Chapter 4: Markov Chains

Markov chains and processes are fundamental modeling tools in applications. The reason for their use is that they natural ways of introducing dependence in a stochastic process and thus more general. Moreover the analysis of these processes is often very tractable. But perhaps an overwhelming importance of such processes is that they can quite accurately model a wide variety of physical phenomena. They play an essential role in modeling telecommunication systems, service systems, and even signal processing applications. In this chapter we will focus on the discrete-time, discrete-valued case, that leads to the appellation Markov chains.

1 Introduction and preliminaries

We restrict ourselves to the discrete-time case. Markov chains (M.C) can be seen first attempt to impose a structure of dependence in a sequence of random variables that is rich enough to model many observed phenomena and yet leads to a tractable structure from which we can perform calculations. Supose we had a sequence of r.v.'s $\{X_i\}$ and we know say X_5, if the $X_i's$ are independent then this information would say nothing about a future value , say X_{10}, other than the a priori assumptions that we have on their distribution. O the other hand if they were dependent, unless we precisely specify how the probability distributions at one time depend on the distributions at other times there is very little we could do because we know to specify a stochastic process we need to specify the joint distributions. Markov chains (or Markov processes in general) are stochastic processes whose future evolution depends only on its current value and not how it reached there. We formalize this idea below.

Definition 1.1. *Let $\{X_n\}$ be a discrete-line stochastic process which takes its values in a space E. Let $A \subset E$. If*

$$\mathbb{P}\{X_{n+1} \in A|X_o, X_1, \dots X_n\} = P\{X_{n+1} \in A \mid X_n\}$$

then $\{X_n\}$ is said to be a discrete-time Markov process.
More generally

$$\mathbb{P}\{X_{n+1} \in A \,|\mathcal{F}_n^X\} = \mathbb{P}\{X_{n+1} \in A \,|X_n\}$$

where $\mathcal{F}_n^X = \sigma\{X_u, u \leq n\}$ the sigma-field of all events generated by the process $\{X_k\}$ up to n.

When

$$E = \{0,\ 1,\ \dots,\ \}$$

i.e., a countable set then $\{X_n\}$ is said to be a Markov chain.

From now on we will always assume E to be a finite or countable (discrete) set. E is said to be the state-space of the Markov chain.

From the definition of a Markov chain it is easy to see that if

$$A \subset \{X_0 \dots X_{n-1}\} \quad , \quad B \subset \{X_{n+1}, \dots\}$$

then

$$\mathbb{P}(A \cap B|X_n) = \mathbb{P}(A|X_n)\ \mathbb{P}\ (B|X_n)$$

Let denote $\mathcal{F}_n = \sigma\{X_k, k \leq n\}$ and $\overline{\mathcal{F}}_n = \sigma\{X_k, K > n\}$ Then more generally if $A \in \mathcal{F}_m$, $B \in \overline{\mathcal{F}}_p$ and $m < n < p$ then:

$$\mathbb{P}(A \bigcap B|\sigma(x_n)) = \mathbb{P}(A|\sigma(X_n))\mathbb{P}(B|\sigma(X_n))$$

In other words, for any $m \leq n-1$, $p \geq n+1$

$$\mathbb{P}\{X_m = i, X_p = j|X_n = k\} = \mathbb{P}\{X_m = i|X_n = k\}\ \mathbb{P}\{X_p = j|X_n = k\}$$

i.e., if $\{X_n\}$ is a Markov chain then the future (represented by the process at times > n) and the past (represented by the process at times < n) are conditionally independent given the process at time n given by X_n. Conditional independence is actually a better way of defining the Markov property since it extends readily to the case when the index set is not necessarily the set of integers but of higher dimension.

From Chapter 1., we know that if we define events $A_n = \{X_n \leq a_n\}$, then if $\{A_n\}$'s are Markovian then the $\mathbb{P}(A_n)$ is determined from the knowledge of $\mathbb{P}(A_0)$ and the conditional probabilities $\mathbb{P}(A_{k+1}|A_k)$. Thus if $\{X_n\}$ is Markov, what we really mean is that the events generated by $\{X_n\}$ have the Markovian property, which is equivalent to the distribution at any time is completely determined by its initial distribution $\pi^{(0)}(i) = \mathbb{P}\{X_0 = i\}$ and the conditional distributions $\mathbb{P}\{X_{k+1} = j|X_k = i\}$ for $k = 1,\ 2,\ \ldots\ n-1$.

The conditional probability

$$P\{X_{k+1} = j|X_k = i\} = P_{ij}(k)$$

is called the transition probability of the Markov chain.

If $P_{ij}(k)$ does not depend on the time $\{k\}$ then we say that the Markov chain is time-homogeneous or simply homogeneous.

For example

$$P\{x_{k+1}, = j/x_k = i\} = P_{ij}$$

Example 1: Consider the following Markov Chain defined as $E = \{0,\ 1,\ 2\}$ with

$$\mathbf{P} = \{P_{ij}\} = \begin{pmatrix} 1 & 0 & 0 \\ 0.5 & 0 & 0.5 \\ \frac{2}{3} & 0 & \frac{1}{3} \end{pmatrix}$$

We can pictorially represent the chain as follows:

In words: If at time 0, the process starts in state 0 then it will stay in state 0 for all time since the transition probability $P_{0j} = 0$ if $j \neq 0$.

If the process starts in state 1 then with probability 1/2 it will be in either (0) or (2) at the next instant. In other words, if we observe the chain for a long time the state 1 will only be observed if the chain starts there, and over the long run it will be in state 0. Since once it goes to 0 it stays there.

In the sequel we will see that the states $\{0\}$, $\{1\}$ and $\{2\}$ have some special properties. We will focus our attention on studying the long-run behavior of Markov chains. In fact, it will be seen that the entire structure is governed by P the matrix of the transition probabilities.

In the following we will use the following notation.

We will denote a Markov chain by (E, P) where E denotes the state space and P the transition probability matrix.

Let us now introduce some rotation which will be used throughout our study of Markov chains.

$$P_{ij} = P\{X_{k+1} = j | X_k = i\}$$

$P_{ij}^{(k)} = P\{X_k = j | X_0 = i\}$, i.e., the conditional probability that the chain is in state j after k-transitions given that it starts in state i.

$$\pi_j^{(k)} = P\{X_k = j\}$$

and

$$\mathbf{P} = \{P_{ij}\}_{i,\ j\ \in\ E}$$

We now state the first fundamental result which is obeyed by the conditional probabilities of a Markov chain. The equation is called the Chapman Kolmogorov equation. We saw (in Chapter 2) that any Markov chain must obey this equation.

Theorem 1.1. *(Chapman-Kolmogorov Equations)*

$$P_{ij}^{(k+l)} = \sum_{\alpha\ \in\ E} P_{i\alpha}^{(k)}\ P_{\alpha j}^{(l)}$$

Proof

$$\begin{aligned}
\mathbb{P}\{X_{k+l} = j / X_0 = i\} &= \sum_{\alpha} \mathbb{P}\{X_{k+1} = j,\ X_k = \alpha | X_0 = i\} \\
\text{(Conditional Probabilities)} &= \sum_{\alpha \in E} \mathbb{P}\{X_{k+1} = j | X_k = \alpha,\ X_0 = i\} \mathbb{P}\{X_k = \alpha | X_0 = i\} \\
\text{(Markov Property)} &= \sum_{\alpha \in E} \mathbb{P}\{X_{k+1} = j | X_{k\ =\ \alpha}\} \mathbb{P}\{X_{k\ =\ \alpha} | X_0 = i\} \\
&= \sum_{\alpha \in E} P_{i\alpha}^{(k)} P_{\alpha j}^{(l)}
\end{aligned}$$

In the above proof we used the fact that the chain is homogeneous. Another way of stating this result is in matrix notation.

Note that by definition

$$P_{ij}^n = (P^n)_{ij}$$

Hence

$$\mathbf{P}^{k+l} = \mathbf{P}^k\mathbf{P}^l .$$

There are two sub-cases of the Chapman-Kolmogorov equation that are important.

$$P_{ij}^{(k+1)} = \sum_{\alpha\in E} P_{i\alpha}P_{\alpha j}^{(k)} \quad : \text{ Backward equation}$$

$$P_{ij}^{(k+1)} = \sum_{\alpha\in E} P_{i\alpha}^{(k)} P_{\alpha j} \quad : \text{ Forward equation}$$

What they state is that to reach the state j after $(k+1)$ steps, the chain starts in state i and goes to state α after the first transition and then goes to j from state α after another k transitions or vice versa.

In a similar vein,

$$\pi_j^{(k+1)} = \sum_{\alpha\in E} \pi_\alpha P_{\alpha j}^{(k)}$$

$$\pi_j^{(k+i)} = \sum_{\alpha} \pi_\alpha^{(k)} P_{\alpha j}^{(i)}$$

Note that by definition of the transition probabilities

$$\sum_{j\in E} P_{ij}^{(n)} = 1 \quad \text{for all n.}$$

Example 2: Consider a homogeneous M.C. $(E\ P)$

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix}$$

Then it is easy to calculate the powers of P as

$$\mathbf{P}^2 = \begin{pmatrix} P_{00}^2 + P_{01}P_{10} & P_{01}(P_{00} + P_{11}) \\ & \\ P_{10}(P_{00} + P_{11}) & P_{11}^2 + P_{01}P_{10} \end{pmatrix}$$

and by induction

$$\mathbf{P}^n = \left(\frac{1}{2 - P_{00} - P_{11}}\right)\begin{pmatrix} 1 - P_{11} & 1 - P_{00} \\ 1 - P_{11} & 1 - P_{00} \end{pmatrix} + \frac{(P_{00} + P_{11} - 1)^n}{2 - P_{00} - P_{11}}\begin{pmatrix} 1 - P_{00} & -(1 - P_{00}) \\ -(1 - P_{11}) & 1 - P_{11} \end{pmatrix}$$

under the hypothesis $|1 - P_{00} - P_{11}| < 1$.

Hence if P_{00} , P_{11} $\neq$ 1 then the hypothesis is always satisfied and

$$\lim_{n\to\infty} P^n = \frac{1}{2-P_{00}-P_{11}} \begin{pmatrix} 1-P_{11} & 1-P_{00} \\ 1-P_{11} & 1-P_{00} \end{pmatrix}$$

which means

$$\lim_{n\to\infty} P\{X_n = 1|X_0=0\} = \frac{1-P_{00}}{2-P_{00}-P_{00}} = P\{X_n = 1/X_0 = 1\}$$

etc. or the chain has the same limiting probabilities irrespective of the initial state i.e., it "forgets" which state it started in.

Another interesting property can also be seen:

Note that $P^* = \lim_{n\to\alpha} P^n$ has columns with identical elements. Also $P^* = PP* = P^*P$. The elements of the columns of P^* are so-called stationary probabilities of the chain that we will study in detail.

Definition 1.2. *The vector $\pi = \{\pi_i\}_{i\in E}$ is said to be the stationary (or invariant) distribution of the chain if*

$$\pi = \pi P$$

or

$$\pi_j = \sum_{i\in E} \pi_i P_{ij}$$

The reason that the vector $\{\pi\}$ is called the stationary distribution can be seen from the following.

