

Stochastic Process

Daniel Mao

Contents

1	Stochastic Process	1
1.1	Definitions	1
1.2	Accessibility and Communication	2
1.3	Periodicity	3
1.4	Transience and Recurrence	4
1.5	Random Walk	8
1.6	Limiting Behaviors	9
1.7	Limiting Behaviors	11
2	Convergence of Random Variables	13
2.1	Definitions	13
2.2	Markov's Inequality	14
2.3	Properties	14
2.4	Law of Large Numbers	15
3	Markov Decision Process	17
4	Poisson Process	19
4.1	Homogeneous Poisson Process	19

Chapter 1

Stochastic Process

1.1 Definitions

Definition (Stochastic Process). Let \mathcal{T} be an index set. Let $X(t)$ be a random variable. We define a **stochastic process** to be the net $(X(t))_{t \in \mathcal{T}}$.

Definition (Discrete-Time Stochastic Process). Let $(X(t))_{t \in \mathcal{T}}$ be a stochastic process. We say that it is a **discrete-time stochastic process** if the index set \mathcal{T} is countable.

Definition (Markov Property). Let \mathcal{S} be a state space. Let $(X_n)_{n \in \mathbb{N}}$ be a discrete-time stochastic process. We say that it has the **Markov property** if

$$\forall n \in \mathbb{N}, \forall x_0 \dots x_{n+1} \in \mathcal{S}, \quad \Pr(X_{n+1} = x_{n+1} \mid (X_n)_{n=0}^n = (x_n)_{n=0}^n) = \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

This property states that the conditional distribution of any future state X_{n+1} given the past states X_0, \dots, X_{n-1} and the present state X_n is independent of the past states.

i.e., if we know the value taken by the process at a given time, we will not get any additional information about the future behavior of the process by gathering more knowledge about the past.

Definition (Markov Chain). We define a **Markov chain** to be a discrete-time stochastic process with the Markov property.

Proposition 1.1.1.

$$\forall n \in \mathbb{N}, \forall j \in \{0 \dots n-1\}, \forall x_0 \dots x_{n+1} \in \mathcal{S}, \quad \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n, (X_i)_{i=1}^{j-1} = (x_i)_{i=1}^{j-1}, (X_i)_{i=j+1}^{n-1} = (x_i)_{i=j+1}^{n-1}) = \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

Proof.

$$\Pr(X_{n+1} = x_{n+1} \mid X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}) \tag{1.1}$$

$$= \frac{\Pr(X_{n+1} = x_{n+1}, X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})}{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})} \quad (1.2)$$

$$= \frac{\sum_{x_j=0}^{\infty} \Pr(X_{n+1} = x_{n+1}, X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}, X_j = x_j)}{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})} \quad (1.3)$$

$$= \frac{\sum_{x_j=0}^{\infty} \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}, X_j = x_j) \Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}, X_j = x_j)}{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})} \quad (1.4)$$

$$= \frac{\sum_{x_j=0}^{\infty} \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n) \Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}, X_j = x_j)}{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})} \quad (1.5)$$

$$= \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n) \frac{\sum_{x_j=0}^{\infty} \Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}, X_j = x_j)}{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})} \quad (1.6)$$

$$= \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n) \frac{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})}{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})} \quad (1.7)$$

$$= \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n). \quad (1.8)$$

That is,

$$\Pr(X_{n+1} = x_{n+1} \mid X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}) = \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

■

Definition (Transition Probability). *Let i and j be a pair of states. Let n be some time step. We define the **transition probability** from state i at time n to state j at time $n+1$, denoted by $P_{n,i,j}$, to be the conditional probability given by*

$$P_{n,i,j} = \Pr(X_{n+1} = j \mid X_n = i).$$

Definition (Stationary / Homogeneous). *We say that a discrete-time Markov chain is **stationary** or **homogeneous** if $\forall i, j \in \mathcal{S}, \forall n \in \mathbb{N}, P_{n,i,j} = P_{i,j}$ for some $P_{i,j}$.*

Theorem 1 (Chapman-Kolmogorov Equations).

$$P^{(n)} = P^{(m)} P^{(n-m)}.$$

1.2 Accessibility and Communication

Definition (Accessible). *Let i and j be two states. We say that state j is **accessible** from state i if $\exists n \in \mathbb{N}$ such that $P_{i,j}^{(n)} > 0$.*

Definition (Communicate). *Let i and j be two states. We say that state i and state j **communicate** if i and j are accessible from each other.*

Proposition 1.2.1. *The communication relation is an equivalence relation. i.e., it is reflexive, symmetric, and transitive.*

Proof. **Transitivity:**

Let i, j, k be states. Assume that $i \leftrightarrow j$ and $j \leftrightarrow k$. We are to prove that $i \leftrightarrow k$. Since $i \leftrightarrow j$, $\exists n \in \mathbb{N}$ such that $P_{i,j}^{(n)} > 0$. Since $j \leftrightarrow k$, $\exists m \in \mathbb{N}$ such that $P_{j,k}^{(m)} > 0$. By the Chapman-Kolmogorov equation, we get

$$P_{i,k}^{(n+m)} = \sum_{l=0}^{\infty} P_{i,l}^{(n)} P_{l,k}^{(m)} \geq P_{i,j}^{(n)} P_{j,k}^{(m)} > 0.$$

