Chapter 2 Discrete Markov Chains

Part I

A **stochastic process** $\{X_t \mid t \in T\}$ is a collection of random variables defined on some probability space. That is, for each $t \in T$, X_t is a random variable. The index *t* is often interpreted as time and, as a result, we refer to X_t as the **state** of the process at time *t*. For example, X_t might equal the total number of customers that have entered a supermarket by time *t*; or the total amount of sales that have been recorded in the market by time *t*; etc.

The set *T* is called the **index** set of the process. When *T* is a countable set the stochastic process is said to be a **discrete-time** process. If *T* is an interval of the real line, the stochastic process is said to be a **continuous-time** process. For instance, $\{X_n \mid n = 0, 1, 2, \cdots\}$ is a discrete-time stochastic process indexed by the nonnegative integers; while $\{X_t \mid t \geq 0\}$ is a continuous-time stochastic process indexed by the nonnegative real numbers.

The **state space** of a stochastic process is defined as the set of all possible values that the random variables X_t can assume. Thus, a stochastic process is a family of random variables that describes the evolution through time of some (physical) process.

In this chapter, we consider a stochastic process $\{X_n \mid n = 0, 1, 2, \cdots\}$ that takes on a finite or countable number of possible values. Unless otherwise mentioned, this set of possible values of the process will be denoted by a countable set *S* of states, which are often labelled by the non-negative integers 0, 1, 2, … (sometimes by positive integers 1,2, …).

In this chapter, we suppose that whenever the process is in state *i*, there is a fixed probability $P_{i,j}$ that it will next be in state *j*. That is, we suppose that

$$P_{i,j} = P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) \tag{1}$$

for all states $i_{n-1}, \cdots, i_0, i, j$ and all $n \geq 0$. Such a stochastic process is known as a Markov chain. Equation (1) may be interpreted as stating that, for a Markov chain, the conditional probability of any future state X_{n+1}, given the past states $X_{n-1}, \cdots, X_0$ and the present state X_n, is independent of the past states and depends only on the present state, i.e.,

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

This is called the Markovian property. The value $P_{i,j}$ represents the probability that the process will, when in state *i*, next make a transition into state *j*. Since probabilities are nonnegative and since the process must make a transition into some state, we have that

$$P_{i,j} \geq 0, \quad i, j \geq 0; \quad \sum_{j \in S} P_{i,j} = 1, \quad i = 0, 1, \cdots$$

Let $\boldsymbol{P}$ denote the one-step transition matrix of probabilities $P_{i,j}$, so that

$$\boldsymbol{P} = \begin{pmatrix} P_{0,0} & P_{0,1} & \cdots \\ P_{1,0} & P_{1,1} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

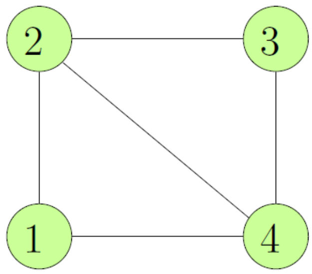
Clearly, sum of each row of $\boldsymbol{P}$ is 1.

Theorem 1.1

Every transition matrix has 1 as an eigenvalue with eigenvector $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$.

Theorem 1.2 (Random walk on a graph)

Consider a graph with undirected edges as below. At every step, a random walker moves from the current vertex to a randomly chosen neighboring vertex (with equal probability).



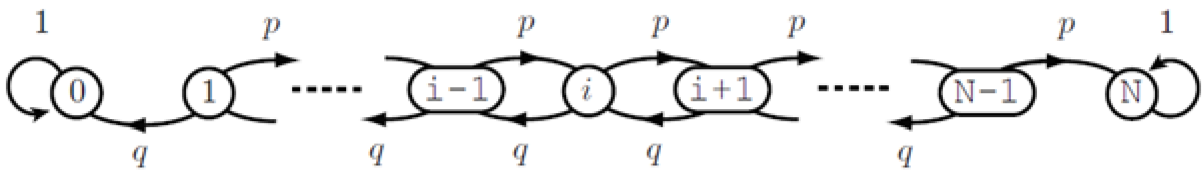
The transition matrix is

$$P = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix}.$$

Example 1.3 (Gambler's Ruin Problem)

Consider a gambler who, at each play of the game, either wins \$1 with probability p or loses \$1 with probability $1-p$. If we suppose that the gambler quits playing either when he goes broke or he attains a fortune of \$$N$, then the gambler's fortune is a Markov chain having transition probabilities

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \quad i = 1, 2, \cdots, N-1, \quad P_{0,0} = P_{N,N} = 1.$$



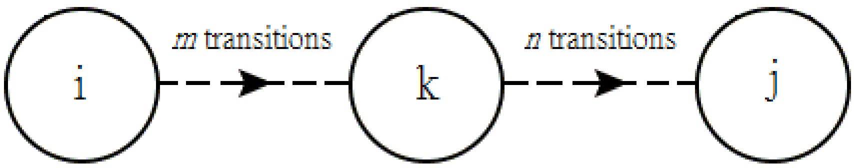
A state i is said to be **absorbing** if once entered they are never left, i.e., $P_{i,i} = 1$. States 0 and N are absorbing. Note that the preceding is a finite state random walk with absorbing barriers (states 0 and N). The transition matrix is

$$\boldsymbol{P} = \begin{pmatrix} 1 & & & & & & \\ 1-p & 0 & p & & & & \\ & 1-p & 0 & p & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1-p & 0 & p \\ & & & & & 1 \end{pmatrix}.$$

We have already defined the one-step transition probabilities $P_{i,j}$. We now define the n-step transition probabilities $P_{i,j}^{n}$ to be the probability that a process in state i will be in state j after n additional transitions. That is,

$$P_{i,j}^{n} = P(X_{n+k} = j \mid X_k = i), \quad n \geq 0, \; i, j \geq 0$$

Obviously, $P_{i,j}^{1} = P_{i,j}$. The **Chapman–Kolmogorov equations** provide a method for computing these n-step transition probabilities. Noting that $P_{i,k}^{m} P_{k,j}^{n}$ represents the probability that starting in i the process will go to state j in $m + n$ transitions through a path which takes it into state k at the m-th transition.



