AMERICAN UNIVERSITY OF BEIRUT

ASYMPTOTIC BEHAVIOUR OF PROBABILISTIC CELLULAR AUTOMATA

by
ALI KHALIL HALAWI

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Sciences
to the Department of Mathematics
at the American University of Beirut

Beirut, Lebanon
September 2022

AMERICAN UNIVERSITY OF BEIRUT

ASYMPTOTIC BEHAVIOUR OF PROBABILISTIC CELLULAR AUTOMATA

by

ALI KHALIL HALAWI

Approved by:

Taati

Dr. Siamak Taati, Assistant Professor
Mathematics — Advisor

Abbas Alhal

Dr. Abbas AlHakim, Associate Professor
Mathematics — Member of Committee

Dr. Bassam Shayya, Professor
Mathematics — Member of Committee

Date of thesis defense: September 2, 2022

AMERICAN UNIVERSITY OF BEIRUT

THESIS RELEASE FORM

Student Name: ____Halawi_____Ali_____Khalil_____
Last First Middle

I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of my thesis; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes:

☒ As of the date of submission

☐ One year from the date of submission of my thesis.

☐ Two years from the date of submission of my thesis.

☐ Three years from the date of submission of my thesis.

Ali Halawi September 8, 2022

Signature Date

Acknowledgements

This project would not have been possible without the support of many people. Many thanks to my adviser, Siamak Taati, who read my numerous revisions and helped make some sense of the confusion. Also thanks to my committee members, Bassam Shayya, Abbas Al Hakim, who offered guidance and support.

Abstract
of the Thesis of

Ali Khalil Halawi for Master of Sciences
Major: Mathematics

Title: Asymptotic Behaviour of Probabilistic Cellular Automata

Cellular automata are mathematical models used to study complex systems in physics, biology and computer science. The automaton iteratively evolves from one configuration to another using local transition rules based on the neighborhood of each cell. The aim of this project is to study the notions of ergodicity and uniform ergodicity in probabilistic cellular automata. Before studying the notion of ergodicity in probabilistic cellular automata, I will start by studying the notion of ergodicity in finite-state Markov chains. The reason behind doing this is that finite- state Markov chains are simpler than probabilistic cellular automata. I will then introduce the notion of a probabilistic cellular automaton. Here, I distinguish between three classes of probabilistic cellular automata: the fully deterministic ones, the fully probabilistic ones, and the rest. Afterwards, I will present the proof of the equivalence between ergodicity and uniform ergodicity in two special cases of PCA which are: fully probabilistic and fully deterministic. But first I will give the needed background to get the result for both cases.

Table of Contents

Illustrations

Tables

Abbreviations

- $\mathbb{N} \coloneqq \{0, 1, 2, \ldots\}$.
- $\mathbb{Z}^+ \coloneqq \{1, 2, 3, \ldots\}$.
- $A \Subset B$ means A is a finite subset of B.

Chapter 1

INTRODUCTION

Cellular automata were first introduced by John von Neumann in 1966 as formal models of self-reproducing organisms. The studied structure was mostly one and two dimensional grids. Also higher dimensions were taken into consideration. After that, physicists and biologists started to study cellular automata in their domains.

Cellular automata (CA) are collection of cells arranged in a grid of specified shape. Time is discrete, and each cell will be in one of the possible states. The set of possible states for each cell is finite and The set of states is the same for all cells. Each cell goes to another state at each time step, and this new state is determined by the the previous one and its neighbors according to a function. The function is called the *local rule*. It is the same for all cells. A configuration is the set of cell states at any time. Applying the local rule to each cell of the cellular automata transforms each configurations into another one.

In this thesis, we are going to talk about cellular automata and discuss two main properties of it which are ergodicity and uniform ergodicity and how they are related.

In the first chapter, I am going to introduce finite-state Markov chains in order to study their ergodicity. The reason behind studying the ergodicity of Markov chains is that its notations and rules are simpler than that in cellular automata. This will illustrate the idea of ergodicity for the next chapters. In this chapter, I will consider finite-state and discrete-time Markov chains and transitions will be time homogeneous.

In the second chapter, I will start by introducing cellular automata by giving some preliminaries that are needed to give mathematical definitions of cellular automata. We are going to consider probabilistic cellular automata. Here we are going to distinguish between two cases: The fully deterministic case (DCA) and the fully probabilistic case . Where every DCA can be considered as a PCA. After that, I will introduce the notions of ergodicity and uniform ergodicity in the general settings where we have a compact metric space, which is more general setting than cellular automata.

