Chapter 2 - Markov Chain

Applied Stochastic Process

CHEUNG Ying Lun

1 Markov chain

Definition 1. A (discrete time) Markov chain with state space $\mathcal{X}$ is a sequence $X_0, X_1, \ldots$ of $\mathcal{X}$-valued random variables such that for all states $i, j, k_0, k_1, \ldots$ and all times $t = 0, 1, 2, \ldots,$

$$\begin{aligned} P(X_{t+1} = j | X_t = i, X_{t-1} = k_{t-1}, \ldots) &= P(X_{t+1} = j | X_t = i) \\ &= p_{i,j} \end{aligned}$$

where $p_{i,j}$ depends only on the states i, j and not on the time t or the previous states $k_{t-1}, k_{t-2}, \ldots$.

A very important property of a Markov chain is that it is Markovian, i.e., only the current state affects the probability distribution of the future values of the process. *History does not matter.*

Definition 2. A stochastic process $\{X(t), t \in \mathcal{T}\}$ is said to be Markovian if

$$P(X(t_n) \in A | X(t), t \leq t_{n-1}) = P(X(t_n) \in A | X(t_{n-1}))$$

where $t_{n-1} < t_n$.

In fact, a Poisson process is also Markovian. The number of arrivals in the next moment $N(t + \delta)$, given the current number of arrivals $N(t)$, depends only on the current level, but not the history. It is a *continuous-time* Markov chain. In this chapter, we will focus on a *discrete-time* Markov chain.

2 Transition matrix

The main ingredient of a Markov chain is the *transition probability*, p_{ij}, which states the probability of a transition from some state i to state j. In a Markov chain, this probability is unchanged throughout. We can collect all transition probabilities in the matrix form and form the *transition matrix* $\mathbf{P}$.

$$\mathbf{P} = (p_{i,j})_{i,j} = \begin{pmatrix} p_{1,1} & \cdots & p_{1,M} \\ \vdots & \ddots & \vdots \\ p_{M,1} & \cdots & p_{M,M} \end{pmatrix}$$

If the initial state is random with probability $\mathbf{q}_0 = (q_{01}, \ldots, q_{0M})$, then the unconditional probability at period 1 is given by

$$\begin{aligned} P(X_1 = i) = q_{1i} &= \sum_{j=1}^{M} P(X_1 = i \cap X_0 = j) \\ &= \sum_{j=1}^{M} P(X_0 = j) P(X_1 = i | X_0 = j) \\ &= \sum_{j=1}^{M} q_{0j} p_{ji}. \end{aligned}$$

The above equation can be interpreted easily: the probability that the Markov chain is at state i at time $n = 1$ is the sum of the probabilities that it is at j at time $n = 0$ *and* transit to state i. In matrix form,

$$\begin{aligned} \mathbf{q}_1 &= \begin{pmatrix} \sum_{j=1}^{M} q_{0j} P_{j1} & \cdots & \sum_{j=1}^{M} q_{0j} P_{jM} \end{pmatrix} \\ &= \begin{pmatrix} q_{01} & \cdots & q_{0M} \end{pmatrix} \begin{pmatrix} p_{11} & \cdots & p_{1M} \\ \vdots & \ddots & \vdots \\ p_{M1} & \cdots & p_{MM} \end{pmatrix} \\ &= \mathbf{q}_0 \mathbf{P} \end{aligned}$$

In general, the $(m+n)$-step transition probability can be obtained by the Chapman-Kolmogorov Equations.

Theorem 2.1. *Let $P_{ij}^{n} = P(X_{n+k} = j | X_k = i)$ be the n-step transition probabilities. Then the Chapman-Kolmogorov equations are*

$$p_{ij}^{(n+m)} = \sum_{k=1}^{M} p_{ik}^{(n)} p_{kj}^{(m)} \qquad \text{for all } n, m \geq 0, \text{ all } i, j.$$

Proof. The $(m + n)$-step transition probability can be written as

$$\begin{aligned} p_{ij}^{(m+n)} &= P(X_{m+n} = j | X_0 = i) \\ &= \sum_{k=1}^{M} P(X_{m+n} = j \cap X_n = k | X_0 = i). \end{aligned}$$

Next, by the definition of conditional probability$P(A \cap B) = P(A|B)P(B)$,

$$P(X_{m+n} = j \cap X_n = k | X_0 = i) = P(X_{m+n} = j | X_n = k \cap X_0 = i) P(X_n = k | X_0 = i).$$