Hence, summing over all intermediate states k yields the probability that the process will be in state j after $m + n$ transitions. Formally, we have

$$
\begin{aligned}
P_{i,j}^{m+n} &= P(X_{m+n} = j \mid X_0 = i) \\
&= \sum_k P(X_{m+n} = j, X_m = k \mid X_0 = i) \\
&= \sum_k \frac{P(X_{m+n} = j, X_m = k, X_0 = i)}{P(X_0 = i)} \\
&= \sum_k \frac{P(X_{m+n} = j, X_m = k, X_0 = i)}{P(X_m = k, X_0 = i)} \frac{P(X_m = k, X_0 = i)}{P(X_0 = i)} \\
&= \sum_k P(X_{m+n} = j \mid X_m = k, X_0 = i) P(X_m = k \mid X_0 = i) \\
&= \sum_k P_{i,k}^m P_{k,j}^n
\end{aligned}
$$

If we let $\boldsymbol{P}^{(n)}$ denote the n-step transition matrix of probabilities $P_{i,j}^n$, then the above equation asserts that

$$\boldsymbol{P}^{(m+n)} = \boldsymbol{P}^{(m)} \boldsymbol{P}^{(n)}.$$

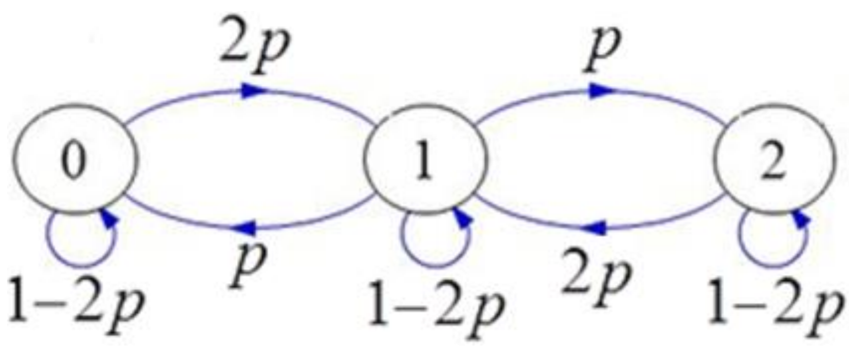
Hence, by induction

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

That is, the n-step transition matrix may be obtained by multiplying the matrix $\boldsymbol{P}$ by itself n times. Clearly, $\boldsymbol{P}^{(0)} = \boldsymbol{P}^0 = I$, an identity matrix.

Example 1.4

Consider the following Markov chain where $0 \le p \le \frac{1}{2}$.



Transition matrix is $\boldsymbol{P}=\begin{pmatrix}1-2p & 2p & 0\\ p & 1-2p & p\\ 0 & 2p & 1-2p\end{pmatrix}$. The characteristic polynomial of $\boldsymbol{P}$ is

$$\begin{aligned}
\det\left(\lambda I_3-\boldsymbol{P}\right)&=\det\begin{pmatrix}\lambda-(1-2p) & -2p & 0\\ -p & \lambda-(1-2p) & -p\\ 0 & -2p & \lambda-(1-2p)\end{pmatrix}\\
&=\left(\lambda-(1-2p)\right)\det\begin{pmatrix}\lambda-(1-2p) & -p\\ -2p & \lambda-(1-2p)\end{pmatrix}+p\cdot\det\begin{pmatrix}-2p & 0\\ -2p & \lambda-(1-2p)\end{pmatrix}\\
&=\left(\lambda-(1-2p)\right)\left(\left(\lambda-(1-2p)\right)^2-2p^2\right)-2p^2\left(\lambda-(1-2p)\right)\\
&=\left(\lambda-(1-2p)\right)\left(\left(\lambda-(1-2p)\right)^2-4p^2\right)\\
&=\left(\lambda-(1-2p)\right)\left(\lambda-(1-4p)\right)\left(\lambda-1\right).
\end{aligned}$$

If $p=0$, then

$$\boldsymbol{P}^{(n)}=\boldsymbol{P}^n=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\ \text{ for all }\ n\ge 1.$$

If $0<p<\frac{1}{2}$, then for $\lambda=1$,

$$N(I_3-\boldsymbol{P})=N\begin{pmatrix}2p & -2p & 0\\ -p & 2p & -p\\ 0 & -2p & 2p\end{pmatrix}=span\left\{\begin{pmatrix}1\\1\\1\end{pmatrix}\right\}$$

For $\lambda=1-4p$,

$$N((1-4p)I_3-\boldsymbol{P})=N\begin{pmatrix}-2p & -2p & 0\\ -p & -2p & -p\\ 0 & -2p & -2p\end{pmatrix}=span\left\{\begin{pmatrix}1\\-1\\1\end{pmatrix}\right\}$$

For $\lambda=1-2p$,

$$N((1-2p)I_3-\boldsymbol{P})=N\begin{pmatrix}0 & -2p & 0\\ -p & 0 & -p\\ 0 & -2p & 0\end{pmatrix}=span\left\{\begin{pmatrix}-1\\0\\1\end{pmatrix}\right\}$$