Suppose we start the chain with initial distribution

$$\pi_i^0 = \pi_i = P\{x_0 = i\}.$$

Then from the Chapman - Kolmogorov equation

$$\pi_j^{(n+1)} = \sum_i \pi_i P_{ij}^{(n)}$$

or in matrix form

$$\left(\pi^{(n+1)}\right)_j = (\pi P^n)_j$$

$$= \left(\pi P^{n-1}\right)_j = \cdots = (\pi P)_j = \pi_j$$

in other words the probability of being in a given state j at any time remains the same implying that the distribution is time-invariant, or the process $\{X_n\}$ is stationary.

To show that $\{X_n\}$ is stationary if it is started with an initial distribution π we need to show the following property:

$$\mathbb{P}(X_{m_1} = i_1, X_{m_2} = i_2, \cdots, X_{m_p} = i_p) = \mathbb{P}(X_{N+m_1} = i_1, X_{N+m_2} = i_2, \cdots, X_{N+m_p} = i_p)$$

for all integers N, p, $m_1, m_2, \ldots, m_p$ and $i_1, i_2, \ldots, i_p$. This follows readily from the fact that the chain is homogeneous and $\mathbb{P}(X_n = i_1) = \pi_{i_1}$ for all n from above. Indeed by the multiplication rule we have both the lhs and the rhs given by: $\pi_{i_1} P_{i_1 i_2}^{(m_2-m_1)} \cdots P_{i_{p-1} i_p}^{(m_p-m_{p-1})}$ showing that the process is stationary.

Through the examples we have considered we have already seen two important aspects of Markov chains: how the structure of the matrix $\mathbf{P}$ determines the behavior of the chain both from the time evolution as well as the existence of stationary distributions.

We will now study these issues in greater generality.

2 Markov Chains - Finite state space

Let us first begin by considering the finite-state case. These are referred to as finite-state Markov chains. Here $|E| < \infty$ (cardinality of the state space is finite or the chain can only take a finite number of values).

Theorem 2.1. *(Ergodicity of Markov chain, $|E| < \infty$)*

Let (E, P) denote a finite state Markov chain. Let $|E| = N + 1$.

1. *If*

$$\exists n_0 \quad s.t. \quad \min_{i,j} \; P_{ij}^{(n_0)} > 0$$

then,

$$\exists \; (\pi_0, \; \pi_j \; , \; \ldots \; \pi_N) \; , \; s.t. \; \pi_i > 0, \qquad \sum_{i=0}^{N} \pi_i \; = \; 1$$

and

$$\lim_{n \to \infty} \; P_{ij}^{(n)} \; \to \; \pi_j \; \forall \; i \in E.$$

2. *Conversely if $\exists \; \pi_i$ satisfying the properties in (a) then*

$$\exists \; n_0 \; s.t. \; \min_{i,j} \; P_{ij}^{(n_0)} \; > \; 0$$

3.

$$\pi_j \; = \; \sum_{k \; = \; 0}^{N} \; \pi_k \; P_{kj}$$

Proof: Let

$$m_j^{(n)} = \min_i \ P_{ij}^{(n)} \qquad \text{and}$$

$$M_j^{(n)} = \max_i \ P_{ij}^{(n)}.$$

By definition

$$m_j^{(n)} \ \leq \ M_j^{(n)}.$$

Since

$$P_{ij}^{(n+1)} \ = \ \sum_{\alpha} \ P_{i\alpha} \ P_{\alpha j}^{(n)}$$

we have

$$m_j^{(n+1)} \ \geq \ m_j^{(n)} \quad \text{and} \quad M_j^{(n+1)} \ \leq \ M_j^{(n)}.$$

Since

$$\begin{aligned} m_j^{(n+1)} \ = \ \min_i \ P_{ij}^{(n+1)} \ &= \\ &\geq \ \textstyle\sum_\alpha \ P_{i\alpha} \ \min_\alpha \ P_{\alpha j}^{(n)} \\ &= \ m_j^{(n)} \end{aligned}$$

hence, $m_j^{(n+1)} \ \geq \ m_j^{(n)}$ and the result $M_j^{(n+1)} \ \geq \ M_j^{(n)}$ follows similarly. This implies that $m_j^{(n)}$ is a monotone non-decreasing sequence and $M_j^{(n)}$ is a monotone non-increasing sequence.

Noting that $m_j^{(n)} \ \leq \ P_{ij}^{(n)}$, if we show that $M_j^{(n)} \ - \ m_j^{(n)} \ \rightarrow \ 0 \infty \ n \ \rightarrow \ \Delta$ then it will imply that $\lim_{n \ \rightarrow \ \infty} \ P_{ij}^{(n)}$ will exist.

Let

$$\varepsilon \ = \ \min_{i,j} \ P_{ij}^{(n_0)} \ > 0.$$

Then

$$\begin{aligned} P_{ij}^{(n_0+n)} \ &= \textstyle\sum_\alpha \ P_{i\alpha}^{(n_0)} \ P_{\alpha j}^{(n)} \\ &= \textstyle\sum_\alpha \ \left[P_{i\alpha}^{(n_0)} \ - \ \varepsilon \ P_{j\alpha}^{(n)}\right] \ P_{\alpha j}^{(n)} \ + \ \varepsilon \ \sum_\alpha \ P_{j\alpha}^{(n)} \ P_{\alpha j}^{(n)} \\ &= \textstyle\sum_\alpha \ \left[P_{i\alpha}^{(n_0)} \ - \ \varepsilon \ P_{j\alpha}^{(n)}\right] P_{\alpha,j}^{(n)} \ + \ \varepsilon \ P_{jj}^{2n)} \end{aligned}$$

But since $P_{i\alpha}^{(n_0)} > \varepsilon$,

$$\Rightarrow \ P_{ij}^{(n_0+n)} \ \geq m_j^{(n)}(1 \ - \ \varepsilon) \ + \ \varepsilon \ P_{jj}^{(2n)}.$$

and hence

$$m_j^{(n_0+n)} \ \geq \ m_j^{(n)}(1-\varepsilon) + \varepsilon \ P_{jj}^{(2n)}.$$

In a similar way

$$M_j^{(n_0+n)} \ \leq \ M_j^{(n)} \ (1-\varepsilon) + \varepsilon P_{jj}^{(2n)}.$$

Hence

$$M_j^{(n_0+n)} \ - \ m_j^{(n_0+n)} \ \leq \ (1-\varepsilon) \ \left(M_j^{(n)} \ - \ m^{(n)}\right).$$

and consequently

$$M_j^{(kn_0+n)} \ - \ m_j^{(kn_0+n)} \ \leq \ (1 \ - \ \varepsilon)^k \ \left(M_j^{(n)} \ - \ m_j^{(n)}\right)$$
$$\to \ 0 \quad \infty \ k \to \ \infty.$$

Hence the subsequence $M_j^{(kn_0+n)} - m_j^{(kn_0+n)}$ converges to 0. But $M_j^{(n)} - m_j^{(n)}$ is monotonic which implies that $M_j^{(n)} - m_j^{(n)} \to 0$ as $n \to \infty$.

Define

$$\pi_j \ = \ \lim_{n\to\infty} \ M_j^{(n)} \ = \ \lim_{n\to\infty} \ m_j^{(n)}$$

Then, since $m_j^{(n)} \leq \pi_j \leq M_j^{(n)}$ we have:

$$\left|P_{ij}^{(n)} \ - \ \pi_j\right| \ \leq \ M_j^{(n)} \ - \ m_j^{(n)} \ \leq \ (1 \ - \ \varepsilon)^{(\frac{n}{n_0})-1}$$

for $n \ \geq \ n_0$ which implies that $P_{ij}^{(n)} \ \to \ \pi_j \ \infty \ n \ \to \ \infty$ geometrically. Since

$$m_j^{(n)} \ \geq \ m_j^{(n_0)} \ \geq \ \varepsilon \ > 0 \quad \Rightarrow \ \pi_j \ > \ 0.$$

The proofs of b) and c) follow in a similar way.

A final remark is that the vector π is unique. Let us show this.

Let $\overline{\pi}$ be another stationary solution, for example

$$\overline{\pi}_j \ = \ \sum_\alpha \ \overline{\pi}_\alpha \ P_{\alpha j} \ = \ \sum_\alpha \ \overline{\pi}_\alpha \ P_{\alpha j}^{(n)}.$$

Since $P_{\alpha j}^{(n)} \ \to \ \pi_j$ we have

$$\overline{\pi}_j \ = \ \sum_\alpha \pi_\alpha \ \lim_{n\to\infty} \ P_{ij} \ = \ \sum_\alpha \ \overline{\pi}_\alpha \ \pi_j \ = \ \pi_j$$

Let us conclude that the theorem is a sufficiency theorem, i.e., there may exist stationary distributions even though there maybe no n_0 s.t. $\min_{ij} \ P_{ij}^{(n_0)}$ 0.

Here is an example: Let

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

Then

$$P^n \ = \ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{if n is odd}$$
$$= \ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{if n is even.}$$

Hence

$$\min_{ij} \ P_{ij}^{(n)} \ = \ 0 \quad \forall \ n.$$

But

$$\begin{aligned} \pi &= \left(\frac{1}{2},\ \frac{1}{2}\right) \qquad \text{satisfies} \\ \pi_i &= \sum_\alpha \pi_\alpha\ P_{\alpha i} \qquad \text{and moreover} \\ \pi_0 &= \frac{1\ -\ P_{11}}{2\ -\ P_{00}\ -\ P_{11}} \qquad \pi_1 \ = \ \frac{1\ -\ P_{00}}{2\ -\ P_{00}\ -\ P_{11}} \end{aligned}$$

This chain is however not ergodic. We will see what this means a little later on.

2.1 Ergodicity and Rate of Convergence

Suppose $\min_{i,j}\ P_{ij}\ =\ \varepsilon > 0$. From the proof of the main result above we have

$$|\pi_i^{(n)}\ -\ \pi_i|\ \leq\ (1\ -\ \varepsilon)^n \quad \forall\ i$$

In fact it can also be shown that:

$$\sum_j\ |P_{ij}^{(n)}\ -\ \pi_j|\ \leq\ \sum_j |\pi_j^{(n)}\ -\ \pi_j|\ <\ 2(1\ -\ \varepsilon)^n$$

Here what this result states is that the transient distribution $\pi_j^{(n)}\ \rightarrow\ \pi_j$ geometrically fast since $1\ -\ \varepsilon\ = \rho\ <\ 1$.

Remark: The quantity $\sum_j |\pi_j^{(n)} - \pi_j|$ is referred to as the *total variation norm*. It measures how *different* $\pi^{(n)}$ and π are. This is a useful metric between two probability measures defined on the same space. This property is often called geometric ergodicity of Markov chains.