However, by the Markovian property of X_n, the information $X_0 = i$ is not useful when we know $X_n = k$, i.e.,

$$P(X_{m+n} = j | X_n = k \cap X_0 = i) = P(X_{m+n} = j | X_n = k).$$

Now gathering terms we get

$$
\begin{aligned}
p_{ij}^{(m+n)} &= \sum_{k=1}^{M} P(X_{m+n} = j \cap X_n = k | X_0 = i) \\
&= \sum_{k=1}^{M} P(X_{m+n} = j | X_n = k) P(X_n = k | X_0 = i) \\
&= \sum_{k=1}^{M} p_{kj}^{(m)} p_{ik}^{(n)}.
\end{aligned}
$$

In matrix form, it gives us $\mathbf{P}^{(m+n)} = \mathbf{P}^{(m)}\mathbf{P}^{(n)}$. □

Now, if we consider the two-step transition matrix,

$$\mathbf{P}^{(2)} = \mathbf{P}^{(1+1)} = \mathbf{P}\mathbf{P} = \mathbf{P}^2.$$

By induction, we can show that

$$\mathbf{P}^{(n)} = \mathbf{P}^n.$$

If the initial state is random with probability $\mathbf{q}_0$, then the unconditional probability at period n is given by $\mathbf{q}_n = \mathbf{q}_0\mathbf{P}^n$. If n is very large, it may converge to a stationary probability.

Theorem 2.2. *The limiting probabilities of a Markov chain always exist and will not depend on the initial state if the chain is aperiodic and irreducible. Moreover, it is given by the solution of*

$$\pi_j = \sum_i \pi_i P_{ij}, \qquad \sum_j \pi_j = 1$$

or in matrix form $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$*,*

The meaning of *periodicity* and *reducibility* will be discussed in the next section. To find out the stationary probability, notice that by taking transpose of the equation $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$,

$$\mathbf{P}'\boldsymbol{\pi}' = \boldsymbol{\pi}' \implies (\mathbf{P}' - \mathbf{I})\boldsymbol{\pi}' = \mathbf{0}.$$

Recalling that the definition of eigenvalue λ and eigenvector $\mathbf{v}$ of any matrix $\mathbf{A}$ is

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}.$$

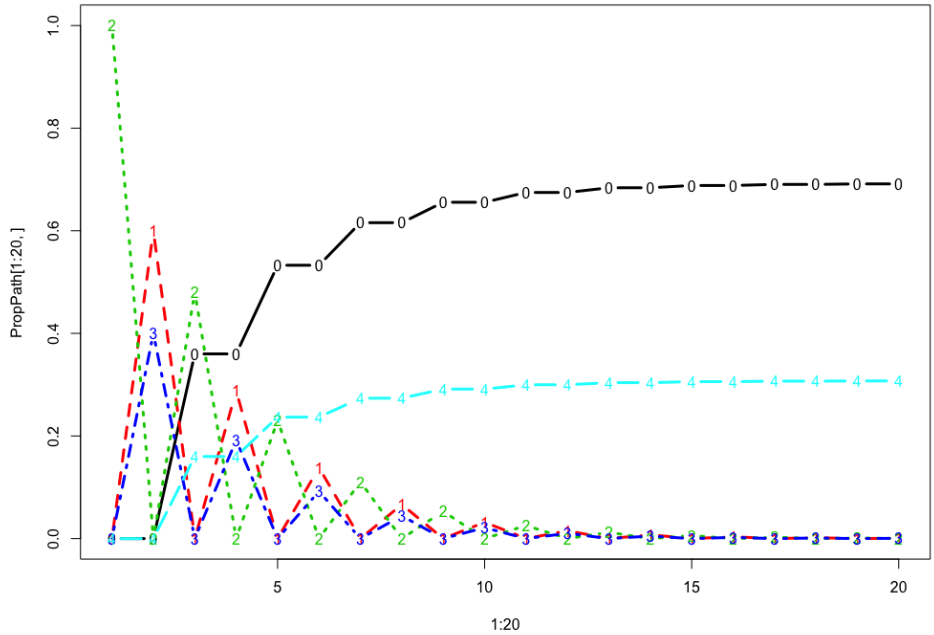
Therefore, the vector $\boldsymbol{\pi}'$, if exists, is actually the eigenvector of matrix $\mathbf{P}'$ corresponding to the eigenvalue $\lambda = 1$, up to normalization $\boldsymbol{\pi}\mathbf{1} = 1$.