That is, $P_{i,k}^{(n+m)} > 0$. So $i \rightarrow k$. Similarly, we can show that $k \rightarrow i$. So $i \leftrightarrow k$. ■

Proposition 1.2.2. *Let i and j be two states. If state j is not accessible from state i , then*

$$\Pr(\text{DTMC ever exists state } j \mid X_0 = i) = 0.$$

Proof. Since state j is not accessible from state i , we have $\forall n \in \mathbb{N}$, $P_{i,j}^{(n)} = 0$.

$$\begin{aligned} & \Pr(\text{DTMC ever exists state } j \mid X_0 = i) \\ &= \Pr\left(\bigcup_{n=0}^{\infty} \{X_n = j\} \mid X_0 = i\right) \leq \sum_{n=0}^{\infty} \Pr(X_n = j \mid X_0 = i) \\ &= \sum_{n=0}^{\infty} P_{i,j}^{(n)} = 0. \end{aligned}$$

That is,

$$\Pr(\text{DTMC ever exists state } j \mid X_0 = i) = 0.$$
■

Definition (Communication Class). *We define a **communication class** to the set of states that communicate with each other.*

Definition (Irreducible, Reducible). *We say that a discrete-time Markov chain is **irreducible** if it has only one communication class; and we say that it is **reducible** otherwise.*

1.3 Periodicity

Definition (Period). *Let i be a state. We define the **period** of i , denoted by $d(i)$, to be the number given by*

$$d(i) := \gcd\{n \in \mathbb{Z}_+ : P_{i,i}^{(n)} > 0\}.$$

Definition (Aperiodic). *We say that a state i is **aperiodic** if $d(i) = 1$. We say that a discrete-time Markov chain is **aperiodic** if $d(i) = 1$ for all state i .*

Proposition 1.3.1. *Let i and j be two states. If $i \leftrightarrow j$, then $d(i) = d(j)$.*

Proof. Since $i \leftrightarrow j$, $\exists n \in \mathbb{Z}_+$ such that $P_{i,j}^{(n)} > 0$; $\exists m \in \mathbb{Z}_+$ such that $P_{j,i}^{(m)} > 0$; and $\exists s \in \mathbb{Z}_+$ such that $P_{j,j}^{(s)} > 0$. Note that

$$P_{i,i}^{(n+m)} \geq P_{i,j}^{(n)} P_{j,i}^{(m)} > 0.$$

and

$$P_{i,i}^{(n+s+m)} \geq P_{i,j}^{(n)} P_{j,j}^{(s)} P_{j,i}^{(m)} > 0.$$

So $d(i) \mid (n+m)$ and $d(i) \mid (n+s+m)$. So $d(i) \mid ((n+s+m) - (n+m)) = s$. Since $\forall s \in \mathbb{Z}_+ : P_{j,j}^{(s)} > 0$, $d(i) \mid s$, we get $d(i) \mid d(j)$. Similarly, we have $d(j) \mid d(i)$. So $d(i) = d(j)$. ■

1.4 Transience and Recurrence

1.4.1 Preliminaries

Notation (First Visit Probability). Let i and j be two states. Let $n \in \mathbb{Z}_+$ be a time step. We define the **first visit probability** to state j , starting from state i , occurs at time step n , denoted by $f_{i,j}^{(n)}$, to be the probability given by

$$f_{i,j}^{(n)} = \Pr(X_n = j, X_{n-1} \dots X_1 \neq j \mid X_0 = i).$$

Proposition 1.4.1.

$$P_{i,j}^{(n)} = \sum_{k=1}^n P_{j,j}^{(n-k)} f_{i,j}^{(k)}.$$

Proof.

$$\begin{aligned} P_{i,j}^{(n)} &= \Pr(X_n = j \mid X_0 = i) \\ &= \sum_{k=1}^n \Pr(X_n = j, \text{ first visit occurs at time } k \mid X_0 = i) \\ &= \sum_{k=1}^n \Pr(X_n = j, X_k = j, X_{k-1} \dots X_1 \neq j \mid X_0 = i) \\ &= \sum_{k=1}^n \Pr(X_n = j \mid X_k = j, X_{k-1} \dots X_1 \neq j, X_0 = i) \Pr(X_k = j, X_{k-1} \dots X_1 \neq j \mid X_0 = i) \\ &= \sum_{k=1}^n \Pr(X_n = j \mid X_k = j) \Pr(X_k = j, X_{k-1} \dots X_1 \neq j \mid X_0 = i), \text{ by the Markov property} \\ &= \sum_{k=1}^n P_{j,j}^{(n-k)} f_{i,j}^{(k)}. \end{aligned}$$