In the last chapter, I will study the relation between ergodicity and uniform ergodicity. It will be clear that uniform ergodicity always implies ergodicity. We are going to discuss and study the other direction. As a result, at the end of this thesis we are going to get a partial answer to the following question: Is it true that every ergodic cellular automata is uniformly ergodic?

Chapter 2

ERGODICITY of MARKOV CHAINS

The objective of this chapter is to illustrate the idea of ergodicity for finite-state Markov chains. First, we will start by introducing discrete-time and finite-state Markov chains. Our discussion will be limited to discrete-time and finite-state Markov chains. Then, I will introduce the transition probabilities that is the probability of going from one state to another. After that, I will classify the types of the states, and I will talk about irreducibly and periodicity. Further, I will introduce stationary distribution that can be interpreted as a state of equilibrium. I will show that if a finite state Markov chain is irreducible, then the stationary distribution is unique. I will end this chapter by discussing the long-run behavior of a finite-state Markov chain. I will prove the convergence theorem: if a (finite-state) Markov chain is both irreducible and aperiodic, then its distribution converges to the unique stationary distribution. This means that it converges to its unique state of equilibrium. In other words, every irreducible and aperiodic finite-state Markov chain is ergodic.

2.1 Introduction to Markov Chains

In this section, I will consider discrete time Markov chains by introducing the Markov property and the importance of transition probability matrices matrices.

2.1.1 *Preliminaries*

I will starting by giving some definitions needed in the next sections.

Definition 2.1.1 (Sigma Algebra)**.** Assume that X is any set, and let $P(X)$ be its power set. Then a subset $\mathcal{F} \subseteq P(X)$ is called sigma-algebra if it has the following properties:

1. $X \in \mathcal{F}$.
2. If $A \in \mathcal{F}$ then $X \setminus A \in \mathcal{F}$.
3. If $A_1, A_2, ... \in \mathcal{F}$ then $A_1 \cup A_2 \cup ... \in \mathcal{F}$.

Definition 2.1.2 (Borel sigma algebra)**.** Assume that X is a topological space and $\mathcal{F}$ the collection of all open subsets of X. The Borel σ-algebra in X is the σ-algebra $\mathcal{B}$ generated by $\mathcal{F}$. Members of $\mathcal{B}$ are called Borel sets.

If $\mathcal{F}_1, \mathcal{F}_2, ...$ are closed subsets of X, then $\mathcal{F}_1 \cup \mathcal{F}_2 \cup ...$ is called a Borel set.
If $\mathcal{G}_1, \mathcal{G}_2, ...$ are open subsets of X, then $\mathcal{G}_1 \cup \mathcal{G}_2 \cup ...$ is called a Borel set.

Definition 2.1.3 (Measurable Space)**.** Consider a set A and a sigma algebra $\mathcal{F}$ on A. Then the tuple $(A, \mathcal{F})$ is said to be a measurable space.

Definition 2.1.4 (Measurable Set)**.** Let $(A, \mathcal{F})$ be a measurable space. Then any set $S \in \mathcal{F}$ is a measurable set.

Definition 2.1.5 (Measurable function)**.** Let $(X, \mathcal{F})$ and $(Y, \mathcal{E})$ be measurable spaces. A function $f : X \to Y$ is said to be measurable if $f^{-1}(E) \in \mathcal{F}$ for every $E \in \mathcal{E}$

Definition 2.1.6 (Probability Measure)**.** A probability measure is a mapping $\mathbb{P} : \mathcal{F} \to [0, 1]$ that assigns a probability $\mathbb{P}(A) \in [0, 1]$ to any event A, with the properties

1. $\mathbb{P}(\Omega) = 1$, and
2. $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ whenever $A_k \cap A_l = \emptyset, k \neq l$

Definition 2.1.7 (Probability Space)**.** A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where $(\Omega, \mathcal{F})$ is a measurable space, and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$.

A probability space is the mathematical model of a random experiment or phenomenon. The set Ω is called the sample space. The elements of Ω are the possible outcomes of the experiment. The measurable subsets of Ω are the events to which probabilities are assigned.