Example (Gambler's ruin). Consider a gambling game in which on any turn you win \$1 with probability $p = 0.4$ or lose \$1 with probability $1 - p = 0.6$. Suppose initially you have \$2 and you will quit the game either if your fortune reaches \$4 or \$0. Let X_n be the amount you have after n plays.

In this example, the state space, i.e., the amount of money you can have at any time n, is $\mathcal{X} = \{0, 1, 2, 3, 4, 5\}$. The transition matrix is given by

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Given the initial probability $\mathbf{q}_0 = (0, 0, 1, 0, 0)$, the unconditional probabilities at different n are plotted in the diagram below:



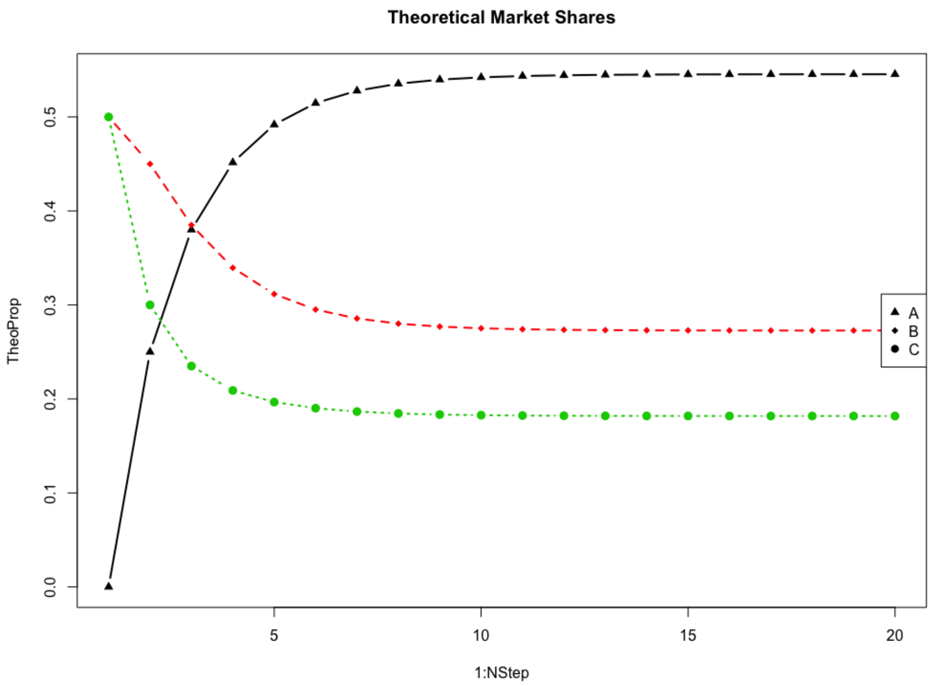
We know that the player will end up with either \$0 or \$4. The probabilities otherwise converge to zero very quickly. Although the unconditional probability seems to stabilize in the above diagram, the limiting probability actually does not exists. In fact, if we compute the eigenvectors of $\mathbf{P}'$, we will find two eigenvectors $\boldsymbol{\pi} = (1, 0, 0, 0, 0)$ and $\boldsymbol{\pi} = (0, 0, 0, 0, 1)$ that correspond to the eigenvalue $\lambda = 1$. With a different initial probability, the unconditional probability will also converge to a different value. (Students are encouraged to play around with the R file I provided.) $\square$

Example (Brand preference). Suppose there are three types of laundry detergent, and let X_n be the brand chosen on the nth purchase. Customers who try these brands are satisfied and choose the same thing again with probability 0.8, 0.6, and 0.4 respectively. When they change they pick one of the other two brands at random. Suppose also that $\mathbf{q}_0 = (0, 0.5, 0.5)$.

In this example, the transition matrix is given by

$$\mathbf{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}.$$

The unconditional probability can be viewed as the market share of each brand, and is plotted below.



We see that the market share stabilizes. After normalization, the eigenvector of $\mathbf{P}'$ corresponding to $\lambda = 1$ is

$$\boldsymbol{\pi} = \begin{pmatrix} \frac{6}{11} & \frac{3}{11} & \frac{2}{11} \end{pmatrix},$$

which is also limiting probability of the Markov chain. □

3 Classification of states

3.1 Accessibility and class

A Markov chain is a model of (random) transitions among states. From the transition matrix $\mathbf{P}$ we can deduce a lot of properties of the states. In this section, we try to classify the states into different groups or types according to their properties.