That is,

$$P_{i,j}^{(n)} = \sum_{k=1}^n P_{j,j}^{(n-k)} f_{i,j}^{(k)}.$$

■

Proposition 1.4.2.

$$f_{i,j}^{(n)} = P_{i,j}^{(n)} - \sum_{k=1}^{n-1} f_{i,j}^{(k)} P_{j,j}^{(n-k)}.$$

Proof.

$$\begin{aligned} f_{i,j}^{(n)} &= f_{i,j}^{(n)} \cdot 1 = f_{i,j}^{(n)} P_{j,j}^{(0)} = f_{i,j}^{(n)} P_{j,j}^{(n-n)} \\ &= \sum_{k=1}^n f_{i,j}^{(k)} P_{j,j}^{(n-k)} - \sum_{k=1}^{n-1} f_{i,j}^{(k)} P_{j,j}^{(n-k)} \\ &= P_{i,j}^{(n)} - \sum_{k=1}^{n-1} f_{i,j}^{(k)} P_{j,j}^{(n-k)}. \end{aligned}$$

That is,

$$f_{i,j}^{(n)} = P_{i,j}^{(n)} - \sum_{k=1}^{n-1} f_{i,j}^{(k)} P_{j,j}^{(n-k)}.$$

■

The above proposition gives a recursive means to compute $f_{i,j}^{(n)}$ for $n \geq 2$.

Notation ($f_{i,j}$). Let i and j be two states.

$$f_{i,j} := \Pr(\text{DTMC ever makes a future visit to state } j \mid X_0 = i) = \sum_{k=1}^{\infty} f_{i,j}^{(k)}.$$

Note that $f_{i,j}$ is an infinite sum of probabilities and that $f_{i,j}$ itself is defined to be a probability. So the infinite sum of probabilities is ≤ 1 .

1.4.2 Definitions

Definition (Transient and Recurrent - 1). Let i be a state. We say that state i is **transient** if $f_{i,i} < 1$; and we say that state i is **transient** if $f_{i,i} = 1$.

Definition (Transient and Recurrent - 2). Let i be a state. We say that state i is **transient** if $\mathbb{E}[M_i \mid X_0 = i] < \infty$; and we say that state i is **transient** if $\mathbb{E}[M_i \mid X_0 = i] = \infty$.

Definition (Transient and Recurrent - 3). Let i be a state. We say that state i is **transient** if $\sum_{n=1}^{\infty} P_{i,i}^{(n)} < \infty$; and we say that state i is **transient** if $\sum_{n=1}^{\infty} P_{i,i}^{(n)} = \infty$.

Proposition 1.4.3. The three definitions of transient and recurrent are equivalent.

Proposition 1.4.4. Let i be a state. Let M_i be a random variable that denotes the number of future visits to state i . Then $\Pr(M_i = k \mid X_0 = i) = f_{i,i}^k (1 - f_{i,i})$.

Proposition 1.4.5. $M_i \sim \text{GEO}_f(1 - f_{i,i})$ and hence $\mathbb{E}[M_i \mid X_0 = i] = \frac{f_{i,i}}{1 - f_{i,i}}$.

Proposition 1.4.6.

$$\mathbb{E}[M_i \mid X_0 = i] = \sum_{n=1}^{\infty} P_{i,i}^{(n)}.$$

Proof. Define a random variable A_n as

$$A_n := \begin{cases} 0, & \text{if } X_n \neq i \\ 1, & \text{if } X_n = i. \end{cases}$$

$$\begin{aligned} \mathbb{E}[M_i \mid X_0 = i] &= \mathbb{E}\left[\sum_{n=1}^{\infty} A_n \mid X_0 = i\right] = \sum_{n=1}^{\infty} \mathbb{E}[A_n \mid X_0 = i] \\ &= \sum_{n=1}^{\infty} \left[0 \cdot \Pr(A_n = 0 \mid X_0 = i) + 1 \cdot \Pr(A_n = 1 \mid X_0 = i)\right] \\ &= \sum_{n=1}^{\infty} \Pr(A_n = 1 \mid X_0 = i) = \sum_{n=1}^{\infty} \Pr(X_n = i \mid X_0 = i) \\ &= \sum_{n=1}^{\infty} P_{i,i}^{(n)}. \end{aligned}$$

That is,

$$\mathbb{E}[M_i \mid X_0 = i] = \sum_{n=1}^{\infty} P_{i,i}^{(n)}.$$

■

1.4.3 Properties

Proposition 1.4.7. Let i and j be two states. If $i \leftrightarrow j$, i is recurrent (transient) if and only if j is recurrent (transient).