Definition 2.1.8 (Random Variable)**.** A random variable X is a measurable function that assigns a value to each possible outcome of the statistical experiment. More precisely, a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a measurable function $X : \Omega \to E$ from $(\Omega, \mathcal{F})$ into another measurable space $(E, \mathcal{E})$. Note that you were using the same symbol $\mathcal{F}$ for two different σ-algebras, one on Ω and one on E. If E is countable then X is called a discrete random variable.

Example 2.1.9. Let us roll a dice twice so we get $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$, and consider the mapping:

$$X : \Omega \to \mathbb{R}$$
$$(m, n) \to m - n$$

Then X is a random variable giving the difference of the two numbers that appears on each die.

Definition 2.1.10 (Random Processes or Stochastic Processes)**.** A stochastic process consists of a collection of random variables, indexed by some set T. More specifically, for a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and a measurable space (S, Σ), a stochastic process is a collection of S-valued random variables

$$(X_t)_{t \in T}$$

S which is called the state space that is the range of values taken by the stochastic process.
A stochastic process $(X_t)_{t\in T}$ is said to be a discrete-time stochastic process when T is discrete (i.e. finite or countable).

In this thesis, we are only concerned with discrete time Markov chains, so T will be the set of non-negative integers.

2.1.2 *Discrete-Time Markov Chains*

For this section I used [1] as a reference.

Definition 2.1.11 (Markov Property)**.** Let S be discrete set, with the discrete σ-algebra. Let us consider an S-valued discrete-time stochastic process $(X_t)_{t\in\mathbb{N}}$. This process is said to be a Markov chain or to have the Markov property if, for all $t \geq 1$, the probability distribution of X_{t+1} is determined only by the state X_t of the process at time t, and the values of X_k for $k \leq t-1$ does not affect it.

Mathematically, for all $t \geq 1$ and $x_0, x_1, ..., x_t, y \in S$, we have:

$$\mathbb{P}(X_{t+1} = y | X_0 = x_0, \ldots, X_t = x_t) = \mathbb{P}(X_{t+1} = y | X_t = x_t)$$

Definition 2.1.12 (Time-Homogeneous)**.** A Markov chain $(X_t)_{t\in\mathbb{N}}$ is called time-homogeneous if for every $x, y \in S$ and every $t \geq 0$, we have

$$\mathbb{P}(X_{t+1} = y | X_t = x) = \mathbb{P}(X_1 = y | X_0 = x)$$

that is, the transition probabilities are independent of time.

From now on we will assume that all Markov chains we talk about are time-homogeneous.

Definition 2.1.13 (Trajectory)**.** A trajectory of a Markov chain is a sequence of values for $X_0, X_1, ...$

Example 2.1.14 (Trajectory)**.** If $X_0 = 1$, $X_1 = 4$, and $X_2 = 7$, then the trajectory up to time $t = 2$ is $1, 4, 7$. ◯

Definition 2.1.15 (Transition probabilities)**.** Let $(X_t)_{t\in\mathbb{N}}$ be a Markov chain with state space S such that S is of finite size N. Then, for every $x, y \in S$, define P_{xy} the transition probability of the Markov chain as

$$P_{xy} = \mathbb{P}(X_{t+1} = y | X_t = x) \qquad \text{for} \qquad x, y \in S \tag{2.1}$$

In words, P_{xy} is the probability of transition to y in one step if the Markov chain is currently at state x.

Now, I will give a proposition that will be useful for our work.

Proposition 2.1.16. *Let $(X_t)_{t\in\mathbb{N}}$ be a Markov chain. Then,*

1. *For every $t \geq 0$ and $x_0, x_2, \ldots, x_t \in S$, we have*

$$\begin{aligned}&\mathbb{P}(X_t = x_t, \ldots, X_0 = x_0)\\ &\quad = \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}) \ldots \mathbb{P}(X_1 = x_1 | X_0 = x_0)\, \mathbb{P}(X_0 = x_0).\end{aligned}$$

2. For every $x \in X$, we have

$$\begin{aligned}
&\mathbb{P}(X_1 = x) \\
&\quad = \sum_{y \in S} \mathbb{P}(X_1 = x, X_0 = y) \\
&\quad = \sum_{y \in S} \mathbb{P}(X_1 = x | X_0 = y)\, \mathbb{P}(X_0 = y), \qquad x \in S.
\end{aligned}$$