Definition 3. 1. State j is said to be *accessible* from state i if $p_{ij}^{(n)} > 0$ for some $n \geq 0$.

2. Two states i and j that are accessible to each other are said to *communicate*, and we write $i \leftrightarrow j$.

3. Two states that communicate are said to be in the same *class*.

4. The Markov chain is said to be *irreducible* if there is only one class, i.e., if all states communicate with each other.

The above definitions are all related with each other. Accessibility means it is possible (i.e., with probability larger than zero) to transit from one state to another after some steps. If $p_{ij}^{(n)} = 0$ for all n,

$$\begin{aligned}
P(\text{ever be in } j|\text{start in } i) &= P\left(\bigcup_{n=0}^{\infty}(X_n = j)|X_0 = i\right) \\
&\leq \sum_{n=0}^{\infty} P(X_n = j|X_0 = i) \\
&= \sum_{n=0}^{\infty} p_{ij}^{(n)} \\
&= 0.
\end{aligned}$$

Therefore, if state j is not accessible from state i, then starting in i, the process will never end up in j.

Example (Gambler's ruin). Continuing with the example, the relations among states in the state space $\mathcal{X} = \{0, 1, 2, 3, 4\}$ are

$$0 \leftarrow 1 \leftrightarrow 2 \leftrightarrow 3 \rightarrow 4$$

It is easy to see that every state is accessible from states $\{1, 2, 3\}$. However, after we get to states $\{0\}$ or $\{4\}$, we cannot get to other states anymore. Therefore, the states $\{1, 2, 3\}$ communicate with each other and form one class .$\{0\}$ and $\{4\}$ do not communicate with each other or other states and each of them forms a class on its own. There are in total three classes:

$$\{0\},\ \{1, 2, 3\}.\ \{4\}.$$

Since there is more than one class, this Markov chain is reducible. This also explains why Theorem 2.2 does not apply: the limiting probabilities do not exist. □

Example (Brand preference). In this example, it is possible to switch from any state to any state. Therefore, all three states are accessible from each other and they all communicate. There is only one class. The Markov chain is irreducible. □

Property. 1. $i \leftrightarrow i$ for all i.

2. If $i \leftrightarrow j$, then $j \leftrightarrow i$.

3. If $i \leftrightarrow j$, and $j \leftrightarrow k$, then $i \leftrightarrow k$.

4. Any two classes of states are either identical or disjoint.

Proof. The first two properties are trivial. They are true by definition. To show the third point, notice that $i \leftrightarrow j$ implies that for some $p_{ij}^{(n)} > 0$ for some n and $j \leftrightarrow k$ implies that $p_{jk}^{(m)} > 0$ for some m. By the Chapman-Kolmogorov equations,

$$p_{ik}^{(n+m)} = \sum_{r=1}^{M} p_{ir}^{(n)} p_{rk}^{(m)} \geq p_{ij}^{(m)} p_{jk}^{(n)} > 0.$$

We can show following the same steps that $p_{ki}^{(s)} > 0$ for some s. Therefore, $i \leftrightarrow k$. The fourth point is a logical consequence of the third point. □

3.2 Recurrent or transient states

Definition 4. For any state i, let $f_{i,j}$ be the probability that, starting in state i, the process will ever enter state j. State i is said to be *recurrent* if $f_{i,i} = 1$ and *transient* if $f_{i,i} < 1$.

A recurrent state is one that once we enter the state, we will eventually re-enter the same state in the future, i.e.,

$$f(i,i) = P(i \to \cdots \to i) = 1.$$

In contrary, a transient state is one that, with positive probability, the process will never re-visit the same state after it leaves the state. This can happen when the process jumps to another class. Remember that if two states, say i and j, are in different classes, they do not communicate. Therefore, if $i \to j$, then $j \nrightarrow i$. If the process enters j after leaving i, it will never visit i again.

