathcal{F}$ satisifying

- P(M)=1;
- $\forall \omega \in M, \bar{X}_n(\omega) \xrightarrow{n\to\infty} \mu$.

As the name of the laws shows, *convergence in probability* is weaker than *convergence with probability* 1. Consider a sequence of independent random variables $X_1, X_2, \dots$ that X_n is 1 with probability $\frac{1}{n}$ and X_n is 0 with probability $1 - \frac{1}{n}$. Then the sequence converges to 0 in probability but not with probability 1 since we cannot find an $M \in \mathcal{F}$ with measure 1 such that $\bar{X}_n(\omega) \xrightarrow{n\to\infty} 0$ for every $\omega \in M$.

Theorem 6 (Strong law of large numbers for Markov chains) *If there is a finite path from state i to j, then*

$$\mathbf{P}_i\left[\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}[X_t = j] = \frac{1}{\mathbf{E}_j[T_j]}\right] = 1.$$

Proof. If j is transient, then the random process will visit j for finite times with probability 1. Thus $\mathbf{P}_i\left[\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}[X_t = j] = \frac{1}{\mathbf{E}_j[T_j]}\right] = \mathbf{P}_i\left[\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}[X_t = j] = 0\right] = 1$.

If j is recurrent, we first prove the theorem for $i = j$. We call a loop from j to j a cycle (we visit j only at the beginning and end of the loop). Denote C_r as the length of the r^{th} cycle during the process. Let $S_k = \sum_{r=1}^{k} C_r$. Let k_n be the number of cycles before the $n+1$ step, that is, $k_n = \max\{k | S_k \le n\}$. Then we have $S_{k_n} \le n < S_{k_n+1}$ and consequently $\frac{S_{k_n}}{k_n} \le \frac{n}{k_n} < \frac{S_{k_n+1}}{k_n}$. Note that with probability 1, $k_n \to \infty$ when $n \to \infty$. We have with probability 1 that

$$\lim_{k\to\infty} \frac{S_k}{k} \le \lim_{n\to\infty} \frac{n}{k_n} < \lim_{k\to\infty} \frac{S_{k+1}}{k}.$$

Note that $S_k = \sum_{r=1}^{k} C_r$ where each C_r is an i.i.d random variable with mean $\mathbf{E}_j\left[T_j\right]$. So by SLLN (Theorem 5), we have $\lim_{k\to\infty} \frac{S_k}{k} = \mathbf{E}_j\left[T_j\right]$ and $\lim_{k\to\infty} \frac{S_{k+1}}{k} = \lim_{k\to\infty} \frac{S_{k+1}}{k+1} \cdot \frac{k+1}{k} = \mathbf{E}_j\left[T_j\right]$. As a result, with probability 1,

$$\mathbf{E}_j\left[T_j\right] = \lim_{n\to\infty} \frac{n}{k_n} = \lim_{n\to\infty} \frac{n}{\sum_{t=1}^{n} \mathbb{1}[X_t = j]}.$$

If j is recurrent and $i \ne j$, let $T_{i\to j}$ be the first time visiting j. Then we have $\frac{S_{k_n} + T_{i\to j}}{k_n} \le \frac{n}{k_n} < \frac{S_{k_n+1} + T_{i\to j}}{k_n}$. Since $\mathbf{P}_i[T_j < \infty] = 1$, $\mathbf{P}_i[\lim_{k\to\infty} \frac{T_{i\to j}}{k} = 0] = 1$. The remaining proof is the same with the situation that $i = j$. $\square$

Corollary 7 *Let P be the transition function of an irreducible Markov chain where $P^t(i,j) = \mathbf{Pr}[X_t = j | X_0 = i]$. Then for any states i, j,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} P^t(i,j) = \frac{1}{\mathbf{E}_j[T_j]}.$$

Proof. By the strong law of large numbers for Markov chains, there exists a set $M \in \mathcal{F}$ such that $P(M) = 1$ and $\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}[X_t(\omega) = j] = \frac{1}{\mathbf{E}_j[T_j]}$ for any $\omega \in M$. Then,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} P^t(i,j) &= \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbf{E}_i\left[\mathbb{1}[X_t = j]\right] \\
&= \lim_{n\to\infty} \mathbf{E}_i\left[\frac{1}{n}\sum_{t=1}^{n} \mathbb{1}[X_t = j]\right] \\
&= \mathbf{E}_i\left[\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}[X_t = j]\right] \\
&= \frac{1}{\mathbf{E}_j[T_j]},
\end{aligned}$$

where the third equation follows from the bounded convergence theorem. □

Bounded Convergence Theorem: If $X_n \xrightarrow{a.s.} X$ and $\mathbf{E}[X] < \infty$, then $\mathbf{E}[X_n] \to \mathbf{E}[X]$.