Proof. To prove the first claim, note that

$$\begin{aligned}
&\mathbb{P}(X_t = x_t, \ldots, X_0 = x_0) \\
&\quad = \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}, \ldots, X_0 = x_0)\, \mathbb{P}(X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \ldots, X_0 = x_0) \\
&\qquad \text{(By definition of conditional probability)} \\
&\quad = \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1})\, \mathbb{P}(X_{t-1} = x_{t-1} | X_{t-2} = x_{t-2}, \ldots, X_0 = x_0) \\
&\qquad \mathbb{P}(X_{t-2} = x_{t-2}, X_{t-3} = x_{t-3}, ..., X_0 = x_0) \\
&\qquad \text{(By Markov property and conditional probability)} \\
&\quad = \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1})\, \mathbb{P}(X_{t-1} = x_{t-1} | X_{t-2} = x_{t-2}) \\
&\qquad \mathbb{P}(X_{t-2} = x_{t-2} | X_{t-3} = x_{t-3}, \ldots, X_0 = x_0)\, \mathbb{P}(X_{t-3} = x_{t-3}, \ldots, X_0 = x_0) \\
&\quad = \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1})\, \mathbb{P}(X_{t-1} = x_{t-1} | X_{t-2} = x_{t-2}) \\
&\qquad \mathbb{P}(X_{t-2} = x_{t-2} | X_{t-3} = x_{t-3})\, \mathbb{P}(X_{t-3} = x_{t-3}, ..., X_0 = x_0) \\
&\quad = \cdots
\end{aligned}$$

So we get,

$$\begin{aligned}
&\mathbb{P}(X_t = x_t, \ldots, X_0 = x_0) \\
&\qquad = \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}) \ldots \mathbb{P}(X_1 = x_1 | X_0 = x_0)\, \mathbb{P}(X_0 = x_0)
\end{aligned} \tag{2.2}$$

for $x_0, x_1, \ldots, x_t \in S$. Using the law of total probability we get also:

$$\begin{aligned}
\mathbb{P}(X_1 = x) &= \sum_{y \in S} \mathbb{P}(X_1 = x, X_0 = y) \\
&= \sum_{y \in S} \mathbb{P}(X_1 = x | X_0 = y)\, \mathbb{P}(X_0 = y), \qquad x \in S
\end{aligned}$$

□

Definition 2.1.17 (Transition Matrix)**.** The transition matrix of a Markov chain is an $N \times N$ is $P = (P_{xy})_{x \in S, y \in S}$ where $S = \{1, 2, \ldots, N\}$.

Going back to (2.1) we note that the row number in the matrix is x which is the initial state while y refers to the column number which is the final state, we have

$$\sum_{y \in S} \mathbb{P}(X_1 = y | X_0 = x) = 1, \qquad x \in S$$

where the rows of the transition matrix satisfy

$$\sum_{y\in S} P_{xy} = 1, \qquad \text{for every} \qquad x \in S$$

Now by the definition of P_{xy} and (2.2), we get

$$\mathbb{P}(X_t = x_t, \ldots, X_0 = x_0) = P_{x_{t-1}x_t x_0 x_1}\, \mathbb{P}(X_0 = x_0),$$

for all $x_0, x_1, \ldots, x_n \in S$ becomes

$$\mathbb{P}(X_1 = x) = \sum_{y\in S} P_{yx}\, \mathbb{P}(X_0 = y), \qquad x \in S \tag{2.3}$$

Finally, we can write (2.3) in matrix and vector notation as follows

$$\mu = \pi P$$

where

$$\pi = [\pi_1, \pi_2,], \pi_N] = [\mathbb{P}(X_0 = 1), ..., \mathbb{P}(X_0 = N)] \in \mathbb{R}^N$$

is the distribution of X_0 and

$$\mu = [\mu_1, \mu_2, ..., \mu_N] = [\mathbb{P}(X_1 = 1), ..., \mathbb{P}(X_1 = N)] \in \mathbb{R}^N$$

is the distribution of X_1.