Example (Gambler's ruin). If the player has \$0 or \$4, they will quit and their fortune will remain unchanged thereafter. Therefore, obviously $\{0\}$ and $\{4\}$ are two recurrent states. For the states $\{1,2,3\}$, since there is positive probability that X_n reaches 0 or 4 after leaving the state and before re-entering the same state, they are transient states. □

Example (Brand preference). Since it is always possible to re-enter any state, all states are recurrent in this example. □

The above examples already give some hints about the properties of recurrent/transient states, which we will discuss below.

Property.

1. If state i is recurrent, then starting in state i, the process will reenter state i infinitely often.

2. If state i is recurrent and $i \leftrightarrow j$, then $f_{i,j} = f_{j,j} = 1$. Therefore, recurrence is a class property.

3. It is impossible to go from a recurrent to a transient state.

4. If state i is transient, then starting in state i, the number of time periods that the process will be in state i has a geometric distribution with finite mean $1/(1 - f_{i,i})$.

5. In a finite state Markov chain, at least one class of state is recurrent.

Proof. 1. It is easy to verify: since the state i is recurrent, starting in i, the process will eventually return to i, i.e., $f(i,i) = P(i \to \cdots \to i) = 1$. Due to the Markovian property of a Markov chain, the process will eventually return to i *twice* is

$$P(i \to \cdots \to i \to \cdots \to i) = f(i,i)^2 = 1.$$

By induction, the probability that the process will return to i infinitely many times after infinitely many steps is still one.

2. Since $i \leftrightarrow j$, there is a positive probability that starting in i, the process will reach j. Therefore, the probability that starting in i, the process reaches j before returning to i is $g_{ij} = P(i \to \cdots \to j \to \cdots \to i) > 0$. The probability that the process starts in i and returns to i without passing through j for n times is $(1 - g_{ij})^n$. From the above, we know that starting in i, the probability that the process never reaches j is therefore $\lim_{n\to\infty}(1 - g_{ij})^n = 0$, implying that the process will eventually reach j, i.e., $f_{i,j} = 1$. The probability that the above happens twice, i.e., $i \to \cdots \to j \to \ldots i \to \cdots \to j \to \cdots \to i$, is $f_{i,j}^2 = 1$, implying that $f_{j,j} = 1$.

3. It can be proved by contradiction. Suppose it is possible to go from a recurrent state i to a transient state j. However, by definition, starting in i, the process will eventually go back to i. It means that it is possible that the process starts in i, then goes to j, and finally comes back to i. This will then imply that i and j communicate and they must be in the same class. Moreover, by the Markovian property, after returning to i, the process will eventually go back to j again, making j also recurrent. Therefore, it contradicts with our initial assumption and must be wrong.

4. First, since recurrence is a class property, starting in i, the process will never come back to i if it jumps to another class. The probability of this happening is $1 - f_{i,i}$. If we define E as the event that the process starts in i and jumps to another class before reaching i again, then the probability of success is $P(E) = 1 - f_{i,i}$. E is a Bernoulli trial and the number of time periods that the process will be in i is the same as the number of failed trials, which follows a geometric distribution with $p = P(E) = 1 - f_{i,i}$.

5. Since the number of time periods that the process will be in a transient state follows a geometric distribution, the probability that it is finite equals one. It means that after a finite amount of time, say T_i, the Markov chain will never go back to state i. Suppose that the state space is $\mathcal{X} = \{1, \ldots, M\}$, then if all states are transient, the process will never go back to any of the state after $T_1 + \cdots + T_M$ periods. However, the process must visit a state. Therefore, at least one class of state is recurrent.

□

The above tells us about the properties of recurrent and transient states. We know that the process will eventually go to a recurrent class. Now, we focus on the transient states and consider the question: on average, how many time periods will the process stay in a transient state? To answer this question, suppose the transient states are collected in $T = \{1, \ldots, t\}$. We define a transition matrix from one transient state to another transient state

$$\mathbf{P}_T = \begin{pmatrix} p_{11} & \cdots & p_{1t} \\ \vdots & \ddots & \vdots \\ p_{t1} & \cdots & p_{tt} \end{pmatrix}.$$

We can compute the mean time spent in transient states using the following theorem.

Theorem 3.1. *Let $\mathbf{P}_T$ specifies only the transition probabilities from transient states into transient states, and let $\mathbf{S} = (s_{ij})_{i,j=1,\ldots,t}$ denote the matrix that collects the expected number of time periods that the Markov chain is in state j, given that it starts in state i. Then,*

$$\mathbf{S} = (\mathbf{I} - \mathbf{P}_T)^{-1},$$

where $\mathbf{I}$ is the identity matrix.