Proof. It suffices to show that i is recurrent implies j is recurrent. Since $i \leftrightarrow j$, $\exists m \in \mathbb{Z}_+$ such that $P_{i,j}^{(m)} > 0$, and $\exists n \in \mathbb{Z}_+$ such that $P_{j,i}^{(n)} > 0$. Since i is recurrent, we have $\sum_{l=1}^{\infty} P_{i,i}^{(l)} = +\infty$. Let $s \in \mathbb{Z}_+$. Then $P_{j,j}^{(n+s+m)} \geq P_{j,i}^{(n)} P_{i,i}^{(s)} P_{i,j}^{(m)}$. Then

$$\begin{aligned} \sum_{k=1}^{\infty} P_{j,j}^{(k)} &\geq \sum_{k=n+s+m}^{\infty} P_{j,j}^{(k)} \\ &= \sum_{s=1}^{\infty} P_{j,j}^{(n+s+m)} \\ &\geq \sum_{s=1}^{\infty} P_{j,i}^{(n)} P_{i,i}^{(s)} P_{i,j}^{(m)} \end{aligned}$$

$$= \underbrace{P_{j,i}^{(n)}}_{>0} \underbrace{P_{i,j}^{(m)}}_{>0} \underbrace{\sum_{s=1}^{\infty} P_{i,i}^{(s)}}_{=+\infty} = +\infty.$$

That is, $\sum_{k=1}^{\infty} P_{j,j}^{(k)} = +\infty$. So j is recurrent. ■

Proposition 1.4.8. *If $i \leftrightarrow j$ and i is recurrent, then $f_{i,j} = 1$.*

Proof. Since $i \leftrightarrow j$ and i is recurrent, we know that j is also recurrent. Since j is recurrent, we have $f_{j,j} = 1$. Assume for the sake of contradiction that $f_{i,j} < 1$. Since $i \leftrightarrow j$, $\exists n \in \mathbb{Z}_+$ such that $P_{j,i}^{(n)} > 0$. Let n_i be the smallest of such. So if the DTMC reaches state i from state j in n_i time steps, there is no state j in between. Then

$$\begin{aligned} 1 - f_{j,j} &= \Pr(\text{DTMC never makes a future visit to state } j \mid X_0 = j) \\ &\geq \underbrace{P_{j,i}^{(n_i)}}_{>0} \underbrace{(1 - f_{i,j})}_{>0} > 0. \end{aligned}$$

That is, $f_{j,j} < 1$. This contradicts to the fact that $f_{j,j} = 1$. So $f_{i,j} = 1$. ■

Based on the above result, we know that starting from any state of a recurrent class, a DTMC will visit each state of the class infinitely many times.

Theorem 2. *A finite-state discrete-time Markov chain has at least one recurrent state.*

Proof. Let $\mathcal{S} = \{1..N\}$ be a state space where $N \in \mathbb{N}$. Assume for the sake of contradiction that all states are transient. Then after some finite amount of time T , all states will never be visited again. However, this is not possible. So there must be some state that is recurrent. ■

Corollary. *An irreducible, finite-state discrete-time Markov chain must be recurrent.*

Proposition 1.4.9. *If i is recurrent and $i \nleftrightarrow j$, then $\forall k \in \mathbb{Z}_+$, $P_{i,j}^{(k)} = 0$.*

Proof. Assume for the sake of contradiction that $\exists k \in \mathbb{Z}_+$ such that $P_{i,j}^{(k)} > 0$. i.e., $i \rightarrow j$. Let k_i be the smallest of such. Since $i \nleftrightarrow j$ and $i \rightarrow j$, $\forall n \in \mathbb{Z}_+$, $P_{j,i}^{(n)} = 0$. So there is a probability of at least $P_{i,j}^{(k_i)}$ that the DTMC starts from state i and never returns to state i . This contradicts to the assumption that i is recurrent. So $\forall k \in \mathbb{Z}_+$, we have $P_{i,j}^{(k)} = 0$. ■

Once a process enters a recurrent class of states, it can never leave that class. For this reason, a recurrent class is often referred to as a **closed class**.

Corollary. *If $P_{i,j}^{(k)} > 0$ for some k and $i \nleftrightarrow j$, then i is transient.*

Proof. Notice this statement is equivalent to the last via some simple logical operations. ■

1.5 Random Walk

Definition (Random Walk). Let $\mathcal{S} = \mathbb{Z}$ be a state space. Let $(X_n)_{n \in \mathbb{N}}$ be a discrete-time Markov chain with state space \mathcal{S} . We say that $(X_n)_{n \in \mathbb{N}}$ is a **random walk** if it satisfies the property that

$$\forall i \in \mathbb{Z}, \quad P_{i,i+1} = p \text{ and } P_{i,i-1} = 1 - p, \text{ for some } p \in (0, 1).$$

Proposition 1.5.1. *Random walks are irreducible.*

Proposition 1.5.2. *The period of all states of a random walk is 2. i.e., $\forall i \in \mathbb{Z}, d(i) = 2$.*

Proof. Note that $P_{i,i}^{(2n-1)} = 0$ for all $n \in \mathbb{Z}_+$ and $P_{i,i}^{(2n)} > 0$ for all $n \in \mathbb{Z}_+$. ■

Proposition 1.5.3. *If $p = 0.5$, then the DTMC is recurrent. If $p \neq 0.5$, then the DTMC is transient.*