Then we have

$$\boldsymbol{P}=\begin{pmatrix}1-2p & 2p & 0\\ p & 1-2p & p\\ 0 & 2p & 1-2p\end{pmatrix}$$

$$=\begin{pmatrix}1 & 1 & -1\\ 1 & -1 & 0\\ 1 & 1 & 1\end{pmatrix}\begin{pmatrix}1 & & \\ & 1-4p & \\ & & 1-2p\end{pmatrix}\begin{pmatrix}1 & 1 & -1\\ 1 & -1 & 0\\ 1 & 1 & 1\end{pmatrix}^{-1}$$

$$=\frac{1}{4}\begin{pmatrix}1 & 1 & -1\\ 1 & -1 & 0\\ 1 & 1 & 1\end{pmatrix}\begin{pmatrix}1 & & \\ & 1-4p & \\ & & 1-2p\end{pmatrix}\begin{pmatrix}1 & 2 & 1\\ 1 & -2 & 1\\ -2 & 0 & 2\end{pmatrix}$$

$$\lim_{n\to\infty}\boldsymbol{P}^{(n)}=\lim_{n\to\infty}\boldsymbol{P}^{n}$$

$$=\lim_{n\to\infty}\frac{1}{4}\begin{pmatrix}1 & 1 & -1\\ 1 & -1 & 0\\ 1 & 1 & 1\end{pmatrix}\begin{pmatrix}1 & & \\ & (1-4p)^n & \\ & & (1-2p)^n\end{pmatrix}\begin{pmatrix}1 & 2 & 1\\ 1 & -2 & 1\\ -2 & 0 & 2\end{pmatrix}$$

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

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

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

If $p=\frac{1}{2}$, then

$$\boldsymbol{P}^{(n)}=\boldsymbol{P}^{n}$$

$$=\frac{1}{4}\begin{pmatrix}1 & 1 & -1\\ 1 & -1 & 0\\ 1 & 1 & 1\end{pmatrix}\begin{pmatrix}1 & & \\ & (-1)^n & \\ & & 0^n\end{pmatrix}\begin{pmatrix}1 & 2 & 1\\ 1 & -2 & 1\\ -2 & 0 & 2\end{pmatrix}$$

$$=\frac{1}{4}\begin{pmatrix}1 & 1 & -1\\ 1 & -1 & 0\\ 1 & 1 & 1\end{pmatrix}\begin{pmatrix}1 & & \\ & 0 & \\ & & 0\end{pmatrix}\begin{pmatrix}1 & 2 & 1\\ 1 & -2 & 1\\ -2 & 0 & 2\end{pmatrix}+\frac{(-1)^n}{4}\begin{pmatrix}1 & 1 & -1\\ 1 & -1 & 0\\ 1 & 1 & 1\end{pmatrix}\begin{pmatrix}0 & & \\ & 1 & \\ & & 0\end{pmatrix}\begin{pmatrix}1 & 2 & 1\\ 1 & -2 & 1\\ -2 & 0 & 2\end{pmatrix}$$

$$=\frac{1}{4}\begin{pmatrix}1 & 1 & -1\\ 1 & -1 & 0\\ 1 & 1 & 1\end{pmatrix}\begin{pmatrix}1 & 2 & 1\\ & & \\ & & \end{pmatrix}+\frac{(-1)^n}{4}\begin{pmatrix}1 & 1 & -1\\ 1 & -1 & 0\\ 1 & 1 & 1\end{pmatrix}\begin{pmatrix} & & \\ 1 & -2 & 1\\ & & \end{pmatrix}$$

$$
= \frac{1}{4}\begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \end{pmatrix} + \frac{(-1)^n}{4}\begin{pmatrix} 1 & -2 & 1 \\ -1 & 2 & -1 \\ 1 & -2 & 1 \end{pmatrix}
$$

$$
= \begin{cases} \dfrac{1}{4}\begin{pmatrix} 0 & 4 & 0 \\ 2 & 0 & 2 \\ 0 & 4 & 0 \end{pmatrix} & \text{if } n \text{ is odd} \\ \dfrac{1}{4}\begin{pmatrix} 2 & 0 & 2 \\ 0 & 4 & 0 \\ 2 & 0 & 2 \end{pmatrix} & \text{if } n \text{ is even} \end{cases}.
$$

To fully determine the distribution of a Markov chain at any time n, we need to specify first the distribution of X_0, which is called the initial distribution of the Markov chain. Assume now that the initial distribution (the probability mass function) is given by

$$\alpha_i := P(X_0 = i), \quad i \ge 0,$$

where $\sum_{i \in S} \alpha_i = 1$. By law of total probability,

$$P(X_n = j) = \sum_{i \in S} P(X_n = j \mid X_0 = i) P(X_0 = i) = \sum_{i=0}^{\infty} P_{i,j}^n \alpha_i.$$

Then, the row of probabilities at time n is given by probability,

$$[P(X_n = i),\ i \in S] = [\alpha_0, \alpha_1, \ldots] \cdot \boldsymbol{P}^n.$$

Example 1.5

Consider the random walk on the graph from Example 1.2. Choose a starting vertex at random and equal probability. (a) What is the probability mass function of the chain at time 2? (b) Compute $P(X_2 = 2, X_6 = 3, X_{12} = 4)$.

As

$$P = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix},$$

we have

$$[P(X_2=1) \quad P(X_2=2) \quad P(X_2=3) \quad P(X_2=4)] = [\tfrac{1}{4} \quad \tfrac{1}{4} \quad \tfrac{1}{4} \quad \tfrac{1}{4}] \cdot P^2 = [\tfrac{2}{9} \quad \tfrac{5}{18} \quad \tfrac{2}{9} \quad \tfrac{5}{18}].$$

The probability in (b) equals

$$P(X_2=2) \cdot P_{23}^4 \cdot P_{34}^6 = \frac{8645}{708588} \approx 0.0122.$$

Definition 1.6

State j is said to be **accessible** from state i if $P_{i,j}^n > 0$ for some $n \geq 0$, i.e., starting in i, it is possible that the process will ever enter state j. Two states i and j that are accessible to each other are said to **communicate**, and we write $i \leftrightarrow j$. Note that $i \leftrightarrow i$, since, by convention, $P_{i,j}^0 = 1$ for all $i \in S$.

Theorem 1.7

Communication is an equivalence relation. That is

(i) $i \leftrightarrow i$.

(ii) If $i \leftrightarrow j$, then $j \leftrightarrow i$.

(iii) If $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$.