Proof. Let $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Conditioning on the initial transition,

$$s_{ij} = \delta_{ij} + \sum_{k=1}^{M} p_{ik} s_{kj} = \delta_{ij} + \sum_{k=1}^{t} p_{ik} s_{kj},$$

where the second equality above comes from the fact that it is impossible to go from a recurrent state to a transient state. Now writing the above in matrix form,

$$\begin{aligned} \mathbf{S} = \begin{pmatrix} s_{11} & \cdots & s_{1t} \\ \vdots & \ddots & \vdots \\ s_{t1} & \cdots & s_{tt} \end{pmatrix} &= \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} + \begin{pmatrix} p_{11} & \cdots & p_{1t} \\ \vdots & \ddots & \vdots \\ p_{t1} & \cdots & p_{tt} \end{pmatrix} \begin{pmatrix} s_{11} & \cdots & s_{1t} \\ \vdots & \ddots & \vdots \\ s_{t1} & \cdots & s_{tt} \end{pmatrix} \\ &= \mathbf{I} + \mathbf{P}_T \mathbf{S}. \end{aligned}$$

Re-arranging terms and solving for $\mathbf{S}$, we obtain $\mathbf{S} = (\mathbf{I} - \mathbf{P}_T)^{-1}$. ☐

Finally, we also want to find out the probability that the probability that a Markov chain ever makes a transition into state j given that it starts in i. We already know that if i is recurrent, then $f_{i,i} = 1$ and $f_{i,j} = 0$ if j is transient. The focus is on the case when both i and j are transient.

Theorem 3.2. *The probability that the Markov chain ever makes a transition into a transient state j given that it starts in a transient state i is given by*

$$f_{i,j} = \frac{s_{ij} - \delta_{i,j}}{s_{jj}}, \qquad \delta_{i,j} = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases}$$

Proof. We can separate s_{ij} into two cases as follows.

$$\begin{aligned} s_{ij} &= \mathbb{E}\left[\text{time in } j | \text{start in } i, \text{ ever transit to } j\right] f_{ij} \\ &\quad + \mathbb{E}\left[\text{time in } j | \text{start in } i, \text{ never transit to } j\right](1 - f_{ij}) \\ &= (\delta_{ij} + s_{jj}) f_{ij} + \delta_{ij}(1 - f_{ij}) \\ &= \delta_{ij} + f_{ij} s_{jj}. \end{aligned}$$

Solving the above equation yields $f_{ij} = s_{jj}^{-1}(s_{ij} - \delta_{ij})$. ☐

3.3 Positive recurrence

In the last part, we introduce the concepts of recurrence and transience, and discuss the properties of the transient states. In this part, we will further separate recurrent states into positive or null recurrent states.

Definition 5. If state j is recurrent, let m_j be the expected number of transitions that it takes the Markov chain when starting in state j to return to that state. Then, the recurrent state j is said to be *positive recurrent* if $m_j < \infty$ and *null recurrent* if $m_j = \infty$.

Let $T_j = \min\{n > 0 : X_n = j\}$ be the number of transitions until the Markov chain first makes a transition into state j, then $m_j = \mathbb{E}\left[T_j | X_0 = j\right]$. The relation between m_j and the long-run proportion of time in state j, π_j, is stated below.

Theorem 3.3. *If the Markov chain is irreducible and recurrent, then for any initial state $\pi_j = 1/m_j$.*

Proof. Suppose the process starts in state i, and that it takes T_1 to reach j for the first time. Then, denote the amount of time for the process to return to j for the k-th time as T_k. Then, at the time when the process return to j for the n-th time, the proportion of time the process spends in state j is given by

$$\pi_{j,n} = \frac{n}{\sum_{i=1}^{n} T_n}.$$

As $n \to \infty$,

$$\pi_j = \lim_{n\to\infty} \frac{n}{\sum_{i=1}^{n} T_n} = \lim_{n\to\infty} \frac{1}{n^{-1}\sum_{i=1}^{n} T_n} = \lim_{n\to\infty} \frac{1}{n^{-1}T_1 + n^{-1}\sum_{i=2}^{n} T_n}.$$

Since the Markov chain is irreducible and recurrent, states i and j communicate and it takes finitely many time for the process to reach j from i. Therefore, $n^{-1}T_1 \to 0$. Next, following the Markovian property of the Markov chain, T_j are independent and identically distributed for all $j \geq 2$. By the Law of Large Number,

$$n^{-1}\sum_{i=2}^{n} T_n = \frac{n-1}{n}\frac{1}{n-1}\sum_{i=2}^{n} T_n \xrightarrow{p} m_j.$$

Therefore, $\pi_j = 1/m_j$. □

Property.