Example 2.1.18. Suppose that in a restaurant there are three main dishes to eat, the first one is Lebanese, the second one is Mexican, and the third one is Italian.
Assume that 30% of those who eat Lebanese food chooses Italian, 20% chooses Mexican, and 50% eats Lebanese food again the next time. From those who eat Mexican first day, 25% go to Lebanese, 35% go to Mexican, and 40% go to Italian next time. Finally, 80% of those who eat Italian food will eat the next time Italian food, and the percentage of eating the other two dishes is equal. Note that the evolution of the food types chosen by a person can be modeled as a Markov chains. This information can be represented in a transition diagram that shows the transition between the states, where here we have 3 states which are the three main dishes, see Figure 2.1. In the above figure, "L" represents Lebanese food, "M" represents Mexican food, and "I" represents Italian food.

The above Markov chain also can be represented by a transition matrix as follows:

$$P = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.25 & 0.35 & 0.4 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

In the above matrix, "now" is represented by the rows and "next" is represented by the columns. Also, the entry (x, y) represents conditional probability which is next $= y$, knowing that now $= x$, that is the probability to go from x to y.

$$P_{xy} = \mathbb{P}(X_{t+1} = y | X_t = x)$$

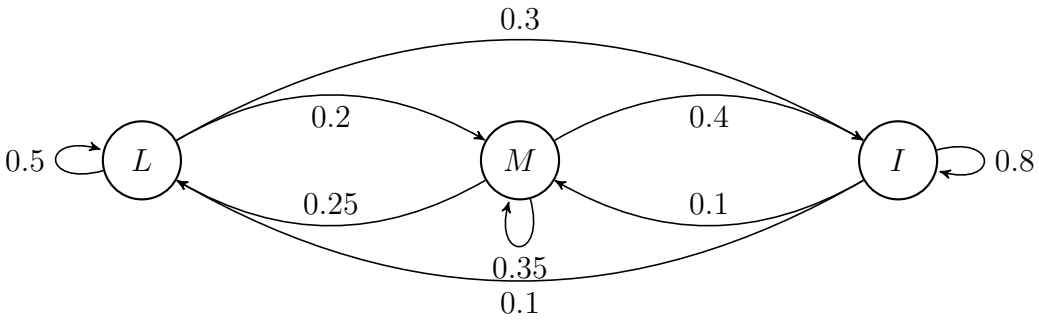


Figure 2.1: Graph of 3 states Markov chain

2.1.3 *Higher Order Transition Probabilities*

From the above explanation and the above example, we can see that the transition matrix only gives us $\mathbb{P}(X_{t+1} = y|X_t = x)$, $x, y \in S$. But what if we want $\mathbb{P}(X_{t+2} = y|X_t = x)$ or $\mathbb{P}(X_{t+n} = y|X_t = x)$ where n is any integer? These data are not given by P, but it can be computed by first step analysis, as follows,

Theorem 2.1.19 (t-step Transition Probabilities)**.** *Let $t \geq 0$. If P denotes the transition matrix of a Markov chain $(X_t)_{t\in\mathbb{N}}$, then the t-step transition probabilities of the Markov chain are represented by the matrix P^t, that is,*

$$\mathbb{P}(X_t = y|X_0 = x) = (P^t)_{xy}$$

More specifically,

$$\mathbb{P}(X_{n+t} = y|X_n = x) = (P^t)_{xy} \qquad \textit{for any } n$$

Proof. Note that if $t = 0$, then the claim is trivially true. The proof is going to be by induction, first I will prove the two-step transition probabilities:

$$\begin{aligned}
\mathbb{P}(X_2 = y|X_0 = x) &= \sum_{k=1}^{N} \mathbb{P}(X_2 = y|X_1 = k, X_0 = x)\,\mathbb{P}(X_1 = k|X_0 = x) \\
&\quad \text{(Law of total probability)} \\
&= \sum_{k=1}^{N} \mathbb{P}(X_2 = y|X_1 = k)\,\mathbb{P}(X_1 = k|X_0 = x) \\
&\quad \text{(Markov Property)} \\
&= \sum_{k=1}^{N} p_{ky}p_{xk} \qquad \text{(By Definitions and time homogeneous)} \\
&= \sum_{k=1}^{N} p_{xk}p_{ky} \qquad \text{(Rearranging)} \\
&= (P^2)_{xy} \qquad \text{(By Definition of Matrix)}
\end{aligned}$$

Hence, the two-step transition probabilities are given by matrix P^2

$$\mathbb{P}(X_2 = y|X_0 = x) = \mathbb{P}(X_{n+2} = y|X_n = x) = (P^2)_{xy} \qquad \text{for any} \qquad n$$