Let us see what this has to do with ergodicity. Recall we usually use the term ergodicity to mean that the SLLN holds for any bounded or integrable function of the process,. More precisely, we use the trem ergodic to imply:

$$\lim_{M\to\infty} \frac{1}{M} \sum_{n=1}^{M} f((X_n) = \mathbb{E}[f(X_0)]$$

where the expectation on the r.h.s is taken under the stationary distribution of the process. Lett is see how the geometric convergence implies this. Note in this finite state setting it is enough to show that the process $\{X_n\}$ satisfies the SLLN. Recall, a stationary process $\{X_n\}$ satisfies the SLLN if $\sum_{k=0}^{\infty} |R(k)| < \infty$ where $R(k)$ is the covariance. Without loss of generality let us take $\min_{i,j} P_{ij} = \varepsilon > 0$.

Let us first compute:

$$\begin{aligned} R(n, n+m) &= \mathbb{E}[X_n X_{n+m}] - \mathbb{E}[X_n]\mathbb{E}[X_{n+m}] \\ &= \sum_{i\in E}\sum_{j\in E} ij\pi_i^{(n)} P_{ij}^{(m)} - \sum_{i\in E} i\pi_i^{(n)} \sum_{j\in E} j\pi_j^{(n+m)} \end{aligned}$$

Now taking limits as $n \to \infty$ and noting that under the conditions of the Theorem $\pi_j^{(n)} \to \pi_j$ we have $R(n, n+m) = \sum_{i,j\in E} ij\pi_i P_{ij}^{(m)} - (\mathbb{E}[X_0])^2$ and the r.h.s is just a function of m so $\lim_{n\to\infty} R(n, n+$

$m) \to R(m)$ (say). Thus it is sufficient to establish that $\sum_{k=0}^{\infty} |R(k)| < \infty$. Now using the fact that $|E| = N+1 < \infty$ we have

$$\begin{aligned} \sum_{k=0}^{\infty} |R(k)| &\leq \sum_{k=1}^{\infty} \sum_{i,j \in E} i\pi_i j |P_{ij}^{(k)} - \pi_j| \\ &= (N+1)^2 \sum_{k=0}^{\infty} (1-\varepsilon)^k < \infty \end{aligned}$$

Therefore $\{X_n\}$ and hence $\{f(X_n)\}$ for all bounded functions $f(.)$ will obey the SLLN establishing ergodicity.

Let us now study some further probabilistic characteristics of Markov chains.

3 Strong Markov Property and recurrence times

So far we have only considered the case where $|E| < \infty$. For this case we saw that if $\exists \; a \; n_0$ such that $\min_{ij} \; P_{ij}^{(n_0)} > 0$ then $P_{ij}^{(n)} \to \pi_i$ which does not depend on i the state the chain started out in.

Our interest is develop results that are also valid when $|E| = \infty$ i.e., the M.C. can take a countable infinite number of values. In this case the simple argument to show that $P_{ij}^{(n)} \to \pi_i$ cannot be carried out since P will now be a matrix of infinite rows and columns, but since $\sum_{j \in E} P_{ij}^{(n)} = 1 \, \forall \, n$ this will necessarily imply $\min_{ij} \; P_{ij}^{(n)} = 0 \; \forall \, n$, and so our previous arguments do not go through. However all is not lost – we can show some interesting properties of the type above but for this we need to undertake a more thorough study of the "structure" of the underlying chain.

Specifically in the case $|E| = \infty$ we will study the following issues:

1. Conditions when limits $\pi_j = \lim_{n \to infty} P_{ij}^{(n)}$ exist and are independent of i.
2. When $\pi = (\pi_0, \; \pi_i, \; \ldots)$ forms a probability distribution i.e., $\pi_i \geq 0 \quad \sum_{i \in E} \pi_i = 1$.
3. Ergodicity i.e., $\pi_i > 0 \quad \sum_{i \in E} \pi_i = 1$,are unique and $\pi_m = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{[X_i = m]}$.

To do so we will begin with a description of the states of a Markov chain. The classification will then enable us to conclude some general properties about states which are members of a class.

We will classify states according to two criteria:

1. Classification of states in terms of the arithmetic (or structural) properties of the transition probabilities $P_{ij}^{(n)}$
2. Classification of states according to the limiting behavior of $P_{ij}^{(n)} \;\; as \; n \to \infty$.

Let us begin by the study on the classification of the states of a M.C. based on the arithmetic properties of $P_{ij}^{(n)}$.

Throughout the discussion we will assume that $|E| = \infty$ although this is not strictly necessary.

Definition 3.1. *A state $i \in E$ is said to be inessential if, with positive probability it is possible to escape from it in a finite number of transitions without ever returning to it, i.e., $\exists$ a m and j s.t. $P_{ij}^{(m)} > 0$ but $P_{ji}^{(n)} = 0 \; \forall \; n$ and j.*

Let us delete all inessential states from E. Then the remaining states are called essential. The essential states have the property that the M.C. once it enters the essential states does not leave them.

Let us now assume that E consists only of essential states.

Definition 3.2. *We say that a state j is accessible form state i if $\exists \; m \geq 0$ s.t. $P_{ij}^{(m)} > 0$ (note by definition $P_{ij}^{(n)} = 1$ if $j = i$, $= 0$ otherwise.*

We denote this property by $i \to j$. States i and j communicate if $j \to i$ (i.e. i is accessible from j) and $i \to j$. In this case we say $i \leftrightarrow j$.

The relation $\leftrightarrow$ is symmetric and reflective i.e., if $i \leftrightarrow j$, $j \leftrightarrow k$ then $i \leftrightarrow k$.

Consequently E separates into classes of disjoint sets E_i , $E = \cup E_i$ with the property that E_i consists of states which will communicate with each other but not with E_j , $j \neq i$.

We say that $E_1, E_2, \ldots$ form indecomposable (or irreducible) classes (of communicating slides).

An example of this is a M.C. with the transition matrix.

$$P = \begin{bmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ 0 & 0 & P_3 \end{bmatrix}$$

where P_i are state transition probability matrices of appropriate dimensions. In this case there are 3 communicating classes.

In such a case since the evolutions of states defined by P_i are not influenced by stats in $P_j \; j \neq i$, the M.C. can be analyzed as 3 separate M.C.'s.

Definition 3.3. *A M.C. is said to be indecomposable or irreducible if E consists of a single indecomposable class (of communicating states).*

Now let us restrict ourselves to a chain which is irreducible (has only one class of communicating states). Even so, there can be a special structure associated with the class.

Consider for example a chain whose transition probability matrix is given by

$$P = \begin{bmatrix} 0 & P_1 & 0 & 0 & 0 \\ 0 & 0 & P_2 & 0 & 0 \\ 0 & 0 & 0 & P_3 & 0 \\ 0 & 0 & 0 & 0 & P_4 \\ P_5 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This chain is indecomposable but has the particular property that if the chain starts in the states corresponding to the first submatrix (0) then it goes to states defined by P_1 at the next transition and so on. So the chain returns to a given set of states only after 5 transitions. This is a so-called cyclic property associated with the states. We can hence sub-classify the states according to such a structure. This is related to the period of a state which we now define formally below.

Definition 3.4. *A state of $j \in E$ is said to have a period $d = d(j)$ if $P_{jj}^{(n)} > 0$ if n is a multiple of a number d and d is the largest number satisfying the property $n = md$ where m is an integer.*

In other words d is the GCD (greatest common divisor) of n for which $P_{jj}^{(n)} > 0$. If $P_{jj}^{(n)} = 0 \,\forall\, n \geq 1$ then we put $d = 0$.

Definition 3.5. *If $d(j) = 1$ then the state is said to be* **a periodic***.*

We will now show that all the states of a single indecomposable chain must have the same period and hence $d = d(j) = d(i)$ and so d is called the period of a class.

Lemma 3.1. *All states in a single indecomposable class of communicating states have the same period.*

Proof: Without loss of generality let E be indecomposable.

If $i,\ j \in E$, then $\exists\, k$ and $i > 0$ *s.t.*

$$P_{ij}^{(k)} > 0, \text{ and } P_{ji}^{(l)} > 0.$$

Hence $P_{ii}^{(k+l)} \geq P_{ij}^{(k)}\, P_{ji}^{(l)} > 0 \Rightarrow$ since $(k+l)$ must be divisible by $d(i)$. Suppose $n > 0$ but not divisible by $d(i)$. Then $n+k+l$ is not divisible by $d(i)$, hence $P_{ii}^{(k+l+n)} = 0$. But since $P_{ii}^{(k+l+n)} \geq P_{ij}^{(k)}\, P_{jj}^{(n)}\, P_{ji}^{(l)} > 0$ if n is divisible by $d(j)$. Hence $k+l+n$ must be divisible by $d(i) \Rightarrow n$ must be divisible by $d(i)$. This $\Rightarrow d(i) \leq d(j)$. By symmetry $d(j) \leq d(i)$. Hence $d(i) = d(j)$.

Definition 3.6. *A M.C. is said to be* **a periodic** *if it is irreducible and the period of the states is 1.*

We will assume that the M.C. is irreducible and a periodic from now on.

If $d > 1$ then the class of states can be subdivided into cyclic subclasses as we saw in our example where $d = 5$.

To show this select any state $i \in E$ and introduce the following subclasses.

$$
\begin{aligned}
C_0 &= \left\{ j \in E \ : \ P_{ij}^{(n)} \ > \ 0 \ \Rightarrow \ n \ = \ 0 \ mod(d) \right\} \\
C_0 &= \left\{ j \in E \ : \ P_{ij}^{(n)} \ > \ 0 \ \Rightarrow \ n \ = \ 1 \ mod(d) \right\} \\
&\vdots \\
C_{d-1} &= \left\{ j \in E \ : \ P_{ij}^{(n)} \ > \ 0 \ \Rightarrow \ n \ : \ (d-1) \ mod(d) \right\}
\end{aligned}
$$

Then it clearly follows that

$$E \ = \ C_0 \ + \ C_1 \ + \quad \ldots \quad + \ C_{d-1}.$$

In particular if $i \ \in \ C_p$ then necessarily $j \ \in \ C_{p+1}$ if $P_{ij} \ > \ 0$. For example, if $P_{ij} \ > \ 0$ then $j \ \in \ (P+1) \ mod(d)$.

Let n be such that $P_{i_0 i}^{(n)} \ > \ 0$. Then since $i \ \in \ C_p \ n \ = \ md + p$ or $n \ = \ p \ mod(d)$. Therefore $n+1 = (p+1) \ mod(d)$ and hence $j \ \in \ (P \ + \ 1) \ mod(d)$ or $j \ \in \ C_{P+1}$.

Finally let us consider a subclass, say C_p. Then the chain will enter class C_p at times given by $n \ = \ p \ mod(d)$ if it starts out in C_0 at time 0.

Consequently if $i, j \ j \ \in \ C_p$ then $P_{ij}^d \ > \ 0$ and thus the chain viewed at instants $0, \ d, \ 2d, \ \ldots$ will be a periodic with transition matrix $\overline{P} \ = \ P_{ij}^d$ which means that without loss of generality we can assume that a M.C. is irreducible and **a periodic**.

Let us summarize the classification so far:

Classification of states in terms of arithmetic properties of $P_{ij}^{(n)}$.

We will now study the second set of classification of states in terms of the asymptotic properties of $P_{ij}^{(n)}$ as $n \to \infty$.

Throughout we will assume that the M.C. is irreducible and **a periodic**.