3.2 Proof of the Fundamental Theorem

We will first prove the existence and uniqueness of the stationary distribution in this lecture.(i.e. [S] and [U])

Theorem 8 *[I]+[PR]⇒[S]+[U].*

Proof. [Proof of [U]] Let $\mathcal{S}$ be the set of states. Assume π is a stationary distribution of the Markov chain, i.e.,

$$\forall j \in \mathcal{S},\ \forall t \geq 0,\ \sum_{i\in\mathcal{S}} \pi(i)P^t(i,j) = \pi(j).$$

This yields that for $n \geq 1$,

$$\frac{1}{n}\sum_{i\in\mathcal{S}} \pi(i)\sum_{t=1}^{n} P^t(i,j) = \pi(j).$$

Taking $n \to \infty$ and applying the dominated convergence theorem, we have

$$\pi(j) = \sum_{i\in\mathcal{S}} \pi(i) \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} P^t(i,j) = \sum_{i\in\mathcal{S}} \pi(i)\cdot\frac{1}{\mathbf{E}_j[T_j]} = \frac{1}{\mathbf{E}_j[T_j]}.$$

□

Dominated Convergence Theorem: If $\int_S |f_n| < \infty$, then $\lim_{n\to\infty}\int_S f_n = \int_S \lim_{n\to\infty} f_n$.

Proof. [Proof of [S]] Then we prove the above π is a stationary distribution.

$\mathcal{S}$ is finite. We first assume $\mathcal{S}$ is finite, so that we can safely exchange the order of taking limitation and summation in the calculations below.

$$\begin{aligned}
\sum_{j\in\mathcal{S}} \pi(j) &= \sum_{j\in\mathcal{S}} \frac{1}{\mathbf{E}_j[T_j]} = \sum_{j\in\mathcal{S}} \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} P^t(i,j) \\
&= \lim_{n\to\infty} \sum_{j\in\mathcal{S}} \frac{1}{n}\sum_{t=1}^{n} P^t(i,j) = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n}\sum_{j\in\mathcal{S}} P^t(i,j) = 1.
\end{aligned}$$

This indicates that π is a legal distribution. We then verify that π is indeed the stationary distribution.

Note that $P^{t+1}(i,j) = \sum_{k\in\mathcal{S}} P^t(i,k)P(k,j)$. Then

$$\begin{aligned}
\frac{1}{\mathbf{E}_j[T_j]} &= \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}P^t(i,j) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}P^{t+1}(i,j)\\
&= \lim_{n\to\infty}\sum_{k\in\mathcal{S}}P(k,j)\frac{1}{n}\sum_{t=1}^{n}P^t(i,k) = \sum_{k\in\mathcal{S}}P(k,j)\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}P^t(i,k)\\
&= \sum_{k\in\mathcal{S}}P(k,j)\cdot\frac{1}{\mathbf{E}_k[T_k]}.
\end{aligned}$$

That is,

$$\pi(j) = \sum_{k\in\mathcal{S}}P(k,j)\pi(k).$$

It is worth noting that [PR] is equivalent to [I] when $\mathcal{S}$ is finite.

$\mathcal{S}$ is infinite. When $\mathcal{S}$ is (countably) infinite, we consider every finite subset A of $\mathcal{S}$. Then

$$\begin{aligned}
\sum_{j\in A}\pi(j) &= \sum_{j\in A}\frac{1}{\mathbf{E}_j[T_j]} = \sum_{j\in A}\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}P^t(i,j)\\
&= \lim_{n\to\infty}\sum_{j\in A}\frac{1}{n}\sum_{t=1}^{n}P^t(i,j) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sum_{j\in A}P^t(i,j) < 1.
\end{aligned}$$

Therefore

$$\sum_{j\in\mathcal{S}}\pi(j) = \sup_{\text{finite } A\subseteq\mathcal{S}}\sum_{j\in A}\pi(j) =: C \le 1.$$

Since $[PR]$, we know that $C \neq 0$. In the following, we will prove that π/C is a stationary distribution. Then $C = 1$ follows from the uniqueness of the stationary distribution we just proved.

For every finite $A \subseteq \mathcal{S}$, we have

$$\sum_{k\in A}P(k,j)\cdot\frac{1}{\mathbf{E}_k[T_k]} \le \frac{1}{\mathbf{E}_j[T_j]}.$$

Therefore,

$$\sum_{k\in\mathcal{S}}P(k,j)\cdot\frac{1}{\mathbf{E}_k[T_k]} = \sup_{\text{finite } A\subseteq\mathcal{S}}\sum_{k\in A}P(k,j)\cdot\frac{1}{\mathbf{E}_k[T_k]} \le \frac{1}{\mathbf{E}_j[T_j]}.$$

We show that indeed the equality holds. Assume for a contradiction that

$$\sum_{k\in\mathcal{S}}P(k,j)\cdot\frac{1}{\mathbf{E}_k[T_k]} < \frac{1}{\mathbf{E}_j[T_j]}.$$

Summing the both sides over all $j \in \mathcal{S}$, we obtain

$$\sum_{k\in\mathcal{S}}\frac{1}{\mathbf{E}_k[T_k]} < \sum_{j\in\mathcal{S}}\frac{1}{\mathbf{E}_j[T_j]},$$

which is a contradiction. As a result, we know

$$\sum_{k \in \mathcal{S}} P(k, j) \cdot \frac{1}{\mathbf{E}_k\left[T_k\right]} = \frac{1}{\mathbf{E}_j\left[T_j\right]},$$

and $\hat{\pi}(j) = \frac{1}{C \cdot \mathbf{E}_j\left[T_j\right]}$ is a stationary distribution. By the uniqueness of the distribution, we have $C = 1$.

$\square$