Proof.

$$\begin{aligned} f_{0,0} &= \Pr(\text{DTMC ever makes a future visit to state 0} \mid X_0 = 0) \\ &= \Pr(X_1 = 1 \mid X_0 = 0) \Pr(\text{DTMC ever makes a future visit to state 0} \mid X_1 = 1, X_0 = 0) \\ &\quad + \Pr(X_1 = -1 \mid X_0 = 0) \Pr(\text{DTMC ever makes a future visit to state 0} \mid X_1 = -1, X_0 = 0) \\ &= \Pr(X_1 = 1 \mid X_0 = 0) \Pr(\text{DTMC ever makes a future visit to state 0} \mid X_1 = 1) \\ &\quad + \Pr(X_1 = -1 \mid X_0 = 0) \Pr(\text{DTMC ever makes a future visit to state 0} \mid X_1 = -1), \text{ by the Markov property} \\ &= pf_{1,0} + (1-p)f_{-1,0}. \end{aligned}$$

That is,

$$f_{0,0} = pf_{1,0} + (1-p)f_{-1,0}. \quad (*)$$

Let \mathfrak{F}_0 denote the event that the DTMC ever makes a future visit to state 0. Then $\mathfrak{F}_0 = \bigcup_{i=1}^{\infty} \{X_i = 0\}$. So

$$\begin{aligned} f_{1,0} &= \Pr(\mathfrak{F}_0 \mid X_0 = 1) \\ &= \Pr(\mathfrak{F}_0 \cap \{X_1 = 0\} \mid X_0 = 1) + \Pr(\mathfrak{F}_0 \cap \{X_1 = 2\} \mid X_0 = 1) \\ &= \Pr(\mathfrak{F}_0 \mid X_1 = 0, X_0 = 1) \Pr(X_1 = 0 \mid X_0 = 1) + \Pr(\mathfrak{F}_0 \mid X_1 = 2, X_0 = 1) \Pr(X_1 = 2 \mid X_0 = 1) \\ &= 1 \cdot \Pr(X_1 = 0 \mid X_0 = 1) + \Pr(\mathfrak{F}_0 \mid X_1 = 2, X_0 = 1) \Pr(X_1 = 2 \mid X_0 = 1) \\ &= \Pr(X_1 = 0 \mid X_0 = 1) + \Pr(\mathfrak{F}_0 \mid X_1 = 2, X_0 = 1) \Pr(X_1 = 2 \mid X_0 = 1) \\ &= (1-p) + \Pr(\mathfrak{F}_0 \mid X_1 = 2, X_0 = 1) \Pr(X_1 = 2 \mid X_0 = 1) \\ &= (1-p) + \Pr(\mathfrak{F}_0 \mid X_1 = 2, X_0 = 1) \cdot p \\ &= (1-p) + p \Pr(\mathfrak{F}_0 \mid X_1 = 2, X_0 = 1) \\ &= (1-p) + p \Pr(\mathfrak{F}_0 \mid X_1 = 2), \text{ by the Markov property} \end{aligned}$$

$$\begin{aligned}
&= (1-p) + p \Pr\left(\bigcup_{i=2}^{\infty} \{X_i = 0\} \cup \{X_1 = 0\} \mid X_1 = 2\right) \\
&= (1-p) + p \Pr\left(\bigcup_{i=2}^{\infty} \{X_i = 0\} \mid X_1 = 2\right) \\
&= (1-p) + p \Pr\left(\bigcup_{i=1}^{\infty} \{X_i = 0\} \mid X_0 = 2\right), \text{ by the stationary assumption} \\
&= (1-p) + p \Pr(\mathfrak{F}_0 \mid X_0 = 2) \\
&= (1-p) + p f_{2,0} \\
&= (1-p) + p \cdot f_{2,1} \cdot f_{1,0} \\
&= (1-p) + p \cdot f_{1,0} \cdot f_{1,0} \\
&= (1-p) + p f_{1,0}^2.
\end{aligned}$$

That is, $f_{1,0} = (1-p) + p f_{1,0}^2$. Solving for $f_{1,0}$ gives $f_{1,0} = \frac{1 \pm |p-q|}{2p}$ where $q := 1-p$. Let $r_1 = \frac{1+|p-q|}{2p}$ and $r_2 = \frac{1-|p-q|}{2p}$.

Case 1: $p = 0.5$.

Then we have $r_1 = r_2 = 1$. So $f_{1,0} = 1$. Similarly, we have $f_{-1,0} = 1$. So $f_{0,0} = p f_{1,0} + (1-p) f_{-1,0} = p + (1-p) = 1$. So the state 0 is recurrent and hence the entire DTMC is recurrent.

Case 2: $p \neq 0.5$.

We can show that if $p < 0.5$, $f_{1,0} = 1$ and $f_{-1,0} = \frac{p}{1-p} < 1$. If $p > 0.5$, $f_{-1,0} = 1$ and $f_{1,0} = \frac{1-p}{p} < 1$. So

$$f_{0,0} = p f_{1,0} + (1-p) f_{-1,0} = \begin{cases} 2p, & \text{if } p < 0.5 \\ 2(1-p), & \text{if } p > 0.5. \end{cases}$$

So $f_{0,0} < 1$ always.