3.1 Classification based on asymptotic properties of $\mathbf{P_{ij}^{(n)}}$

Before we begin our study of the classification based on the asymptotic properties, we will discuss the issue of the strong Markov property.

The strong Markov property implies that a M.C. continues to inherit its Markov structure when viewed at instants beyond a random time instant.

Of course the above is a very imprecise statement and so let us try to understand what it means.

Let us begin by considering a simple example.

Let

$$E = \{0,\ 1\} \quad P = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix}$$

with $P_{00} > P_{11} > 0$ and $P_{00} < 1,\ P_{11} < 1$.

Let us define the following random time $\tau(\omega)$.

$\tau = \min\{n > 0 : X_{n+1} = 0\}$. For example, τ is the time instant before the time that it first reaches 0. Then for any initial distribution $\pi^{(0)} = \left(\pi_0^{(0)},\ \pi_1^{(0)}\right)$.

$P\{X_{\tau+1} = 0 \mid X_m,\ m < \tau, |X_{\tau} =1\} = 1 \neq P_{10}$. What this means is that the Markov transition structure is not inherited by $\{X_n\}$ after the random time τ.

A natural question is when does $P\{X_{\tau+1} = j\ X_{\tau=i}\} = P_{ij}$ if τ is random? It turns out it holds when τ is a so-called Markov or stopping time which we define below.

Definition 3.7. *A random τ is said to be a Markov or stopping time if the event $\{\tau = n\}$ can be completely determined by knowing $\{X_0,\ X_1,\ \ldots\ X_n\}$, for example*

$$P\{\tau = n / X_m,\ m \geq 0\} = P\{\tau = n / X_m,\ m \leq n\}$$

Example: Let X_n be a M.C. Define

$$\tau = \min\{n > 0 : X_n = i \mid X_0 = i\}.$$

Clearly by observing $\{X_n\}$ we can determine whether τ defined as above is a stopping line.

As an example of τ which is not a stopping line is the case we considered earlier because to determine τ we need to know the future value of the process beyond τ.

We now state the strong Markov property and give a proof of it.

Proposition 3.1. *(Strong MarkovProperty)*

Let $\{X_n\}$ be a homogeneous M.C. on $(E,\ P)$

1. *The processes X_n before and after τ are independent given X_τ.*
2. $P\{X_{\tau+k_{11}}\ =\ j\ /\ X_{\tau+k}\ =\ i\}\ =\ P_{ij}$
   (i.e. the process after τ is an M.C. with transition probability P).

Proof

1. To show (a) and (b) it is enough to show

$$P\{X_{\tau+1}\ =\ j\ /\ X_m\ ;\ m\ <\ \tau,\ X_\tau\}\ =\ P\{X_{\tau+1}\ =\ j\ /\ x_\tau\ =\ \}\ =\ P_{ij}$$

for all $i,\ j\ \ \in E$.

Now (with abusage of notation)

$$P\{X_{\tau+1}\ =\ j\ /\ X_m,\ m\ <\ \tau,\ X_\tau\ =\ i\}\ =\ \frac{\{PX_{\tau+1}\ =\ j,\ X_\tau\ =\ i,\ X_m,\ m\ \ \tau\}}{P\{X_m,\ m\ <\ \tau,\ X_\tau\ =\ i\}}$$

The numerator is just

$$\begin{aligned} P\{X_{\tau+1}\ =\ j\ X_2\ &=\ i,\ X_m,\ m\ <\ \tau\} \\ &=\ \sum_{\gamma\geq 0} P\{X_{\gamma+1}\ =\ j,\ X_\gamma\ =\ i,\ X_m,\ m\ <\ \gamma,\ \tau\ =\ \gamma\} \end{aligned}$$

Now we will use the following result that follows from the definition of conditional probabilities:

$$P(A,\ B,\ C)\ =\ P(A)\ P(B/A)\ P(C/AB).$$

to write

$$\begin{aligned} \sum_{\gamma>\tau 0} P\{X_{\gamma+1}\ &=\ j,\ X_\gamma\ =\ i,\ X_m, m\ <\ \gamma,\ \tau\ =\ \gamma\} \\ &=\ \sum_{\gamma} P\{X_\gamma\ =\ i,\ X_m\ <\ \tau\} P\{X_{\gamma k}\ =\ j\ /\ X_\gamma\ =\ i,\ X_m,\ m\ <\ \gamma\} \\ &\cdot\ \ P\{\tau\ =\ \gamma\ /\ X_\tau\ =\ i,\ X_m,\ m\ <\ \gamma,\ X_{\gamma+1}\ =\ j\} \end{aligned}$$

Now $\{X_n\}$ is Markov,therefore

$$P\{X_{\gamma+1}\ =\ j\ /X_\gamma\ =\ i, X_m,\ m\ <\ \gamma\}\ =\ P\{X_{\gamma+1}\ =\ j\ /\ X_{\gamma=1}\}\ =\ P_{ij}$$

and since τ is a stopping time

$$P\{\tau\ =\ \gamma\ /\ X_\gamma\ =\ i,\ X_m,\ m\ <\ \gamma,\ X_{\gamma+'}\ =\ j\}\ =\ P\{\tau\ =\ \gamma\ /\ X_\gamma\ =\ i,\ X_m\ m\ <\ \gamma\}$$

(i.e., it does not depend on $X_{\gamma+k}\ k \geq 1\}$ so the numerator is just

$$= P_{ij}\ P\{X_\tau = i,\ X_m, m < \tau\}$$

proving the statement

$$P\{X_{\tau+1} = j\ /\ X_{\tau\,=\,i}, X_m,\ m < \tau\} = P_{ij}$$

On the other hand

$$\begin{aligned} P\{X_{\tau+1} &= j\ /\ X_\tau = i\} = \sum_{\gamma\geq 0} \frac{P\{X_{\gamma+1} = j,\ X_\gamma = i,\ \tau = \gamma\}}{P\{X_\tau = i\}} \\ &= \sum_{\gamma} P\{X_\gamma = i\}\ P\{X_{\gamma+1\,=\,j}\ /\ X_\gamma = i\}\ P\{\tau = \gamma\ /\ X_\gamma = i,\ X_m\} \\ &= P_{ij} \end{aligned}$$

showing that

$$\begin{aligned} P\{X_{\tau+1} &= j\ /\ X_\tau = i,\ X_m\ ;\ m < \tau\} \\ &= P\{X_{\tau+1} = j\ /\ X_{\tau=i}\} = P_{ij} \end{aligned}$$

or $\{X_{\tau+k}\}$ is Markov for $k \geq 0$ with the same transition matrix P.

Examples of stopping times

1. All constant (non-random) times are stopping times.
2. First entrance times such as

$$\tau_{\mathcal{F}} = \inf_n \{n \geq 0\ : X_n \in \mathbf{F}\}$$

An example of a random time which is not a stopping time is a last exit time of the type

$$\tau_E\ :\ \sup_n \{X_n \in E\ /\ X_0 \in E\}$$

Stopping times play a very important role in the analysis of Markov chains. They also play an important role in some practical situations where we can only observe certain transitions such as the so-called M.C. "watched" in a set which is the following.

Define

$$\tau_0 = \inf_n \{n \geq 0\ :\ X_n \in Y\}$$

and recursive by defining

$$\tau_{n+1} = \inf \{M > \tau_n \mid X_m \in Y\}$$

Then $Y_n = X_{\tau n}$ is a Markov chain (why?) since $\{\tau_n\}$'s are stopping times.

We will re-visit this example in more detail a little bit later. Let us now focus on first establishing the long-term behavior of M.C.

Define

$$\tau_i = \{n \geq 1 : \quad X_n = i\}$$

τ_i: First return time to state i and $\tau_i = \infty$ if $X_n \neq i \quad \forall\, n$.

Note

$$\{\tau_i = k\} = \{X_1 = \neq i,\ X_2 \neq i,\ \ldots,\ X_{k-1} = \neq i,\ X_k = i\}$$

so τ_i is a stopping time.

Define $f_{ij} = P\{T_j < \infty \mid X_0 = i\}$

f_{ij} denotes the probability the process starting in i enters state j at some finite time.

Let

$$N_i = \sum_{1}^{\infty} 1_{\{X_{n=i}\}}$$

N_i just counts the number of times the chain visits state i in an infinite sequence of moves

Define $f_{ij}^k = P\{\tau_j = k \mid X_0 = i\}$

Then we have the following result which is a direct consequence of the strong Markov property.

Lemma 3.2.

$$P_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)}\, P_{jj}^{(n-k)} = \sum_{k=0}^{n-1} P_{ii}^{(k)}\, f_{ij}^{(n-k)} \ \textit{with}\ P_{ii}(0) = 1.$$

Note

$$P_{ij}^{(n)} = P\{X_n = j \mid X_0 = i\} = \sum_{1 \leq K \leq n} P\{X_n = j,\ \tau = k \,/\, X_0 = i\}$$

$$= \sum_{1 \leq k \leq n} P\{X_{\tau+n-k} = j,\ \tau = k \mid X_0 = i\}$$

$$= \sum_{1 \leq k \leq n} P\{X_{\tau+n-k} = j \,/\, \tau = k,\ X_0 = i\}\, P\{\tau = k \mid X_0 = i\}$$

But

$$\{\tau = k\} = \{X_1 \neq j,\ X_2 \neq j,\ \ldots\ X_{k-1} \neq j,\ X_k = j\}$$

$$= \sum_{1 \leq K \leq n} P_{jj}^{(n-k)}\, P\{\tau = k \,/\, X_0 = i\} = \sum_{1}^{n} f_{ij}^{(k)}\, P_{jj}^{(n-k}$$

from the Markov property.

On the other hand,

$$P\{X_n = j \,/ X_0 = i\} = \sum_{1 \leq K \leq n-1} P\{X_n = j,\ \tau = n - K \mid X_0 = i\}$$

from which the other result follows.

This result allows us to compute the transition probability from i to j in n-steps from the first return time probability.

The return time probability plays an important role in determining the long-term behavior of M.C.'s.

Lemma 3.3. *Let N_i be the number of visits to state i defined earlier.*
Then

$$P\{N_i = k \mid X_0 = j\} = f_{ji}\, f_{ii}^{k-1}(1 - f_{ii}) \text{ if } k \geq 1$$

$$= 1 - f_{ji} \text{ if } k = 0$$

Proof: For $k = 0$ this is just the definition of f_{ji}.

Let us show the proof by induction. Suppose it is true for k. Now

$$P\{N_i > k| X_0 = j\} = 1 - \sum_{r=0}^{k} P\{N_i = r\}$$
$$= f_{ji}\, f_{ii}^{r}$$

Let τ_m denote the m_{th} return time.

$$P\{N_i = m+1 \,|X_0 = j\} = P\{N_i = m + 1,\ X_{\tau_{m+1}} = i \mid X_0 = j\}$$

$$= P\{\tau_{m+2} - \tau_{m+1} = \infty,\ X_{\tau_{m+1}} = i \mid X_0 = j\}$$

$$= P\{\tau_{m+2} - \tau_{m+1} = \infty \mid X_{\tau_{m+1}} = i,\ X_0 = i\}\ \{X_{\tau_{m+1}} = i \mid X_{i=j}\}$$

$$= P\{\tau_{m+2} - \tau_{m+1} = \infty \mid X_{\tau_{m+1}} = i\}\ P\{X_{\tau_{m+1}} = i \,/\, X_m = j\}$$

$$= P\{T_i = \infty \mid X_0 = i\}\ P\left(X_{\tau_{m+1}} = i \mid X_0 = j\right)$$

$$= (1 - f_{ii})\ f_{ii}^{m}\ f_{ji}$$

Note $P\{X_{\tau_{m+1}} = i \mid X_0 = j\} = P\{N_i > m \mid X_0 = j\}$ and therefore the proof is done.