Two states that communicate are said to be in the same **class**. It is an easy consequence of (i), (ii), and (iii) that any two classes of states are either identical or disjoint. In other words, the concept of communication divides the state space up into a number of separate classes. The Markov chain is said to be **irreducible** if there is only one class, that is, if all states communicate with each other.

Example 1.8

Consider the Markov chain in Example 1.3. The classes of this Markov chain are $\{0\}$,

$\{1,\cdots,N-1\}$, $\{N\}$ if $0 < p < 1$. State 0 is accessible from state 1, but the reverse is not true. The classes of this Markov chain are $\{0\},\{1\},\{2\},...,\{N-1\}$, $\{N\}$ if $p = 0$ or 1.

Definition 1.9

For any states i and j define $f_{i,j}^{n}$ to be the probability that, starting in i, the first transition into j occurs at time n. Formally,

$$f_{i,j}^{0}=0, \quad f_{i,j}^{n}=P(X_n=j, X_k \neq j, k=1,\cdots,n-1 \mid X_0=i) \text{ for } n \geq 1.$$

Let τ_j denote the **hitting time** that the chain reaches state j; i.e.,

$$\tau_j=\begin{cases} \min\{n\geq 1: X_n=j\}, & \text{if } X_n=j \text{ for some } n, \\ \infty, & \text{otherwise.} \end{cases}$$

Then we can also express $f_{i,j}^{n}$ in terms of τ_j as

$$f_{i,j}^{n}=P(\tau_j=n \mid X_0=i) \text{ for } n\geq 0.$$

Note that the last identity is true also for $n=0$, since, by definition, it holds that $\tau_j \geq 1$, so $P(\tau_j=0 \mid X_0=i) =0=f_{i,j}^{0}$.

Let $f_{i,j}=\sum_{n=1}^{\infty} f_{i,j}^{n}$. Then $f_{i,j}$ denote the probability that, starting in state i, the process will ever make a transition into state j. (Note that for $i\neq j$, $f_{i,j}$ is positive if, and only if, j is accessible from i.) State i is said to be **recurrent** if $f_{i,i}=1$ and **transient** if $f_{i,i}<1$.

Theorem 1.10

Using the same notation as above, we have

$$P_{i,j}^n = f_{i,j}^n + f_{i,j}^{n-1}P_{j,j}^1 + \cdots + f_{i,j}^1 P_{j,j}^{n-1} \quad n \geq 1.$$

Let $P_{i,j}(z) = \sum_{n=0}^{\infty} P_{i,j}^n z^n$ be the generating function of number of transition entering state j starting from state i and let $\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$ Then $P_{i,j}(z) = \delta_{i,j} + f_{i,j}(z)P_{j,j}(z)$. In particular,

$$P_{i,i}(z) = \frac{1}{1 - f_{i,i}(z)}, \quad z \in (0,1).$$

Proof

$$\begin{aligned} P_{i,j}(z) &= \sum_{n=0}^{\infty} P_{i,j}^n z^n \\ &= P_{i,j}^0 + \sum_{n=1}^{\infty} (f_{i,j}^n + f_{i,j}^{n-1}P_{j,j}^1 + f_{i,j}^{n-2}P_{j,j}^2 + \cdots + f_{i,j}^1 P_{j,j}^{n-1}) z^n \\ &= \delta_{i,j} + \left(\sum_{n=1}^{\infty} f_{i,j}^n z^n \right) \left(\sum_{n=0}^{\infty} P_{j,j}^n z^n \right) \\ &= \delta_{i,j} + f_{i,j}(z) P_{j,j}(z) \end{aligned}$$

Corollary 1.11

State i is recurrent if $\sum_{n=0}^{\infty} P_{i,i}^n = \infty$, and transient if $\sum_{n=0}^{\infty} P_{i,i}^n < \infty$.

Proof

$$\sum_{n=0}^{\infty} P_{i,i}^n = P_{i,i}(1) = \frac{1}{1 - f_{i,i}(1)} \begin{cases} = \infty & \text{if state } i \text{ is recurrent} \\ < \infty & \text{if state } i \text{ is transient} \end{cases}$$

Corollary 1.11 shows that a transient state will only be visited a finite number of times (hence the name transient), while a recurrent state will be visited infinitely many times.

We call a set S_0 of states closed if $P_{i,j} = 0$ for each $i \in S_0$ and $j \notin S_0$. In plain language, once entered, a closed set cannot be exited.

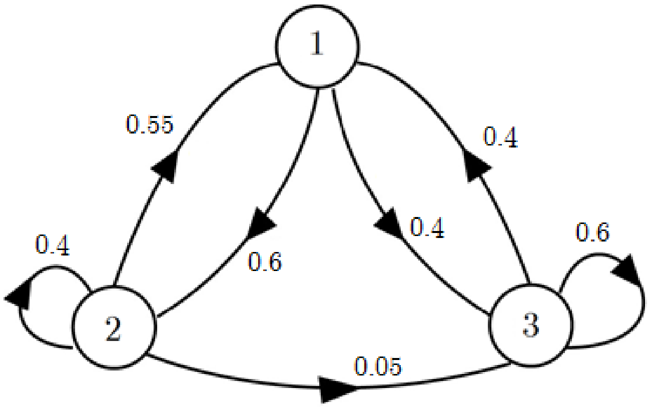
Proposition 1.12

If a closed set S_0 has only finitely many states, then there must be at least one recurrent state. In particular, any finite Markov chain must contain at least one recurrent state.

Proof. Start from any state from S_0. By definition, the chain stays in S_0 forever. If all states in S_0 are transient, then each of them is visited either not at all or only finitely many times. This is impossible.

Example 1.13

Consider the following Markov chain.