Summary:

So

$$f_{0,0} = 2 \min\{p, 1-p\} \begin{cases} = 1, & \text{if } p = 0.5 \\ < 1, & \text{if } p \neq 0.5. \end{cases}$$

■

1.6 Limiting Behaviors

Proposition 1.6.1. *Let \mathcal{S} be a state space. Let $(X_n)_{n \in \mathbb{N}}$ be a discrete-time Markov chain with state space \mathcal{S} . Let i and j be two states in \mathcal{S} . Suppose that state j is transient. Then $\lim_{n \rightarrow \infty} P_{i,j}^{(n)} = 0$.*

Proof.

$$\begin{aligned}
\sum_{n=1}^{\infty} P_{i,j}^{(n)} &= \sum_{n=1}^{\infty} \sum_{k=1}^n f_{i,j}^{(k)} P_{j,j}^{(n-k)} = \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} f_{i,j}^{(k)} P_{j,j}^{(n-k)} \\
&= \sum_{k=1}^{\infty} f_{i,j}^{(k)} \sum_{n=k}^{\infty} P_{j,j}^{(n-k)} = \sum_{k=1}^{\infty} f_{i,j}^{(k)} \sum_{l=0}^{\infty} P_{j,j}^{(l)} \\
&= f_{i,j} \sum_{l=0}^{\infty} P_{j,j}^{(l)} = f_{i,j} \left(1 + \sum_{l=1}^{\infty} P_{j,j}^{(l)} \right) \\
&\leq 1 \cdot \left(1 + \sum_{l=1}^{\infty} P_{j,j}^{(l)} \right) = 1 + \sum_{l=1}^{\infty} P_{j,j}^{(l)} \\
&< 1 + \infty, \text{ since state } j \text{ is transient.} \\
&= +\infty.
\end{aligned}$$

That is, $\sum_{n=1}^{\infty} P_{i,j}^{(n)} < +\infty$. So $\lim_{n \rightarrow \infty} P_{i,j}^{(n)} = 0$. ■

1.6.1 Mean Recurrent Time

Notation.

$$N_i := \min\{n \in \mathbb{Z}_+ : X_n = i\}.$$

Definition (Mean Recurrent Time). *Let i be a recurrent state of a discrete-time Markov chain. We define the **mean recurrent time** of state i , denoted by m_i , to be the condition mean given by*

$$m_i := \mathbb{E}[N_i \mid X_0 = i].$$

In words, m_i represents the average time it takes the discrete-time Markov chain to make successive visits to state i .

Definition (Positive Recurrent and Null Recurrent). *Let i be a recurrent state. We say that i is **positive recurrent** if $m_i < +\infty$; and we say that i is **null recurrent** if $m_i = +\infty$.*

Proposition 1.6.2. *Let i and j be two states. Suppose $i \leftrightarrow j$ and i . Then i is positive recurrent (null recurrent) if and only if j is positive recurrent (null recurrent).*

Proposition 1.6.3. *A finite-state discrete-time Markov chain has no null recurrent states.*

Definition (Ergodic). *We say a state is **ergodic** if it is positive recurrent and aperiodic.*

1.6.2 Stationary Distribution

Definition (Stationary Distribution). *Let $\{p_i\}_{i=0}^{\infty}$ be a probability distribution over \mathbb{Z}_+ . We say that it is **stationary** if $\sum_{i=1}^{\infty} p_i = 1$ and $p = pP$ where P is the transition probability matrix.*

If a discrete-time Markov chain started according to a stationary distribution, then the probability of being in a given state remains unchanged over time.

Proposition 1.6.4. *If all the states of a discrete-time Markov chain are either null recurrent or transient, then there is no stationary distribution.*

Proposition 1.6.5. *An irreducible discrete-time Markov chain is positive recurrent if and only if a stationary distribution exists.*

1.6.3 Doubly Stochastic Transition Probability Matrix

Definition (Doubly Stochastic). *Let P be a transition probability matrix. We say that P is **doubly stochastic** if the column sums are all equal to 1.*

1.7 Limiting Behaviors

Theorem 3 (Basic Limit Theorem). *Let $\mathcal{S} = \mathbb{N}$ be a state space. Consider a irreducible, recurrent, and aperiodic discrete-time Markov chain with state space \mathcal{S} .*

$$\forall i, j \in \mathcal{S}, \quad \lim_{n \rightarrow \infty} P_{i,j}^{(n)} = \frac{1}{m_j},$$

where m_j is the mean recurrent time of state j . Moreover, if the discrete-time Markov chain is positive recurrent, then $\{\frac{1}{m_j}\}_{j \in \mathcal{S}}$ is the unique stationary distribution over \mathcal{S} .