Noting that

$$f_{ii} = P\{\tau_i < \infty \,|X_0 = i\}$$

and hence

$$f_{ii} \in (0,\ 1).$$

Now

$$P\{N_i = k \mid X_0 = i\} = f_{ii}^k(1 - f_{ii}).$$

Let

$$P_i(N_i = k) = P(N_i = k \mid X_0 = i).$$

Therefore if $f_{ii} = 1$ then

$$P\{N_i = k \mid X_0 = i\} = 0 \; \forall K.$$

Hence

$$\{N_i = \infty \mid X_0 = i\} = 1.$$

On the other hand if $f_{ii} < 1$ then

$$\begin{aligned} E\left[N_i \mid X_0 = i\right] &= \sum_{k=0}^{\infty} k \; f_{ii}^{(k)} \; (1 - f_{ii}) \\ &= \frac{f_{ii}}{1 - f_{ii}} < \infty \\ \Rightarrow \; P_i\{N_i = \infty\} &= 0. \end{aligned}$$

So

$$P_i\{N_i = \infty\} = 1 \quad \Leftrightarrow \quad f_{ii} = 1$$

It also follows that:

$$f_{ii} < 1 \quad \Leftrightarrow \quad E_i[N_i] < \infty.$$

These two quantities define a class of properties called recurrence associated with the states of a M.C.

Definition 3.8. *A state i is said to be recurrent if $N_i = \infty$ a.s.. Let T_i be the first return to i. If $\mathbb{E}_i[T_i] < \infty$ the the state is said to be positive recurrent while if $\mathbb{E}_i[T_i] = \infty$ the i is said to be null recurrent. A state that is not recurrent is said to be transient.*

Let us see one of the implications of the property $N_i = \infty \; a.s.$ Define $\tau_1 = T_i$ and

$$\tau_{n+1} = \inf_m \{m > \tau_n : |X_m = i\}$$

The $\{\tau_n\}$'s are the successive visits to state i. Define $S_n = \tau_{n+1} - \tau_n$.

We can then show the following.

Proposition 3.2. *The sequence $\{S_n\}$ is i.i.d and moreover the pieces of the trajectory*

$$\{X_{\tau_k} - X_{\tau_{k-1}}, \; X_{\tau_{k+1}} - X_{\tau_k}, \cdots\}$$

are independent and identically distributed.

Proof This is just a consequence of the strong Markov property. This is because the process after τ_k and the process before τ_k are independent. Furthermore since τ_k are the return times to the state i. We know $X_{\tau_{k+n}}$ has the same distribution as X_n given $X_0 = i$ by the strong Markov property. Also $S_k \equiv T_0$ in distribution since the chain starts off afresh in state i.

Remark 3.1. *Such pieces* $\{X_{\tau_k} - X_{\tau_{k-1}},\ X_{\tau_{k+1}} - X_{\tau_k}, \ldots\}$ *are called regenerative cycles and* τ_k *the regeneration times or epochs.*

Remark 3.2. *A consequence of these results is that if a M.C. is irreducible (indecomposable) (all states form a single communicating class), then all states are either transient or recurrent.*

Later on we will show that positive and null recurrence i.e., when the return times have finite mean or not, are also a class property.

The next result establishes the limiting behavior of $P_{ij}^{(n)}$ when the states are transient.

Lemma 3.4. *If j is transient then for every i*

$$\sum_{n-1}^{\infty} P_{ij}^{(n)} < \infty$$

and hence $\lim_{n\to\infty} P_{ij}^{(n)} = 0$.

Proof:

$$\sum_{n=1}^{\infty} P_{ij}^{(n)} = E_i[N_j]$$

and so the sum being finite means that on the average the chain visits j only a finite number of times.

Now

$$\begin{aligned}
\sum_{n=1}^{\infty} P_{ij}^{(n)} &= \sum_{n=1}^{\infty} \sum_{k=1}^{n} P\{T_j = k \mid X_0 = i\}\, P_{jj}^{(n-k)} \\
&= \sum_{k=1}^{\infty} P(T_j = k \mid X_0 = i) \sum_{1}^{\infty} P_{jj}^{(m)} \\
&= f_{ij} \sum_{1}^{\infty} P_{jj}^{(n)} < \infty
\end{aligned}$$

since j is transient. Since $0 \le f_{ij} \le 1$. Hence

$$\sum_{n=1}^{\infty} P_{ij}^{(m)} < \infty \Rightarrow P_{ij}^{(n)} \to 0$$

as $n \to \infty$ if j is transient.

Thus, with this partition of states into recurrent or transient we now show that recurrent states can be further decomposed into those where the expected return time is finite called positive recurrent and those whose expected return time is infinite , called null recurrent. Positive or null recurrence are closely associated with ergodicity of a MC.

The following figure summarizes a classification of states based on the temporal behavior.

Classification of states in terms of temporal properties of a MC .

4 Classification of state of M.C. based on temporal behavior

We saw that recurrence is a property which is dependent on whether f_{ii} is 1 or < 1 and $f_{ii} = P\{T_i < \infty \mid X_m = i\}$. This is usually not easy to calculate so we seek an alternative criterion.

To do so let us define the so called potential matrix

$$G = \sum_{n=0}^{\infty} P^{(m)}$$

Then

$$\begin{aligned} g_{ij} &= \sum_{n=0}^{\infty} P_{ij}^{(n)} = \sum_{n=0}^{\infty} P\left(X_n = j \mid X_0 = i\right) \\ &= E_i \left[\sum_{0}^{\infty} 1_{\{X_{n=j}\}} \right] \end{aligned}$$

which is just the average number of visits to j starting from state i.

We can then state the following proposition.

Proposition 4.1. *A state $i \in E$ is recurrent if*

$$\sum_{n=0}^{\infty} P_{ii}^{(n)} = \infty.$$

Proof: This is just equivalent to stating $\mathbb{E}_i[N_i] = \infty$ or the fact that the chain visits i an infinite number of times a.s..

With this equivalent condition we can now show that recurrence is a class property i.e., if $i \Leftrightarrow j$ (they belong to the same class) and i is recurrent then j is recurrent.

Proposition 4.2. *Let j be recurrent and $i \Leftrightarrow j$, then i is recurrent.*

Proof: If $i \Leftrightarrow j \ \exists\, s, t > 0$ such that

$$P_i^{(j)} > 0, \ P_{ji}^{(t)} > 0$$

Hence since

$$P_{ii}^{(s+n+t)} \geq P_{ij}^{(s)} P_{jj}^{(n)} P_{ji}^{(t)}$$

so if

$$\sum P_{jj}^{(n)} \geq \infty \Rightarrow \sum P_{ii}^{(n)} = \infty \Rightarrow i$$

is recurrent.

Reversing the arguments shows the reverse implication.

4.1 Recurrence and Invariant Measures

As we have seen if a M.C. is irreducible then either all states are recurrent or transient. Let us now study conditions for recurrence without calculating f_{ii}.

To do so we now introduce the notion of invariant measures. Invariant measures extend the notion of stationary distributions – M.C. can have invariant measures even when no stationary distribution exists. an example of such a case is a M.C. we have seen where

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

Hence $(\frac{1}{2}, \ \frac{1}{2})$ is an invariant measure.

Let us now define it formally:

Definition 4.1. *A non-null vector $\mu = \{\mu_i, \ i \in E\}$ is said to be an invariant measure for X_n if $\mu \geq 0$ and $\mu = \mu P$. i.e.,*

$$\mu_i = \sum_{j \in E} \mu_j P_{ji}$$

An invariant measure is said to be a stationary measure if $\sum_i \mu_i < \infty$. In particular we can define the stationary distribution as

$$\pi_i = \frac{\mu_i}{\sum_{i \in E} \mu_i}$$

in this case.

Let us now define a canonical invariant measure for X_n.

Proposition 4.3. *Let P be the transition matrix of a M.C. $\{X_n\}$. Assume X_n is irreducible and recurrent. Let 0 be an arbitrary state and T_0 to be the return time to 0. For each $i \in E$, define*

$$\mu_i = E_0 \left[\sum_{n \geq 1} \mathbb{1}_{\{X_n = i\}} \mathbb{1}_{\{n \leq T_o\}} \right]$$

(This is the expected number of visits to state i before returning to 0). Then for all $i \in E$

$$0 < \mu_i < \infty$$

and $\mu = \{\mu_i\}$ is an invariant measure of P.

Before we give the proof a few comments are in order.

Remark 4.1. *Note by definition: $\mu_0 = 1$. Since for*

$$n \in [1, T_0] \quad X_n = 0 \quad \textit{if and only if } n = T_0.$$

Also since

$$\begin{aligned} \sum_{i \in E} \sum_{n \geq 1} 1_{\{X_n = i\}} 1_{\{n \leq T_0\}} &= \sum_{n \geq 1} \left\{ \sum_{i \in E} 1_{X_n = i\}} 1_{\{n \leq T_0\}} \right\} \\ &= \sum_{n \geq 1} 1_{\{n \leq T_0\}} = T_0. \quad \textit{We have} \\ \sum_{i \in E} \mu_i &= E_0 [T_0] \end{aligned}$$

<u>**Proof:**</u> Let us first show that if μ_i is invariant then $\mu_i > 0$.

Let $\mu = \mu P$. Then iterating gives $\mu = \mu P^n$. So suppose $\mu_i = 0$

Then

$$0 = \sum_{j \in E} \mu_j P_{ji}^{(n)}$$

Now since $\mu_0 = 1$ we have $P_{oi}^{(n)} = 0$. Hence 0 cannot communicate with i which contradicts the hypothesis that the chain is irreducible.

On the other hand suppose $\mu_i = \infty$: Then

$$\mu_0 = 1 = \sum_{j \in E} \mu_j P_{j0}^{(n)} \geq \mu_i P_{i0}^{(n)}$$

which can only happen if $P_{i0}^{(n)} = 0 \quad \forall\, n$ which once again contradicts the irreducibility hypothesis. Hence $0 < \mu_i < \infty \quad \forall\, i \in E$.

Let us know show that μ_i as defined in an invariant measure.

Then by definition of μ_i we have

$$\mu_i = \sum_{k \geq 1} G_{0,i}^{(k)}$$

where $G_{0,i}^{(k)} = \mathbb{P}(X_k = i, T_0 > k | X_0 = 0)$ Applying the result of Lemma 5.5 we obtain for all $k \geq 2$

$$\begin{aligned} \sum_{2}^{\infty} G_{0,i}^{(k)} = \mu_i - G_{0,i}^{(1)} &= \sum_{j \neq 0} \sum_{k=2}^{\infty} G_{0,j}^{(k-1)} P_{ji} \\ &= \sum_{j \neq 0} \mu_j P_{ji} \end{aligned}$$

Noting that by definition $\mu_0 = 1$, and $G_{0,i}^{(1)} = P_{0i}$ we see

$$\mu = \mu P.$$

or μ is an invariant measure for P.