Now, the 3-step transition probabilities:

$$\begin{aligned}
\mathbb{P}(X_3 = y|X_0 = x) &= \sum_{k=1}^{N} \mathbb{P}(X_3 = y|X_2 = k, X_1 = k-1, X_0 = x)\, \mathbb{P}(X_2 = k|X_0 = x) \\
&= \sum_{k=1}^{N} \mathbb{P}(X_3 = y|X_2 = k)\, \mathbb{P}(X_2 = k|X_0 = x) \\
&= \sum_{k=1}^{N} p_{ky}(P^2)_{xk} \\
&= (P^3)_{xy}
\end{aligned}$$

So we get that the 3-step transition probabilities are given by matrix P^3 such that:

$$\mathbb{P}(X_3 = y|X_0 = x) = \mathbb{P}(X_{n+3} = y|X_n = x) = (P^3)_{xy} \qquad \text{for any} \qquad n$$

Now let us apply induction to get the t-step transition. Assume that it is true for $(t-1)$-step transition. Then:

$$\begin{aligned}
\mathbb{P}(X_t = y|X_0 = x) &= \sum_{k=1}^{N} \mathbb{P}(X_t = y|X_{t-1} = k, \ldots, X_0 = x)\, \mathbb{P}(X_{t-1} = k|X_0 = i) \\
&= \sum_{k=1}^{N} \mathbb{P}(X_t = y|X_{t-1} = k)\, \mathbb{P}(X_{t-1} = k|X_0 = x) \\
&= \sum_{k=1}^{N} p_{ky}(P^{(t-1)})_{xk} \\
&= (P^t)_{xy}
\end{aligned}$$

Therefore, the t-step transition probabilities are given by the matrix P^t, that is,

$$\mathbb{P}(X_t = y|X_0 = x) = \mathbb{P}(X_{n+t} = y|X_n = x) = (P^t)_{xy} \qquad \text{for any} \quad n$$

□

Example 2.1.20 (2-step transition probability)**.** Going back to Figure 2.1 let us find:

1. $\mathbb{P}(X_2 = I|X_0 = L) =?$
2. $\mathbb{P}(X_2 = L|X_0 = I) =?$

We have

$$P = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.25 & 0.35 & 0.4 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

and we can find by hand that

$$P^2 = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.25 & 0.35 & 0.4 \\ 0.1 & 0.1 & 0.8 \end{pmatrix} \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.25 & 0.35 & 0.4 \\ 0.1 & 0.1 & 0.8 \end{pmatrix} = \begin{pmatrix} 0.33 & 0.2 & 0.47 \\ 0.2525 & 0.2125 & 0.535 \\ 0.155 & 0.135 & 0.71 \end{pmatrix}$$

Now we can find the result by using P^2, and it will be as follows

1. $\mathbb{P}(X_2 = I | X_0 = L) = [P^2]_{LI} = 0.47$
2. $\mathbb{P}(X_2 = L | X_0 = I) = [P^2]_{IL} = 0.155$

○

2.2 First Step Analysis

In this section, I will show how can we use first step analysis , where we mainly use it in finding hitting probability, return probability and expected return time. For this section I used [2] as a reference

Consider a Markov chain $(X_t)_{t\in\mathbb{N}}$ with state space S and transition matrix P. The hitting time of $F \subseteq S$ is a random variable H_F, given by,

$$H_F = \min\{t \in \{0, 1, 2, \ldots\} : X_t \in F\}$$

In a special case, where the set F consists of only one state that is $F = \{y\}$, we write

$$H_y = \min\{t \in \{0, 1, 2, \ldots\} : X_t = y\}$$

and mainly we will deal with this case.

We use the convention that $H_F = \infty$ if $X_t \notin F$ for all t. The hitting probability h_{xF} of F starting from x is

$$h_{xF} = \mathbb{P}(X_t \in F \text{ for some } t \geq 0 | X_0 = x) = \mathbb{P}(H_F < \infty | X_0 = x)$$

In a special case, where the set F consists of only one state that is $F = \{y\}$, we write

$$h_{xy} = \mathbb{P}(X_t = y \text{ for some } t \geq 0 | X_0 = x) = \mathbb{P}(H_y < \infty | X_0 = x)$$

which is the probability that we hit state y starting from state x.