Characteristic polynomial of $\boldsymbol{P}$ is

$$\det(\lambda I_3 - \boldsymbol{P}) = \det\begin{pmatrix} \lambda & -0.6 & -0.4 \\ -0.55 & \lambda - 0.4 & -0.05 \\ -0.4 & 0 & \lambda - 0.6 \end{pmatrix}$$
$$= \lambda^3 - \lambda^2 - 0.25\lambda + 0.25$$
$$= (\lambda - 1)\left(\lambda^2 - 0.25\right) \quad [1 \text{ is an eigenvalue}]$$
$$= (\lambda - 1)(\lambda - 0.5)(\lambda + 0.5)$$

Its eigenvalues are $\lambda_1 = 1, \lambda_2 = 0.5$ and $\lambda_3 = -0.5$.

For $\lambda_1 = 1$, the corresponding eigenvector is $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

For $\lambda_2 = 0.5$, consider $N(0.5I_3 - \boldsymbol{P}) = N\begin{pmatrix} 0.5 & -0.6 & -0.4 \\ -0.55 & 0.1 & -0.05 \\ -0.4 & 0 & -0.1 \end{pmatrix} = span\left\{\begin{pmatrix} 2 \\ 7 \\ -8 \end{pmatrix}\right\}$. The corresponding eigenvector is $\begin{pmatrix} 2 \\ 7 \\ -8 \end{pmatrix}$.

For $\lambda_3 = -0.5$, consider $N(-0.5I_3 - \boldsymbol{P}) = N\begin{pmatrix} -0.5 & -0.6 & -0.4 \\ -0.55 & -0.9 & -0.05 \\ -0.4 & 0 & -1.1 \end{pmatrix} = span\left\{\begin{pmatrix} 22 \\ -13 \\ -8 \end{pmatrix}\right\}$. The corresponding eigenvector is $\begin{pmatrix} 22 \\ -13 \\ -8 \end{pmatrix}$.

Then

$$\boldsymbol{P} = \begin{pmatrix} 0 & 0.6 & 0.4 \\ 0.55 & 0.4 & 0.05 \\ 0.4 & 0 & 0.6 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & -0.5 \end{pmatrix}\begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix}^{-1}$$

$$\boldsymbol{P}^n = \begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & -0.5 \end{pmatrix}^n\begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix}^{-1}$$
$$= \frac{1}{100}\begin{pmatrix} 1 & 2 & 22 \\ 1 & 7 & -13 \\ 1 & -8 & -8 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5^n & 0 \\ 0 & 0 & (-0.5)^n \end{pmatrix}\begin{pmatrix} 32 & 32 & 36 \\ 1 & 6 & -7 \\ 3 & -2 & -1 \end{pmatrix}$$

$$P_{1,1}^n = \frac{1}{100}\begin{pmatrix} 1 & 2 & 22 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5^n & 0 \\ 0 & 0 & (-0.5)^n \end{pmatrix}\begin{pmatrix} 32 \\ 1 \\ 3 \end{pmatrix} = \frac{1}{100}\left(32 + 2(0.5^n) + 66(-0.5)^n\right)$$

$$\sum_{n=0}^{\infty} P_{1,1}^n = \sum_{n=0}^{\infty}\left(0.32 + 0.02(0.5^n) + 0.66(-0.5)^n\right)$$
$$= 0.32\sum_{n=0}^{\infty} 1 + 0.02\sum_{n=0}^{\infty} 0.5^n + 0.66\sum_{n=0}^{\infty}(-0.5)^n$$
$$= 0.32\sum_{n=0}^{\infty} 1 + \frac{0.02}{1-0.5} + \frac{0.66}{1+0.5} = \infty.$$

So state 1 is recurrent according to Corollary 1.11.

Theorem 1.14

If state i is recurrent and $i \to j$, then state j is recurrent and $f_{j,i} = 1$. Moreover, recurrence (transience) is a class property.

Theorem 1.15

Consider two states i and j. If i is recurrent and j is transient, then $f_{i,j} = 0$.

Example 1.16

Consider the Markov chain in Example 1.3 for $N = 3$.

$$f_{0,0} = f_{0,0}^1 = f_{3,3} = f_{3,3}^1 = 1, \quad f_{1,1} = f_{1,1}^2 = f_{2,2} = f_{2,2}^2 = p(1-p) < 1.$$

States 0 and 3 are recurrent and states 1 and 2 are transient. On the other hand, $P_{0,0}^n = P_{3,3}^n = 1$ for all $n \geq 0$. Hence $\sum_{n=1}^{\infty} P_{0,0}^n = \sum_{n=1}^{\infty} P_{3,3}^n = \infty$.

$$P_{1,1}^n = P_{2,2}^n = \begin{cases} p^{n/2}(1-p)^{n/2} & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases} \quad \Rightarrow \quad \sum_{n=1}^{\infty} P_{1,1}^n = \sum_{n=1}^{\infty} P_{2,2}^n = \frac{p(1-p)}{1-p(1-p)} < \infty.$$

By Definition 1.9 and Corollary 1.11, states 0 and 3 are recurrent and states 1 and 2 are transient.

Clearly, $f_{0,1} = 0$. This is in line with Corollary 1.11 and Theorem 1.15.

Example 1.17

Consider Example 1.13. Note that we have shown that state 1 is recurrent. It follows from Theorem 1.14 that states 2 and 3 are also recurrent, since $1 \to 2$ and $1 \to 3$.

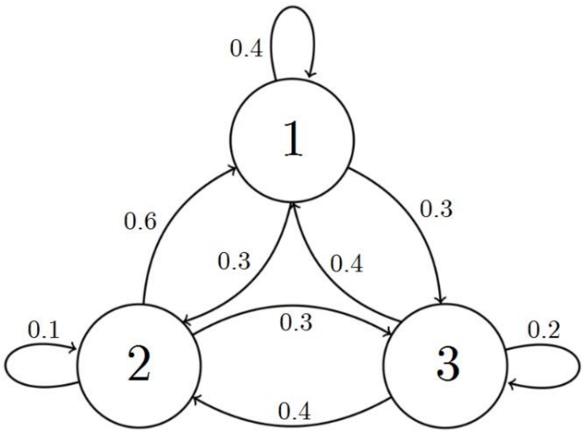
Remark: For finite Markov chains, we don't really need to bother to do complicated computations to determine recurrence and transience of states. In fact, for finite Markov chains, the recurrence of a communication class is equivalent to its closedness.