Remark 4.2. *Note that in invariant measure is only defined up to a multiplicative factor. Let us show this formally.*

Proposition 4.4. *An invariant measure of an irreducible stochastic matrix P is unique up to a multiplicative constant.*

Proof: Let y be a recurrent measure. Then we have seen that $\infty > y_i > 0 \quad \forall\, i$.

Define

$$q_{ji} = \frac{y_i}{y_j} P_{ij}$$

Then

$$\sum_i q_{ji} = \frac{1}{y_i} \sum_i y_i P_{ij} = \frac{y_i}{y_k} = 1$$

So $Q = \{q_{ij}\}$ is < stochastic matrix with

$$q_{ji}^{(n)} = \frac{y_i}{y_i} P_{ij}^{(n)}$$

Since P is irreducible Q is irreducible.

Also

$$\sum_{n \geq 0} q_{ii}(n) \;=\; \sum_{n \geq 0} P_{ii}(n).$$

So if $\sum\; P_{ii}(n)\; \infty \Rightarrow\; \sum\; q_{ii}^{(n)} \;=\; \infty$ so Q. Let

Let $g_{ji}^{(n)}$ = Prob {the chain defined by Q returns for the first time to state i at time n starting from j}

Then

$$g_{i0}^{(n+1)} \;=\; \sum_{j \neq 0} q_{ij}\; g_{j0}^{(n)}.$$

Hence

$$y_i\; g_{i0}^{(n+1)} \;=\; \sum_{j \neq 0} y_j\; g_{j0}^{(n)}\; P_{ji}$$

and, in particular, noting

$$f_{0i}^{(n+1)} \;=\; \sum_{j \neq 0} g_{0j}^{(n)}\; P_{ji}$$

we see that $f_{0i}^{(n)}$ and $y_i\; g_{i0}^{(n)}$. Satisfy the same recurrence with $f_{0,i}^{(1)} \;=\; y_i\; g_{i0}^{(n)}$. Therefore we see

$$X_i \;=\; \frac{y_i}{y_0} \text{ is also the invariant distribution}$$

$\Rightarrow\; X_i$ is obtained up to a multiplicative factor.

We can now state the Markov result for positive recurrence.

Theorem 4.1. *An irreducible M.C. is positive recurrent if its invariant measures μ satisfy*

$$\sum_{i \in E} \mu_i \;<\; \infty$$

Proof: The proof follows directly from the fact that

$$\sum_{i} \mu_i \;=\; E_0\,[T_0] \quad <\; \infty$$

Remark: Noting that

$$\pi_j \;=\; \frac{\mu_j}{\sum\; \mu_j}$$

we see that π_j when defined is unique since the multiplicative factors cancel out.

We state this as a theorem.

Theorem 4.2. *An irreducible M.C. is positive recurrent if and only if $\exists$ a stationary distribution. Moreover the stationary distribution is unique.*

Proof: The first part follows from the previous Theorem and the remark above.

Let π be the stationary distribution.

Then

$$\pi = i P^n$$

or

$$\pi_i = \sum_j \pi_i P_{ji}^{(n)}$$

Now if i were transient then $P_{ji}^{(n)} \rightarrow 0 \infty n \rightarrow 0$ then $P_i = 0$. Since the chain is irreducible then $\pi_i = 0 \forall_i$ which contradicts $\sum \pi_i = 1$. Hence the chain is the recurrent. Uniqueness follows from the argument in the remark.

Definition 4.2. *An irreducible* **a periodic** *Markov chain that is recurrent is said to be ergodic.*

Let us show that every finite state case, every homogeneous Markov chain that is irreducible is necessarily positive recurrent.

The idea is the following.

If all states are transient then (suppose these are $K + 1$)

$$1 = \lim_{n\to\infty} \sum_{j=0}^{K} P_{ij}^{(n)} = \sum_{j=0}^{K} \lim_{n\to\infty} P_{ij}^{(n)} = 0$$

which is a contradiction.

On the other hand if it is recurrent it possesses an invariant measure $\{\mu_i\}$ with $0 < \mu_i < \infty$. So $\sum_0^K \mu_i < \infty$ (finite sum) so the chain is positive recurrent.

We can now show the following result that shows the importance of the mean return time w.r.t. the stationary distribution

Theorem 4.3. *Let π be the unique stationary distribution of a +ve recurrent chain. Let T_i be the return time to state i.*

Then

$$\pi_i = \frac{1}{E_i [T_i]}$$

Proof: Since in the definition of μ_i we considered an arbitrary state 0 for which $\mu_0 = 1$, we know $\sum \mu_i < \infty$ and

$$\pi = \frac{\mu_i}{\sum \mu_j}$$

Taking $i = 0$ we obtain

$$\pi_0 = \frac{1}{\sum_i \mu_i} = \frac{1}{E_0 [T_0]}$$

but 0 is an arbitrary state. Therefore

$$\pi_i = \frac{1}{E_i[T_i]}$$

Remark 4.3. *Suppose the MC is stationary, define:*

$$\tau = \min_n \{n \geq 1 : X_n = X_0\}$$

the first return time to a given state. Suppose $|E| = N < \infty$. Then:

$$\mathbb{E}[\tau] = \sum_i \mathbb{E}[\tau | X_0 = i]\pi_i = \sum_i \mathbb{E}_i[T_i]\pi_i = N$$

since $\mathbb{E}_i[T_i] = \frac{1}{\pi_i}$. Hence if the cardinality of E is infinite then $\mathbb{E}[\tau] = \infty$. Does this contradict positive recurrence? It does not, since X_0 can be any one of the states, all the statement says that the MC cycles *through all the states on average before returning to the state it started out in. If we condition on a particular state the average return time is finite.*

So far we have only discussed the positive recurrent case and the transient case. The null recurrent case corresponds to the case when

$$\sum_i \mu_i = \infty.$$

In this case it can be shown that $P_{ij}^{(n)} \to 0 \infty n \to \infty$ if j is null recurrent. The proof of this is much more technical and so we approach it differently.

An alternate approach to showing conditions of positive recurrence and null recurrence is as follows:

Recall

$$P_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} P_{jj}^{(n-k)}.$$

Now

$$\begin{aligned} \lim_{n\to\infty} P_{ij}^{(n)} &= \lim_{n\to\infty} \sum_{k=1}^{\infty} f_{ij}^{(k)} P_{jj}^{(n-k)} \\ &= f_{ij} \lim_{n-\infty} P_{jj}^{(n)} \text{ (by monotone convergence)} \end{aligned}$$

Now if $i \Leftrightarrow j$ then $f_{ij} = 1$. Therefore it is enough to show $P_{jj}^{(n)} \to 0$. For this we use the following result.

<u>**Lemma:**</u> Let

$$U_0 = 1, \sum_{k=1}^{\infty} f_k = 1, \quad f_0 = 0$$

and

$$U_n = \sum_{k=1} f_k U_{n-k}$$

Then

$$\lim_{n \to \infty} U_n = \frac{1}{\sum_{k=1}^{\infty} k \ f_k}$$

Proof: Take Z transforms on both sides

$$U(z) = \sum_{0}^{\infty} U_n \ z^n \quad F(z) = \sum_{1}^{\infty} f_k \ z^k.$$

Then

$$U(Z) - 1 = \sum_{1}^{\infty} U_n \ z^n$$

Hence using the fact that U is a convolution of U with f

$$U(z) - 1 = F(z) \ U(z)$$

or

$$U(z) = \frac{1}{1 - F(z)}$$

The final value theorem for Z-transforms states that

$$\lim_{n \to \infty} U_n = [1 - z] \ U(z)|_{z=1}.$$

Hence

$$\lim_{n - \infty} U_n = \frac{1 - z}{1 - F(z)} |_{Z=1}$$

But

$$F(1) = \sum_{k=1}^{\infty} f_k = 1$$

so using L'Hopital's rule rule.

$$\lim_{n - \infty} U_n = \frac{-1}{-F^1(z)}|_{z=1} = \frac{1}{\sum_{K=1}^{\infty} K \ f_K \ z^{K-1}}|_{Z=1}$$

$$= \frac{1}{\sum_{k=1}^{\infty} k \ f_k}$$

Using this lemma write

$$U_n = P_{ji}^{(n)}$$

$$f_n = f_{jj}^{(n)}$$

we obtain

$$\lim_{n \to \infty} P_{jj}^{(n)} = \frac{1}{\sum_1^{\infty} n \ f_{jj}^{(n)}} = \frac{1}{E_j \ [T_j]}$$

and so if j is null recursive $E_j[T_j] = \infty$ so lim $P_{ij}^{(n)} \to 0$. On the other hand if j is +ve recurrent then

$$E_j \ [T_j] < \infty$$

then

$$\lim_{n \to \infty} P_{ij}^{(n)} = \frac{1}{E_i[T_j]} = \pi_j$$

In the above result we have shown that if j is recurrent then $\lim_{n\to\infty} P_{ij}^{(n)}$ always exists. The limit is 0 if j is null recurrent and the limit is Pi_j if j is +ve recurrent.

Actually we can show that if the chain is **a periodic** and irreducible i.e. (1 class of communicating states) then if $i \Leftrightarrow j$ i is positive recurrent then j is positive recurrent.

Let us show this. Suppose i is positive recurrent and j is not. Since i and j communicate

$$\mathcal{F}\ n,\ m > 1 \quad P_{ij}^{(n)} > 0, \quad P_{ji}^{(m)} > 0$$

Now

$$P_{jj}^{(n+m+K)} > P_{ji}^{(m)} P_{ii}^{(K)} P_{ij}^{(n)}.$$

Hence lim $k \to \infty$ $P_{jj}^{(n+m+k)} \to 0$ (null recurrence) which $P_{ij}^{(k)} \to \pi_i > 0$ which is a contradiction.

This establishes the class property of positive and null recurrent states.

So far we have concentrated on understanding how a MC behaves on the long-term. We identified these properties as related to how the return times behave. A natural question is whether there is a simple way of determining conditions on whether a chain is ergodic.

Let us consider some simple examples :

Examples:

1. (Random Walk). This is a 1-dim process constructed as follows:

   $$X_{n+1} = X_n + Z_n$$

   where $\{Z_n\}$ is i.i.d sequence and takes values in $\{-1, 1\}$ with $\mathbb{P}(Z_n = 1) = p = 1 - \mathbb{P}(Z_n = -1)$. Clearly since the chain can only return to 0 at even steps $P_{0,0}^{(2n+1)} = 0$ and $P_{00}^{(2n)} = \binom{2n}{n} p^n (1-p)^n$. Hence if $p = 0.5$ we see $\sum_n P_{00}^{(n)} = \infty$ implying 0 is recurrent. With some further analysis it can be shown that the process is actually null recurrent.