The expected hitting time δ_{xF} of F starting from x is

$$\delta_{xF} = \mathbb{E}(H_F | X_0 = x)$$

In a special case, where the set F consists of only one state that is $F = \{y\}$, we write

$$\delta_{xy} = \mathbb{E}(H_y | X_0 = x)$$

which is the expected time until we hit state y starting from state x. Clearly we have

$$\delta_{xF} = 0 \qquad \text{for} \qquad x \in F$$

In addition, for all $x \in S \setminus F$, using the first step analysis we have

$$\begin{aligned}
\delta_{xF} &= \mathbb{E}(H_F|X_0 = x) \\
&= \sum_{l\in S} \mathbb{P}(X_1 = l|X_0 = x)(1 + \mathbb{E}(H_F|X_0 = l)) \\
&= \sum_{l\in S} \mathbb{P}(X_1 = l|X_0 = x) + \sum_{l\in S} \mathbb{P}(X_1 = l|X_0 = x)\mathbb{E}(H_F|X_0 = l) \\
&= 1 + \sum_{l\in S} \mathbb{P}(X_1 = l|X_0 = x)\mathbb{E}(H_F|X_0 = l) \\
&= 1 + \sum_{l\in S} P_{xl}\delta_{lF} \qquad x \in S \setminus F
\end{aligned}$$

So we have

$$\delta_{xF} = 1 + \sum_{l\in S} P_{xl}\delta_{lF} \qquad \text{for} \qquad x \in S \setminus F \tag{2.4}$$

Under the boundary conditions

$$\delta_{lF} = \mathbb{E}(H_F|X_0 = l) = 0, \qquad l \in F \tag{2.5}$$

Conditions (2.5) implies that (2.4) becomes

$$\delta_{xF} = 1 + \sum_{l\in S\setminus F} P_{xl}\delta_{lF} \qquad \text{for} \qquad x \in S \setminus F.$$

Proposition 2.2.1. *Given $x \in S$ and $F \subseteq S$, let δ_{xF} denote the expected hitting time of F starting from x. If $x \notin F$, then for all $x \in S \setminus F$, we have:*

$$\delta_{xF} = 1 + \sum_{l\in S\setminus F} P_{xl}\delta_{lF} \qquad \textit{for} \qquad x \in S \setminus F.$$

Note that from above we get that $h_{xx} = 1$ and $\delta_{xx} = 0$ since we are already there but these cases are trivial.

In this case , it becomes more meaningful to look at a random variable representing return time,

$$M_x = \min\{t \in \{1, 2, 3, \ldots\} : X_t = x\}$$

Finally we have the return probability and expected return time

$$\begin{aligned}
m_x &= \mathbb{P}(X_t = x \text{ for some } t \geq 1|X_0 = x) = \mathbb{P}(M_x < \infty|X_0 = x) \\
\mu_{xz} &= \mathbb{E}(M_z|X_0 = x)
\end{aligned}$$

where the expected return time means the expected time for returning to state z after starting from state x. The expected return time can also be computed by first step

analysis. We have:

$$\begin{aligned}
\mu_{xz} &= \mathbb{E}(M_z|X_0 = x) \\
&= \mathbb{P}(X_1 = z|X_0 = x) + \sum_{l\in S, l\neq z} \mathbb{P}(X_1 = l|X_0 = x)(1 + \mathbb{E}(M_z|X_0 = l)) \\
&= P_{xz} + \sum_{l\in S, l\neq z} P_{xl}(1 + \mu_{lz}) \\
&= P_{xz} + \sum_{l\in S, l\neq z} P_{xl} + \sum_{l\in S, l\neq z} P_{xl}\mu_{lz} \\
&= \sum_{l\in S} P_{xl} + \sum_{l\in S, l\neq z} P_{xl}\mu_{lz} \\
&= 1 + \sum_{l\in S, l\neq z} P_{xl}\mu_{lz}
\end{aligned}$$

hence

$$\mu_{xz} = 1 + \sum_{l\in S, l\neq z} P_{xl}\mu_{lz} \qquad x, z \in S. \tag{2.6}$$

Note that for the computation of return probabilities in (2.6) we do not work with any boundary condition. In addition, the time M_x to go back to state x is always at least one by construction, hence $\mu_{xx} \geq 1$ cannot vanish and μ_{xz} cannot vanish too.