Theorem 1.18

Suppose a Markov chain has only finitely many states. Consider a state i and let $H_i = \{j \in S : i \leftrightarrow j\}$ be the communication class corresponding to i. Then H_i is a recurrent class iff H_i is closed.

Example 1.19

Consider the following Markov Chain:



Clearly, $f_{1,1}^1 = 0.4$, $f_{1,1}^{n+1} = 0.3 f_{2,1}^n + 0.3 f_{3,1}^n$, $n \geq 1$. We have $f_{1,1}(z) = 0.4z + 0.3z f_{2,1}(z) + 0.3z f_{3,1}(z)$.

Similarly,

$$f_{2,1}(z) = 0.6z + 0.1z f_{2,1}(z) + 0.3z f_{3,1}(z) \quad \text{and} \quad f_{3,1}(z) = 0.4z + 0.4z f_{2,1}(z) + 0.2z f_{3,1}(z)$$

$$\begin{pmatrix} 1-0.1z & -0.3z \\ -0.4z & 1-0.2z \end{pmatrix}\begin{pmatrix} f_{2,1}(z) \\ f_{3,1}(z) \end{pmatrix} = \begin{pmatrix} 0.6z \\ 0.4z \end{pmatrix}$$

$$\begin{aligned}
\begin{pmatrix} f_{2,1}(z) \\ f_{3,1}(z) \end{pmatrix} &= \begin{pmatrix} 1-0.1z & -0.3z \\ -0.4z & 1-0.2z \end{pmatrix}^{-1}\begin{pmatrix} 0.6z \\ 0.4z \end{pmatrix} \\
&= \frac{1}{(1-0.1z)(1-0.2z)-0.12z^2}\begin{pmatrix} 1-0.2z & 0.3z \\ 0.4z & 1-0.1z \end{pmatrix}\begin{pmatrix} 0.6z \\ 0.4z \end{pmatrix} \\
&= \frac{1}{1-0.3z-0.1z^2}\begin{pmatrix} 0.6z \\ 0.4z+0.2z^2 \end{pmatrix} \\
&= \frac{1}{(1-0.5z)(1+0.2z)}\begin{pmatrix} 0.6z \\ 0.4z+0.2z^2 \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
f_{1,1}(z) &= 0.4z + 0.3zf_{2,1}(z) + 0.3zf_{3,1}(z) \\
&= 0.4z + 0.3z\frac{0.6z}{(1-0.5z)(1+0.2z)} + 0.3z\frac{0.4z+0.2z^2}{(1-0.5z)(1+0.2z)} \\
&= 0.4z + \frac{0.3z^2+0.06z^3}{(1-0.5z)(1+0.2z)} = 0.4z + \frac{0.3z^2}{1-0.5z} \\
&= 0.4z + 0.3z^2 + (0.3)0.5z^3 + \cdots + (0.3)0.5^{n-2}z^n + \cdots
\end{aligned}$$

Definition 1.20

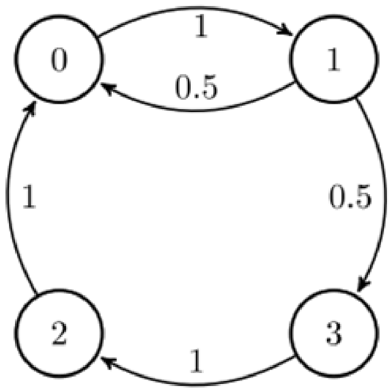
State i is said to have **period** d if $d = \gcd\{n \ge 1 : P_{i,i}^n > 0\}$. If $P_{i,i}^n = 0$ for all $n > 0$, then period of i is ∞. Let $d(i)$ denote the period of i. A state with period 1 is said to be **aperiodic**. In particular, any absorbing state is both aperiodic and recurrent.

Theorem 1.21

If $i \leftrightarrow j$, then $d(i) = d(j)$. That is, if state i has period d, and states i and j communicate, then state j also has period d.

Example 1.22

(a) Consider the following Markov chain:

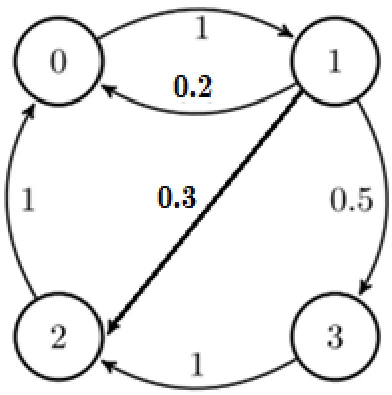


Here we have

$$\{n \geq 1 : P_{0,0}^n > 0\} = \{2, 4, 6, 8, 10, \cdots\},$$
$$\{n \geq 1 : P_{1,1}^n > 0\} = \{2, 4, 6, 8, 10, \cdots\},$$
$$\{n \geq 1 : P_{2,2}^n > 0\} = \{4, 6, 8, 10, 12, \cdots\},$$
$$\{n \geq 1 : P_{3,3}^n > 0\} = \{4, 6, 8, 10, 12, \cdots\}.$$

Hence all states have period 2.

(b) Next, consider the modification of (a)



Here the chain is aperiodic since we have

$$\{n \geq 1 : P_{0,0}^n > 0\} = \{2, 3, 4, 5, 6, 7, \cdots\},$$
$$\{n \geq 1 : P_{1,1}^n > 0\} = \{2, 3, 4, 5, 6, 7, \cdots\},$$
$$\{n \geq 1 : P_{2,2}^n > 0\} = \{3, 4, 5, 6, 7, 8, \cdots\},$$
$$\{n \geq 1 : P_{3,3}^n > 0\} = \{4, 6, 7, 8, 9, 10, \cdots\}$$

Hence all states have period 1.