Notation.

$$\pi_j := \lim_{n \rightarrow \infty} P_{i,j}^{(n)}.$$

Theorem 4. *Consider a finite-state, irreducible, and aperiodic discrete-time Markov chain. Suppose that it is doubly stochastic. Then the limiting probabilities are $\pi_i = \frac{1}{|\mathcal{S}|}$ for each $i \in \mathcal{S}$, where \mathcal{S} is the state space.*

Proof. Since the DTMC has finite state, it must have a positive recurrent state. Since the DTMC is irreducible and has a positive recurrent state, the entire DTMC is positive recurrent. Since the DTMC is irreducible, positive recurrent, and aperiodic, by the Basic Limit Theorem, we know that the limiting probabilities $\{\pi_j\}_{j \in \mathcal{S}}$ are $\frac{1}{m_j}$ and that $\{\frac{1}{m_j}\}_{j \in \mathcal{S}}$ is the unique stationary distribution over \mathcal{S} . So it remains to show that the probability distribution given by $\{\frac{1}{|\mathcal{S}|}\}_{j \in \mathcal{S}}$ is stationary. Note that

$$\begin{aligned} \sum_{k \in \mathcal{S}} \frac{1}{|\mathcal{S}|} P_{k,j} &= \frac{1}{|\mathcal{S}|} \sum_{k \in \mathcal{S}} P_{k,j} \\ &= \frac{1}{|\mathcal{S}|} \cdot 1, \text{ since } P \text{ is doubly stochastic} \end{aligned}$$

$$= \frac{1}{|\mathcal{S}|}.$$

So $\{\frac{1}{|\mathcal{S}|}\}_{j \in \mathcal{S}}$ is stationary. This completes the proof. \blacksquare

Proposition 1.7.1. *Let j be a state. Define a random variable A_k as $A_k := \mathbb{I}[X_k = j]$. Then*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n A_k \mid X_0 = i \right] = \pi_j.$$

Proof.

$$\begin{aligned} \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n A_k \mid X_0 = i \right] &= \frac{1}{n} \sum_{k=1}^n \mathbb{E}[A_k \mid X_0 = i] \\ &= \frac{1}{n} \sum_{k=1}^n \left[0 \cdot \Pr(A_k = 0 \mid X_0 = i) + 1 \cdot \Pr(A_k = 1 \mid X_0 = i) \right] \\ &= \frac{1}{n} \sum_{k=1}^n \Pr(A_k = 1 \mid X_0 = i) = \frac{1}{n} \sum_{k=1}^n \Pr(X_k = j \mid X_0 = i) \\ &= \frac{1}{n} \sum_{k=1}^n P_{i,j}^{(k)}. \end{aligned}$$

That is,

$$\mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n A_k \mid X_0 = i \right] = \frac{1}{n} \sum_{k=1}^n P_{i,j}^{(k)}.$$

So

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n A_k \mid X_0 = i \right] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P_{i,j}^{(k)} = \lim_{n \rightarrow \infty} P_{i,j}^{(n)} = \pi_j.$$

That is,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n A_k \mid X_0 = i \right] = \pi_j,$$

as desired. \blacksquare

The above proposition is essentially saying that $\pi_j := \lim_{n \rightarrow \infty} P_{i,j}^{(n)}$ is the long-run mean fraction of time that the process spends in state j .

Chapter 2

Convergence of Random Variables

2.1 Definitions

Definition (Convergence in Distribution). *Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of random variables. Let F_n be the cumulative distribution function of X_n . Let X be a random variable. Let F_X be the cumulative distribution function of X . We say that the sequence $\{X_n\}_{n \in \mathbb{N}}$ **converges in distribution** to X , denoted by $X_n \xrightarrow{d} X$, if $\forall x$ at which F is continuous,*

$$\lim_{n \rightarrow \infty} F_n(x) = F_X(x).$$

In this case, we say F_X is the asymptotic distribution of $\{X_n\}_{n \in \mathbb{N}}$.

Definition (Convergence in Probability). *Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of random variables. Let X be a random variable. We say that the sequence $\{X_n\}_{n \in \mathbb{N}}$ **converges in probability** to X , denoted by $X_n \xrightarrow{p} X$, if*

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0.$$

Or equivalently,

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1.$$

Definition (Almost Sure Convergence). *Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of random variables. Let X be a random variable. We say that the sequence $\{X_n\}_{n \in \mathbb{N}}$ **converges almost surely** to X if*

$$P(\lim_{n \rightarrow \infty} X_n = X) = 1.$$

Definition (Sure Convergence). *Let Ω be a sample space of the underlying probability space. Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of random variables. Let X be a random variable. We say that the sequence $\{X_n\}_{n \in \mathbb{N}}$ **converges surely** to X if*

$$\forall \omega \in \Omega, \quad \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

Definition (Convergence in Mean). *Let $r \geq 1$. Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of random variables. Let X be a random variable. We say that the sequence $\{X_n\}_{n \in \mathbb{N}}$ **converges in the r^{th} mean** to X , denoted by $X_n \xrightarrow{L^r} X$, if the r^{th} absolute moments $\mathbb{E}[|X_n|^r]$ and $\mathbb{E}[|X|^r]$ of X_n and X exists and*