   Now if $p \neq 0.5 = q = (1-p)$ it is easy to see that $4pq < 1$ and using the fact that n is large and Stirling's formula: we have for large n, $P_{00}^{(2n)} \approx \frac{(4pq)^n}{\sqrt{\pi n}}$ and thus $\sum_n P_{00}^{(n)} < \infty$ or 0 is transient. Thus a simple random walk is not ergodic and has no stationary distribution.

2. (Reflected random walk).

   Let us now consider the same example except that when the chain hits 0 it either stays there or moves to the right. Now:

   $$X_{n+1} = (X_n + Z_n)^+$$

   where $(x)^+ = x$ if $x > 0$ or 0 otherwise.

Now it is easy to see that the period is 2 and $f_{i0} = \left(\frac{q}{p}\right)^i < 1$ if $q < p$. Hence we have all states are transient. On the other hand if $p < q$ it is easy to see $f_{i0} = 1$ implying 0 is recurrent and moreover it can be shown $\pi = \pi P$ gives:

$$\pi_j = \frac{\left(\frac{p}{q}\right)^j}{1 - \frac{p}{q}} > 0$$

establishing the chain is positive recurrent.

3. Random Walk with returns to 0 Here:

$$X_{n+1} = X_n + Z_n$$

where Z_n is and independent seq. with $\mathbb{P}(Z_n = 1 | X_n = m) = p_m = 1 - \mathbb{P}(Z_n = -X_n | X_n = m)$.

Now we can see:

$$\begin{aligned} f_{00}^{(1)} &= p_0 \\ f_{00}^{(n)} &= p_{n-1} \prod_{j=0}^{n-2} q_j \end{aligned}$$

Thus: $\mathbb{P}_0(T_0 < m) = \mathbb{P}(T_0 < m | X_0 = 0) = 1 - U_m$ where $U_m = \prod_{i=1}^{m-1} q_i$. Now we know $\lim_{n\to\infty} \prod q_j (1 - p_j) = 0 \Leftrightarrow \sum_{j=0}^{\infty} p_j = \infty$. Hence 0 is recurrent iff $\sum_j p_j = \infty$. Consider the special case $p_j = p = 1 - q_j = 1 - q$. In this case we can see $\mathbb{E}_0[T_0] < \infty$ establishing positive recurrence.

We now state the general ergodic theorem for MC. and provide a proof:

Theorem 4.4. *Let $X(0)_n\}$ be a homogeneous, irreducible, and recurrent MC. Let μ denote the invariant distribution and let $f : E \to \Re$ such that: $\sum_{i \in E} |f(i)| \mu_i < \infty$. Then:*

$$\lim_{n\to\infty} \frac{1}{\nu(n)} \sum_{k=1}^{n} f(X_k) = \sum_{i \in E} f(i) \mu_i \tag{4.1}$$

where:

$$\mu_i = \mathbb{E}_0[\sum_{k=1}^{T_0} \mathbb{I}_{[X_k = i]}$$

and

$$\nu(n) = \sum_{k=1}^{n} \mathbb{I}_{[X_k = 0]}$$

Proof: It is sufficient to show proof for positive functions. We now exploit the regenerative property of MC to prove this.

Let τ_i be the successive return times to 0. Define:

$$Y_p = \sum_{k=\tau_p+1}^{\tau_{p+1}} f(X_k)$$

Then from the strong Markov property $\{Y_p\}$ are i.i.d. and

$$\begin{aligned}
\mathbb{E}[Y_p] &= \mathbb{E}[\sum_{k=\tau_p+1}^{\tau_{p+1}} f(X_k)] = \mathbb{E}_0[\sum_{k=1}^{\tau_1} f(X_k)] \\
&= \mathbb{E}_0[\sum_{k=1}^{\tau_1}\sum_{i\in E} f(i)\mathbb{I}_{[X_k=i]}] = \sum_{i\in E} f(i)\mathbb{E}_0[\sum_{k=1}^{\tau_1}\mathbb{I}_{[X_k=i]}] \\
&= \sum_{i\in E} f(i)\mu_i
\end{aligned}$$

where we have used the definition that $\mu_i = \mathbb{E}_0[\sum_{k=1}^{\tau_0}\mathbb{I}_{[X_k=i]}]$ Therefore: by the SLLN:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} Y_i = \mathbb{E}[Y_1] = \sum_{i\in E} f(i)\mu_i$$

Now by definition: $\tau_{\nu(n)} \le n < \tau_{\nu(n)+1}$ by definition of $\nu(n)$. Noting $\nu_(n) \to \infty$ as $n \to \infty$ if the states are recurrent the result follows by noting $\sum_{k=1}^{\nu(n)} f(X_k) \le \sum_{k=1}^{n} f(X_k) \le \sum_{k=1}^{\nu(n)+1} f(X_k)$.

Corollary 4.1. *If the MC is positive recurrent then the SLLN reads:*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f(X_k) = \mathbb{E}[f(X_0)] = \sum_{i\in E} f(i)\pi_i \tag{4.2}$$

where π is the stationary distribution of the MC.

Proof: The only thing to note that if X_n is positive recurrent then $\sum_i \mu_i < \infty$ and hence:

$$\lim_{n\to\infty}\frac{n}{\nu(n)} = \sum_{i\in E}\mu_i$$

by definition of the invariant measure.

We now conclude this discussion with an easy to verify suffiiciency theorem to check whether or not a MC is positive recurrent. This is called te Foster-Lyapunov theorem and is just a consequence of the strong Markov property.

Lemma 4.1. *Let $\{X_n\}$ defined on (E,P) be a hmomgeneous MC Let $F \subset E$ and $\tau_F = \inf\{n \ge 0 : X_n \in F\}$ be the hitting or first entrance time to the set F. Define:*

$$m(i) = \mathbb{E}[\tau_F | X_o = i]$$

Then:

$$\begin{aligned}
m(i) &= 1 + \sum_{j\in E} P_{ij} m(j); \quad i \notin F \\
&= 0 \ i \in F
\end{aligned}$$

Proof: Clearly if $i \in F$ the result is trivial. Now suppose $X_0 \notin F$ then τ_F being a stopping time is a function of X_n i.e.

$$\tau_F(X_n) = 1 + \tau_F(X_{n+1})$$

by the Markov property since in 1 step it goes from X_n to X_{n+1}. Hence:

$$\begin{aligned}
\mathbb{E}[\tau_F(X_n)|X_0 = i] &= \mathbb{E}[1 + \tau_F(X_{n+1})|X_0 = i] \\
&= 1 + \sum_{j \in E} \mathbb{E}[\tau_F(X_{n+1})\mathbb{1}_{X_1=j]}|X_o = i] \\
&= 1 + \sum_{j \in E} \mathbb{E}[\tau_F(X_{n+1})|X_1 = j]\mathbb{P}(X_1 = j|X_0 = i) \\
&= 1 + \sum_{j \in E} P_{ij} m(j)
\end{aligned}$$

where we used the strong Markov property in the 3rd step.

With the help of this lemma we now state and prove the Foster-Lyapunov theorem.

Theorem 4.5. *(Foster-Lyapunov Criterion) Let $\{X_n\}$ be an irreducible, homogeneous MC on (E, P). Then a sufficient condition for $\{X_n\}$ to be positive recurrent is that $\exists$ a function $h(.)$: $E \to \Re$ and a finite subset F of E and a $\varepsilon > 0$ such that:*

a) $\inf_{i \in E} h(i) > -\infty$

b) $\sum_{k \in E} P_{ik} h(k) < \infty \quad \forall\, i \in F$

c) $\sum_{k \in E} P_{ik} h(k) \le h(i) - \varepsilon \;\forall\, i \notin F$

Proof: First note since $\inf_i h(i) > -\infty$, by addiding a constant we can assume that $h(i) \ge 0 \;\forall\, i \in E$. By the definition of transition probabilities c) can be written as:

$$\mathbb{E}[h(X_{n+1})|X_n = j] \le h(j) - \varepsilon \quad \forall\, j \in F \tag{4.3}$$

which is equivalent to :

$$\mathbb{E}[h(X_{n+1}) - h(X_n)|X_n = j] \le -\varepsilon < 0$$

or the conditional drift in state $j \in F$ is negative.

Let $\tau_F = \inf_n\{n \ge 1 : X_n \in F\}$ and define $Y_n - h(X_n)\mathbb{1}_{[n<\tau_F]}$.

Note τ_F is a stopping time. Let $i \notin F$ and $\mathbb{E}_i[.]$ denote $\mathbb{E}[.|X_0 = i]$, then:

$$\begin{aligned}
\mathbb{E}_i[Y_{n+1}|X_0, X_1, \ldots, X_n] &= \mathbb{E}_i[T_{n+1}\mathbb{1}_{[n<\tau_F]}|X_0, \ldots, X_n] + \mathbb{E}_i[Y_{n+1}\mathbb{1}_{[\tau_F \le n]}X(0)_0, \ldots, X_n] \\
&= \mathbb{E}_i[Y_{n+1}\mathbb{1}_{[n<\tau_F]}|X_0, \ldots, X_n] \\
&\le \mathbb{E}_i[h(X_{n+1})\mathbb{1}_{[n<\tau_F]}|X_0, \ldots, X_n] \\
&= \mathbb{1}_{[n<\tau_F]}\mathbb{E}_i[h(X_{n+1}|X_n] \\
&\le \mathbb{1}_{[n<\tau_F]}(h(X_n) - \varepsilon)
\end{aligned}$$

where we used the fact that $\mathbb{1}_{[n<\tau]}$ is completely known given $X_0, \ldots, X_n$ and if $n < \tau_F$ then $X_n \notin F$.

So taking expectations w.r.t. $\mathbb{E}_i$ once more,

$$0 \le \mathbb{E}_i[Y_{n+1}] \le \mathbb{E}_i[Y_n] - \varepsilon\mathbb{P}_i(\tau_F > n)$$

Iterating this inequality startng from 0 we obtain:

$$0 \leq E_i[Y_0] - \varepsilon \sum_{k=1}^{n} \mathbb{P}_i(\tau_F > k)$$

But we know: $\sum_{k=1}^{\infty} \mathbb{P}_i(\tau_F > k) = \mathbb{E}_i[\tau_F]$, but $\mathbb{E}_i[Y_0] = h(i)$ and hence: $\mathbb{E}_i[\tau_F] \leq \frac{h(i)}{\varepsilon} < \infty$.

On the other hand using the previous lemma we have for $j \in F$:

$$\mathbb{E}_j[\tau_F]1 + \sum_{i \notin F} P_{ji}\mathbb{E}_i[\tau_F]$$

and hence:

$$E_j[\tau_F] \leq 1 + \frac{1}{\varepsilon} \sum_{i \notin F} P_{ji} j(i)$$

which is finite by condition b).

Thus $\mathbb{E}_i[\tau_F] < \infty$ for all states $i \in F$. Since F is finite it immediately implies that for any $i \in F$, $\mathbb{E}_i[T_i] < \infty$ where T_i is the return time to state i and hence the states are positive recurrent. Since by assumption the chain is irreducible all states are therefore positive recurrent and thus the chain is ergodic.

In many applications $E = Z_+ = \{0, 1, 2, \ldots, \}$. In this case there is a much simpler version known as Pakes' theorem that applies. We state this below.