1. If i is positive recurrent and $i \leftrightarrow j$, then j is positive recurrent.
2. Null recurrence is also a class property.
3. An irreducible finite state Markov chain must be positive recurrent.

Proof.

1. Suppose i is positive recurrent and $i \leftrightarrow j$. Then, there exists an n such that $p_{i,j}^{(n)} > 0$. Now let $\pi_i > 0$ be the long-run proportion of time that the chain is in i. Then, $\pi_i p_{i,j}^{(n)}$ represents the long-run proportion that the process is in i and then in j after n transitions. Equivalently, it is the long-run proportion that the process is in j and was in

i n transitions before, which is smaller than or equal to the long-run proportion that the chain is in j. Finally,

$$\pi_j \geq \pi_i p_{i,j}^{(n)} > 0.$$

Since $\pi_j > 0$, $m_j = 1/\pi_j < \infty$.

2. It follows directly from the above. Since recurrence is a class property, if i is recurrent and $i \leftrightarrow j$, then j is also recurrent. Since positive recurrence is also a class property, i and j can either be both positive or null recurrent. Therefore, null recurrence is also a class property.

3. We already know that an irreducible finite state Markov chain must be recurrent and that positive/null recurrence is a class property. If every state is null recurrence, the long-run proportion of each state will be $\pi_i = 1/m_i = 0$, which is impossible since there are only finitely many states. Therefore, the chain can only be positive recurrent.

□

More can be said about the long-run proportion.

Theorem 3.4. *The long-run proportions of an irreducible Markov chain are the unique solution of the equations*

$$\pi_j = \sum_i \pi_i p_{ij}, \qquad \sum_j \pi_j = 1$$

or in matrix form $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$, *if the chain is positive recurrent. Moreover, if there is no solution of the preceding linear equations, then the Markov chain is either transient or null recurrent and all* $\pi_j = 0$.

Proof. Noting that π_i is the long-run proportion of transition from state i, the long-run proportion of transitions from state i to j is $\pi_i p_{ij}$. Therefore, the long-run proportion in state j is equivalent to adding up the long-run proportion of transitions from all states i to j, i.e.,

$$\pi_j = \sum_{i=1} \pi_i p_{ij}.$$

Writing in matrix form, $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$. □

The long-run proportions are also called the *stationary* probabilities. The reason is that if the initial state is chosen according to π_j, then the probability of being in state j at any time n is also equal to π_j. You may also notice that they are the same as the limiting probabilities, if they exist. The relations between them will be discussed next.

3.4 Periodicity

Definition 6. The *period* of a state is the largest number that will divide all the $n \geq 1$ for which $p_{i,i}^{(n)} > 0$. A Markov chain is said to be *aperiodic* if all of its states have period one, and it is *periodic* otherwise.

Example (Gambler's ruin). In this example, one can only either earn one dollar or lose one dollar unless one has already quit the game. We have for example

$$p_{22} = 0, \qquad p_{22}^{(2)} > 0, \qquad p_{22}^{(3)} = 0, \qquad p_{22}^{(4)} > 0...$$

Then, we say the period of state 2 is 2, and that the process is periodic. □

Example (Brand preference). In this example, $p_{ii} > 0$ for all i. Obviously, the process is aperiodic. □

Theorem 3.5.

1. *A periodic Markov chain does not have limiting probabilities.*
2. *The limiting probabilities of an irreducible, aperiodic chain always exist and do not depend on the initial state.*
3. *The limiting probabilities, when they exist, will equal the long-run proportions.*

Combining the second and third part of the above theorem, the limiting probabilities of an irreducible, positive recurrent, aperiodic Markov chain is always the same as the long-run proportion. In other words, as long as we observe the random process for a long enough period of time, we will be able to estimate the limiting probabilities be simply computing the long-run proportion of each state. Such process is said to be ergodic.

Definition 7. The stochastic process $\{X(t), t \in \mathcal{T}\}$ is said to be *ergodic* if any characteristics of the process can be obtained, with probability 1, from a single realization.