Definition 1.23

Starting at a recurrent state *i*, i.e., $\sum_{n=1}^{\infty} f_{i,i}^{n} = 1$. Let $\mu_{i,i}$ denote the expected number of transitions needed to return state *i*. In other words, $\mu_{i,i}$ is the expected hitting time of state *i*, started from *i*, and $\mu_{i,i} = \sum_{n=1}^{\infty} n f_{i,i}^{n} = f_{i,i}'(1)$. If the expected hitting time is finite then this is called **positive-recurrent**; if the expected hitting time is infinite then this is called **null-recurrent**. Positive recurrent, aperiodic states are called **ergodic**.

Example 1.24

Consider the Markov chain in Example 1.19.

$$\frac{d}{dz} f_{1,1}(z) = \frac{d}{dz}\left(0.4z + \frac{0.3z^2}{1-0.5z}\right) = 0.4 + \frac{0.6z(1-0.5z) - 0.3z^2(-0.5)}{(1-0.5z)^2} = 0.4 + \frac{0.6z - 0.15z^2}{(1-0.5z)^2}$$

$$\mu_{1,1} = \sum_{n=1}^{\infty} n f_{1,1}^{n} = \left.\frac{d}{dz} f_{1,1}(z)\right|_{z=1} = 0.4 + \frac{0.6-0.15}{(1-0.5)^2} = 2.2$$

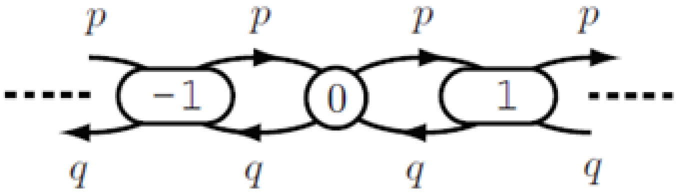
The expected time of the first return is finite, so state 1 is positive recurrent. In fact, one can show that in a finite-state Markov chain, all recurrent states are positive recurrent.

Example 1.25 (1-dimensional Simple Random Walk)

Consider a Markov chain whose state space consists of the integers $i = 0, \pm 1, \pm 2, \cdots,$ and have transition probabilities given by

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \quad i = 0, \pm 1, \pm 2, \cdots$$

where $0 < p < 1$. In other words, on each transition the process either moves one step to the right (with probability *p*) or one step to the left (with probability $q = 1-p$). When $p = 1/2$, the preceding process is called a **symmetric random walk**. It can represent the winnings of a gambler who on each play of the game either wins or loses one dollar.



Since all states clearly communicate, it follows from Theorem 1.14 that they are either all transient or all recurrent. So let us consider state 0.

Fix an integer *m*, and let τ_m denote the hitting time of the state *m*; i.e.,

$$\tau_m = \min\{n \geq 1 : X_n = m\},$$

and if the random walk never reaches the state *m*, we define τ_m to be infinity.

Define $p_n = P(\tau_1 = n) = f_{0,1}^n$. Clearly, $p_{2n} = 0, p_1 = p$. Let $f_{0,1}(z) = p_0 + p_1 z + p_2 z^2 + \cdots$ be the probability generating function of τ_1. Then we have

$$p_n = q(p_1 p_{n-2} + p_2 p_{n-3} + \cdots + p_{n-2} p_1) \text{ for } n \geq 3.$$

Then

$$f_{0,1}(z) = pz + (qz) f_{0,1}(z)^2.$$

We obtain

$$f_{0,1}(z) = \frac{1 - \sqrt{1 - 4pqz^2}}{2qz}.$$

Similarly, define $q_n = P(\tau_{-1} = n) = f_{0,-1}^n$. Let $f_{0,-1}(z) = q_0 + q_1 z + q_2 z^2 + \cdots$ be the probability generating function of τ_{-1}. We obtain

$$f_{0,-1}(z) = \frac{1 - \sqrt{1 - 4pqz^2}}{2pz}.$$

Let $f_{0,0}(z) = f_{0,0}^0 + f_{0,0}^1 z + f_{0,0}^2 z^2 + \cdots$ be the probability generating function. Then

$$f_{0,0}(z)=(pz)f_{1,0}(z)+(qz)f_{-1,0}(z)=(pz)f_{0,-1}(z)+(qz)f_{0,1}(z)=1-\sqrt{1-4pqz^2},$$
$$f_{0,0}=f_{0,0}(1)=1-\sqrt{1-4pq}=1-\sqrt{(p+q)^2-4pq}=1-\sqrt{(p-q)^2}=1-|p-q|$$

and

$$\mu_{0,0}=\sum_{n=1}^{\infty}nf_{0,0}^{n}=f'_{0,0}(1)=\frac{d}{dz}\left(1-\sqrt{1-z^2}\right)_{z=1}=\left.\frac{z}{\sqrt{1-z^2}}\right|_{z=1}=\infty \text{ if } p=q.$$

Theorem 1.26

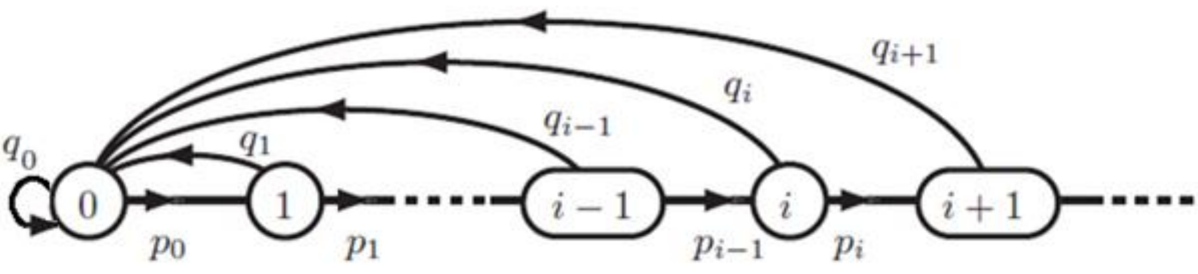
If the random walk is non-symmetric, that is $p\neq q$, then $f_{0,0}=\min\{2p,2q\}<1$. It is transient. If the random walk is symmetric, that is $p=q=1/2$, then $f_{0,0}=1$. It is null-recurrent.