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^r] = 0.$$

2.2 Markov's Inequality

Theorem 5 (Markov's Inequality). *Let X be a random variable. Let k and c be arbitrary positive numbers. Then*

$$P(|X| \geq c) \leq \frac{\mathbb{E}[|X|^k]}{c^k}.$$

Corollary.

$$P(|X - \mathbb{E}[X]| > k\sqrt{\text{var}[X]}) \leq \frac{1}{k^2}.$$

2.3 Properties

Proposition 2.3.1. *Convergence in probability implies convergence in distribution.*

Proposition 2.3.2. *Almost sure convergence implies convergence in probability.*

Proposition 2.3.3. *Convergence in the r^{th} mean for $r \geq 1$ implies convergence in probability.*

Proposition 2.3.4. *Let $\{X_i\}_{i \in \mathbb{N}}$ be a sequence of random variables. Let c be a constant. Then $\{X_i\}_{i \in \mathbb{N}}$ converges to c in distribution if and only if $\{X_i\}_{i \in \mathbb{N}}$ converges to c in probability.*

Sketch Proof.

$$\begin{aligned} P(|X_i - c| \geq \varepsilon) &= P(X_i \geq c + \varepsilon) + P(X_i \leq c - \varepsilon) \\ &= 1 - P(X_i < c + \varepsilon) + F_i(c - \varepsilon) \\ &\leq 1 - P(X_i \leq c + \varepsilon/2) + F_i(c - \varepsilon) \\ &= 1 - F_i(c + \varepsilon/2) + F_i(c - \varepsilon) \end{aligned}$$

$$\begin{aligned}
& \lim_{i \rightarrow \infty} [1 - F_i(c + \varepsilon/2) + F_i(c - \varepsilon)] \\
&= 1 - F(c + \varepsilon/2) + F(c - \varepsilon) \\
&= 1 - 1 + 0 \\
&= 0.
\end{aligned}$$

■

Proposition 2.3.5 (Continuous Map). *Let $\{X_i\}_{i \in \mathbb{N}}$ be a sequence of random variables. Let g be a continuous function on the X_i 's. Then*

- (1) *if $X_i \xrightarrow{d} X$, we have $g(X_i) \xrightarrow{d} g(X)$.*
- (2) *if $X_i \xrightarrow{p} c$, we have $g(X_i) \xrightarrow{p} g(c)$.*

Proposition 2.3.6 (Slutsky's Theorem). *Let $\{X_i\}_{i \in \mathbb{N}}$ and $\{Y_i\}_{i \in \mathbb{N}}$ be sequences of random variables. Suppose $X_i \xrightarrow{d} X$ for some random variable X and $Y_i \xrightarrow{p} c$ for some constant c . Then*

- (1) $X_i + Y_i \xrightarrow{d} X + c$.
- (2) $X_i Y_i \xrightarrow{d} cX$.
- (3) $X_i / Y_i \xrightarrow{d} X / c$.

2.4 Law of Large Numbers

Theorem 6 (Strong Law of Large Numbers). *Let $\{X_i\}_{i \in \mathbb{N}}$ be a sequence of independent and identically distributed random variables. Suppose that $\mathbb{E}[X_i] = \mu$ for some $\mu \in \mathbb{R}$ for all $i \in \mathbb{N}$. Then their cumulative average \bar{X}_n converges almost surely to μ . That is,*

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{almost surely}} \mu.$$

Chapter 3

Markov Decision Process

Chapter 4

Poisson Process

4.1 Homogeneous Poisson Process

4.1.1 Definitions

Definition (Homogeneous Poisson Process). *We say a counting process is a **homogeneous Poisson counting process** with rate $\lambda > 0$ if it has the following three properties:*

- $N(0) = 0$;
- *it has independent increments; and*
- *the number of events in any interval of length t is a Poisson random variable with parameter λt .*

Definition (Homogeneous Poisson Process). *We say a point process is a **homogeneous Poisson point process** with rate $\lambda > 0$ if the following two conditions hold:*

- *The probability $\mathbb{P}\{N(a, b] = n\}$ of the number $N(a, b]$ of points of the process in the interval $(a, b]$ being equal to some counting number n is given by*

$$\mathbb{P}\{N(a, b] = n\} = \frac{[\lambda(b-a)]^n}{n!} e^{-\lambda(b-a)}.$$

i.e. the number of arrivals in each finite interval has a Poisson distribution.

- *For any positive integer k and non-overlapping intervals $(a_1, b_1], \dots, (a_k, b_k]$,*

$$\mathbb{P}\left\{\bigwedge_{i=1}^k N(a_i, b_i] = n_i\right\} = \prod_{i=1}^k \mathbb{P}\{N(a_i, b_i] = n_i\}.$$

i.e. the number of arrivals in disjoint intervals are independent random variables.