Corollary 4.2. *Let $E = Z_+$. Define the conditional drift in state i as follows:*

$$r_i = \mathbb{E}[X_{n+1} - X_n | X_n = i] \tag{4.4}$$

Suppose:

i)$\sup_{i \in E} |r_i| < \infty$

ii) There exists a $i_0 < \infty$ such that for all $i \geq i_0$, $r_i < -\varepsilon$ for some $\varepsilon > 0$.

Then the chain is ergodic.

Proof: This just follows from above by taking $h(X_n) = X_n$ and $F = \{i \in Z+ : i \leq i_0 - 1\}$. Then all conditions of the Foster-Lyapunov theorem are satisfied.

We conclude our discussion to show how these results apply on a canonical example the represents a discrete-time queue.

Example:

Let :

$$X_{n+1} = (X_n - 1)^+ + \nu_{n+1}$$

where ν_{n+1} is a i.i.d sequence with $0 < \mathbb{E}[\nu_{n+1}]] < 1$.

Then applying Pakes theorem we see for all $i > 1$:

$$\mathbb{E}[X_{n+1} - X_n | X_n = i] = -1 + \mathbb{E}[\nu_{n+1}] < 0$$

implying that the chain is ergodic,

In the next section we study the convergence to stationary state a bit further as in the finite state case.

4.2 Coupling and Convergence to Steady State

Suppose X_n is an **a periodic, irreducible M.C.** which is positive recurrent. We have shown $\frac{1}{N}\sum_{n=1}^{N} P_{ij}^{(n)} \rightarrow \pi_j$ *as* $N \rightarrow \infty$ from the ergodic theorem by noting that $P_{ij}^{(n)} = \mathbb{E}_i[\mathbb{I}_{X_n=j}]$. When $|E| < \infty$ we actually showed that $P_{ij}^{(n)} \to \pi_j$ as $n \to \infty$ independent of i and the convergence rate was geometric. This convergence is actually related to the notion of *stability* . We discuss this issue in detail now. Specifically:

How does $P_{ij}^{(n)} \rightarrow \pi_j$? In the finite case we have seen the convergence is geometric. Under what conditions is this true for infinite chains?

We can show something stronger. X_n converges to a stationary process in a finite but random time. This is called the setting time or coupling time. The ramification of this is that when we try to simulate stationary MC we need not wait for an infinite time for the chain to be stationary, we can observe certain events, and once they occur we can conclude that after that time the chain is stationary.

But first let us recall the result we showed for finite state Markov chains.

Let P be at time $n \times n$ and let

$$\min_{(i)} P_{ij} = \varepsilon > 0$$

$$\begin{aligned} \text{Let} \quad \pi_i^{(n)} &= P\{X_n = i\} \\ \text{and} \quad \pi_i &= P\{X_n = i\} \quad \text{(stationary dist)} \\ \text{where} \quad \pi_i &= \sum_j \pi_j P_{ji} \end{aligned}$$

Define

$$\|\overline{\pi}^{(n)} - \pi\| = \frac{1}{2}\sum_{1} | \pi_i^{(n)} - \pi_i |$$

This is called the 'total variation' metric and convergence under this is called total variation convergence. The factor $\frac{1}{2}$ is just to normalize the metric.

Now in the proof of Theorem 5.2 we saw

$$m_j(n) \leq P_{ij}(n) \leq M_j^{(n)}$$

and since

$$\min_{ij} P_{ij} > 0$$

$$\Rightarrow \pi_j^{(n)} \downarrow \pi_j, \quad m_j^{(n)} \uparrow \pi_j$$

Hence

$$\begin{aligned} \sum_j | \pi_j^{(n)} - \pi_i| &= \sum_j | \sum_i \pi_i^{(0)} P_{ij}^{(n)} - \pi_j | \\ &\leq \sum_j | M_j^{(n)} - m_j^{(n)} | \\ &\leq (1-\varepsilon)^n \sum_j | M_j - m_j | \leq 2(1 - \varepsilon)^n. \end{aligned}$$

Note

$$\sum_{j} \mid M_j \ - \ m_j| = \ \sum_{j \in E} | \ \max_{i \in E} \{ \mid P(ij)\} - \min_{i \in E} \ P(ij) \leq 2$$

Hence

$$\|\overline{\pi}^{(n)} \ - \ \pi \ \| \ \leq \ \ 2(1 \ - \ \varepsilon)^n$$

This convergence is geometric.

Indeed $(1 \ - \ \varepsilon)^n$ is related to the tail distribution of a fundamental quality associated with the convergence: a so-called coupling time which we will now discuss. Coupling is a powerful technique which can be used to establish existence of stationary distributions, rate of convergence, etc.

The basic approach is the following: suppose $\{X_n^1\}$ and $\{X_n^2\}$ are two homogeneous irreducible and **a periodic** M.C.'s with the same P which are independent.

Define $Z_n \ = \ (X_n^1, \ X_n^2)$ on $E \ X \ E$. Then Z_n is a M.C. with transition problem matrix

$$\overline{P}_{ij,kl} P \{ Z_{n+1} \ = \ ((k,l) \ / \ Z_n \ = \ (i,j)) \ \ = \ P_{ik} \ P_{jl} \}$$

Suppose the chain is positive recurrent then, $\mathcal{F}$ a finite τ such that starting for any states i and j the chain goes to a diagonal state where the two co-ordinates are equal i.e.,

$$X_\tau^1 \ = \ X_\tau^2$$

Define

$$\begin{aligned} X_n \ &= \ X_n^1 \quad n \ \leq \ \tau \\ &= \ X_n^2 \quad n \geq \ \tau. \end{aligned}$$

Then we can show the following.

Proposition: $\{X_n\}$ is a +ve recurrent M.C. with transition probability matrix $\overline{P}$ defined above .

Proof: This follows directly from the strong M.C. proposition. Let us formally define coupling.

Definition 4.3. *2 stochastic processes $\{X_n^1\}, \{X_n^2\}$ and with values in E are said to couple if there exists a $\tau(\omega) \ < \ \infty$ s.t. for all*

$$n \ \geq \ \tau : \Rightarrow \ X_n^1 \ = \ X_n^2.$$

Lemma 4.2. *(The coupling inequality)*

Let $\{X_n^1\}$ and $\{X_n^2\}$ be two independent processes defined on $(\Omega, \mathcal{F}, P)$ and let τ be a coupling time . Then for any $A \in \mathcal{F}$ we have:

$$|P(X_n^1 \in A) - P(X_n^2 \in A)| \leq P(\tau > n) \tag{4.5}$$

Proof:

$$
\begin{aligned}
P(X_n^1 \in A) - P(X_n^2 \in A) &= P(X_n^1 - A,\ \tau \leq n) \\
&- P(X_n^2 \in A,\ \tau \leq n) \\
&+ P(X_n^1 \in A,\ \tau > n) \\
&- P(X_n^2 \in A, \tau > n)
\end{aligned}
$$

Now if $\tau \leq n \Rightarrow X_n^1 = X_n^2$ by definition of τ. Therefore

$$
\begin{aligned}
P(X_n^1 \in A) - P(X_n^2 \in A) &= P(X_n^1 \in A,\ \tau > n) \\
&- P(X_n^2 \in A)\ \tau > n) \\
&\leq P(\tau > n).
\end{aligned}
$$

By symmetry we have the $P(X_n^2 \in A) - P(X_n^1 \in A) \leq P(\tau > n)$ and so the result follows.

Using this inequality we will now prove the convergence in the true recurrent sense.

Now suppose X_n^1 is a +ve recurrent chain independent of X_n^2 which is also +ve recurrent (we assume both are **a periodic and irreducible**). Then $Z_n = (X_n^1,\ X_n^2)$ is +ve recurrent.

We are now ready to state the main convergence or stability result.

Proposition 4.5. *Let $\{X_n\}$ be an irreducible, aperiodic and positive recurrent with stationary distribution π. Then:*

$$\lim_{n\to\infty} \pi_j^{(n)} = \pi_j \tag{4.6}$$

uniformly in $j \in E$ for any initial distribution. In particular

$$\lim_{n\to\infty} P_{ij}^{(n)} = \pi_j$$

for all $i, j \in E$.

Proof:

Construct to independent MC on $E \times E$ with transition probability $\overline{P}$.

Let τ be a coupling time state the chains meet at $X_n^1 = X_n^2$.

Then if X_n^2 has an initial distribution π then $P\{X_n^2 = j\} = \pi_j$ for all j.

Using the coupling inequalities

$$
\begin{aligned}
\sum_j \mid P\{X_n^1 = j\} - \pi_j \mid &\leq \sum P(X_n^1 = j,\ \tau > n) + P(X_n^2 = j, \tau > n) \\
&\leq 2\ P(\tau > n).
\end{aligned}
$$

Therefore since $\tau < \infty \quad P(\tau \to n) \to 0 \infty n \to \infty$. So

$$|P_{ij}^{(n)} - \pi_j| \to 0$$

From this we see $\sum_{j\in E} |P_{ij}^{(n)} - \pi_j| \to 0 \ as\ n \to \infty$

In fact after τ the chain can be considered to have converged to the stationary distribution.

Remark 4.4. *The aperiodicity and irreducible assumption is important. Otherwise is is very easy to construct periodic chains that never meet at a diagonal especially if they start out in different cyclic subclasses. Hence the periodic case can be treated by considering the transition probability matrices* P^d *.*

How do we get convergence rates from these results?

Lemma 4.3. *Suppose* $E[\varphi(\tau)] < \infty$ *for a non-decreasing function* $\varphi(.)$*. Then*

$$|P_{ij}^n - \pi_j| = 0\left(\frac{1}{\varphi(n)}\right)$$

<u>**Proof:**</u> Since $\varphi(\tau)$ is non-decreasing

$$\varphi(\tau)\ 1\ (\tau > n) > \varphi(n) 1_{\{\tau > n\}}$$

So

$$\varphi(n) P(\tau > n) \leq E\left[\varphi(\tau)\ 1_{\{\tau > n\}}\right]$$

Now since $E\left[\varphi(\tau)\ 1_{\{\tau>n\}}\right] \to 0 > n \uparrow \infty$ by finiteness of $E[\varphi(\tau)]$ we have $\Rightarrow \varphi(n)\ P(\tau > n) \to 0 \infty\, n \to \infty\ \ P(\tau > n) = 0\left(\frac{1}{\varphi(n)}\right)$.

Of course, depending on the MC we need to establish that $\mathbb{E}[\varphi(\tau)] < \infty$. When $|E| < \infty$ it is easy to show the following:

Lemma 4.4. *Let* $\{X_n\}$ *be a finite state M.C. on* (E, P)*, then there exists* $\alpha > 0$ *s.t.*

$$\mathbb{E}\left[e^{\alpha\tau}\right] < \infty.$$

Proof: The proof follows from taking $\varphi(\tau) = e^{\alpha\tau}$ and since the MC is finite the hitting time to the diagonal state can be shown to have a geometric tail distribution. Hence convergence of the distribution to the steady state is geometric.

With this we conclude our study of discrete-time Markov chains. In the next part we will study continuous-time Markov chains where these results will play an important part.