Theorem 1.27

Positive (null) recurrence is a class property.

Example 1.28

Dr. Wong plays video game, starting from level 0. This game has countable infinitely many levels. He needs to finish one by one in ascending order and cannot skip any level.



The probability for levelling up from level i to level $i+1$ is p_i and $q_i=1-p_i$. However, there is no save button. If he cannot level up, then the game will go back to level 0.

(a) Let $u_n = \prod_{i=0}^{n} p_i$, the probability that Dr. Wong levels up consecutively $n+1$ times. Express $f_{0,0}^n$ in terms of u_n. Hence show that level 0 is recurrent if and only if $\prod_{i=0}^{\infty} p_i = 0$.

(b) Show that $f_{0,0}(z)-1=(z-1)(1+p_0z+p_0p_1z^2+\cdots)$. Hence show that $\mu_{0,0}=1+p_0+p_0p_1+\cdots$. Furthermore show that level 0 is positive recurrent if $\lim_{n\to\infty} p_n < 1$.

(c) Determine which of the following is transient or recurrent. If it is recurrent, calculate $\mu_{0,0}$. Is it null recurrent or positive recurrent?

(i) $q_n = 0.1$

(ii) $q_n = \dfrac{1}{n+2}$

(iii) $q_n = \dfrac{1}{(n+2)^2}$.

Solution

(a) $f_{0,0}^1 = q_0 = 1-p_0 = 1-u_0$

$$f_{0,0}^n = \left(\prod_{i=0}^{n-2} p_i\right) q_{n-1} = \left(\prod_{i=0}^{n-2} p_i\right)(1-p_{n-1}) = \left(\prod_{i=0}^{n-2} p_i\right) - \left(\prod_{i=0}^{n-1} p_i\right) = u_{n-2}-u_{n-1} \quad \text{for } n>1$$

$$\sum_{n=1}^{m+1} f_{0,0}^n = f_{0,0}^1 + \sum_{n=2}^{m+1} f_{0,0}^n = (1-u_0)+\sum_{n=2}^{m+1}(u_{n-2}-u_{n-1}) = (1-u_0)+(u_0-u_1)+\cdots+(u_{m-1}-u_m) = 1-u_m$$

Level 0 is recurrent if and only if $\sum_{n=1}^{\infty} f_{0,0}^n = 1$, i.e, $\lim_{m\to\infty} u_m = 0$.

(c)
$$\begin{aligned} f_{0,0}(z) &= f_{0,0}^1 z + f_{0,0}^2 z^2 + \cdots + f_{0,0}^n z^n + \cdots \\ &= (1-p_0)z + (p_0 - p_0p_1)z^2 + (p_0p_1 - p_0p_1p_2)z^3 + \cdots \\ &= z + \left(p_0z^2 + p_0p_1z^3 + p_0p_1p_2z^4 + \cdots\right) - \left(p_0z + p_0p_1z^2 + p_0p_1p_2z^3 + \cdots\right) \\ &= z + (z-1)\left(p_0z + p_0p_1z^2 + p_0p_1p_2z^3 + \cdots\right) \end{aligned}$$

$$f_{0,0}(z)-1 = (z-1)+(z-1)\left(p_0z+p_0p_1z^2+p_0p_1p_2z^3+\cdots\right) = (z-1)(1+p_0z+p_0p_1z^2+\cdots).$$

$$\mu_{0,0} = f_{0,0}'(1) = 1+p_0+p_0p_1+\cdots$$

Since $\lim_{n\to\infty} \frac{p_0 \cdots p_{n-1} p_n}{p_0 \cdots p_{n-1}} = \lim_{n\to\infty} p_n < 1$, by Ratio test, $\mu_{0,0}$ is finite and level 0 is positive recurrent.

(c) (i) $\lim_{m\to\infty} u_m = \lim_{m\to\infty} \prod_{i=0}^{m} p_i = \lim_{m\to\infty} (0.9)^{m+1} = 0.$ It is recurrent.

$\mu_{0,0} = 1 + u_0 + u_1 + u_2 + \cdots = 1 + 0.9 + 0.9^2 + 0.9^3 + \cdots = \frac{1}{1-0.9} = 10.$ Level 0 is positive recurrent. Clearly, this Markov chain is irreducible. Since positive recurrence is a class property, all other levels are positive recurrent.

(ii) $\lim_{m\to\infty} u_m = \lim_{m\to\infty} \prod_{i=0}^{m} p_i = \lim_{m\to\infty} \prod_{i=0}^{m} \left(1 - \frac{1}{i+2}\right) = \lim_{m\to\infty} \prod_{i=0}^{m} \left(\frac{i+1}{i+2}\right) = \lim_{m\to\infty} \frac{1}{m+2} = 0.$ It is recurrent.

$\mu_{0,0} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots = \infty.$ Level 0 is null recurrent. Clearly, this Markov chain is irreducible. Since null recurrence is a class property, all other levels are null recurrent.

(iii)
$$\begin{aligned}
\lim_{m\to\infty} u_m &= \lim_{m\to\infty} \prod_{i=0}^{m} p_i \\
&= \lim_{m\to\infty} \prod_{i=0}^{m} \left(1 - \frac{1}{(i+2)^2}\right) \\
&= \lim_{m\to\infty} \prod_{i=0}^{m} \left(1 - \frac{1}{i+2}\right)\left(1 + \frac{1}{i+2}\right) \\
&= \lim_{m\to\infty} \prod_{i=0}^{m} \left(\frac{i+1}{i+2}\right)\left(\frac{i+3}{i+2}\right) \\
&= \lim_{m\to\infty} \frac{m+3}{2(m+2)} = \frac{1}{2}.
\end{aligned}$$

Level 0 is transient. Clearly, this Markov chain is irreducible. Since transience is a class property, all other levels are transient.