$$\delta_{zx} = \mathbb{E}(M_x|X_0 = z) = \mathbb{E}(H_x|X_0 = z) = \mu_{zx}$$

for $x \neq z$. In case $x = z$ the expected return time μ_{zz} can be found from the expected hitting times δ_{lz} for $l \neq z$,as follows

$$\mu_{zz} = 1 + \sum_{l\in S, l\neq z} P_{zl}\delta_{lz},$$

Proposition 2.2.2. *Given $x, z \in S$, let μ_{xz} denote the expected return time to z starting from x. Then,*

$$\begin{aligned}
\mu_{xz} &= 1 + \sum_{l\in S, l\neq z} P_{xl}\mu_{lz} \\
&= 1 + \sum_{l\in S, l\neq z} P_{xl}\delta_{lz}
\end{aligned}$$

2.3 Classification of states

In this section, we will introduce the meaning of communicating states. Also, we will talk about two important notions which are irreducibly and periodicity of Markov chains. These two notions will be important later on in analysing the long run behaviour of Markov chains. In this section I will use [2] as a reference.

2.3.1 *Communicating states*

A state $y \in S$ is said to be accessible from a state $x \in S$ and is represented as

$$x \longrightarrow y$$

if there exists a integer $t \geq 0$ such that:

$$P_{xy}^t = \mathbb{P}(X_t = y | X_0 = x) > 0$$

The above definition means that we can go from state x to state y with a probability greater than zero in a finite number of steps.

Note that since $P^0 = I$, we get from definition that state x is accessible to state x for all states $x \in S$, even if $P_{xx} = 0$.

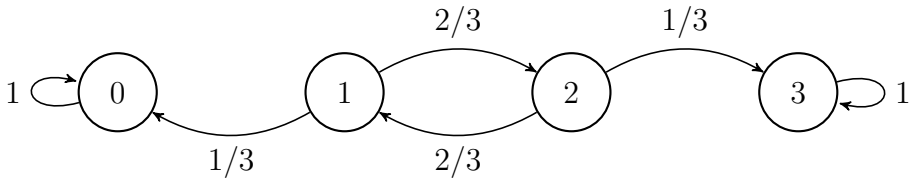
If state x is accessible to state y, and vice versa, then we say that the two states communicate, and write it as $x \longleftrightarrow y$.

Note that this relation $\longleftrightarrow$ is an equivalence relation. Therefore, it induces a partition of S into disjoint subsets $K_1, ..., K_k$ such that $S = K_1 \cup ... \cup K_k$.

In other words, the states of the Markov chain can be divided into communication classes. Each class contains only members communicating with each other. That is, two states x and y belong to the same class if and only if $x \longleftrightarrow y$.

Definition 2.3.1 (Reducibility)**.** A Markov chain is called irreducible if all states communicate with each other, otherwise it is reducible.

Example 2.3.2 (Reducible and Irreducible Markov Chains)**.** Going back to our Markov chain in Figure 2.1, we see that all the states communicate with each other so they form a unique communicating class. In other words, the Markov chain of Figure 2.1 is irreducible. Here is another example where the Markov chain is reducible.



We have 3 communicating states. Namely, states 1 and 2 communicate with each other, but they do not communicate with the other states in the graph. Moreover, states 0 and 3 do not communicate with any other state. So, the communication classes are $K_1 = \{1, 2\}, K_2 = \{0\}$, and $K_3 = \{3\}$. ◯

2.3.2 *Periodicity and Aperiodicity*

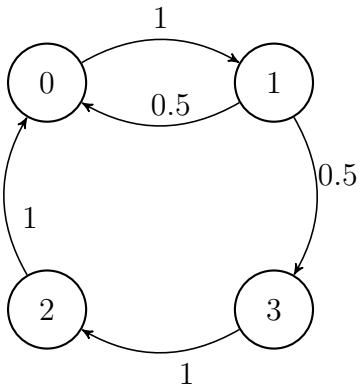
The period of a state $x \in S$ is:

$$d(x) = \gcd\{t \in \mathbb{Z}^+ : P_{xx}^t > 0\}$$

State x is aperiodic if $d(x) = 1$ and periodic if $d(x) > 1$.

A Markov chain is called aperiodic if all its states are aperiodic.

Example 2.3.3 (Periodic and Aperiodic Markov Chains)**.** Back to our Markov chain in Figure 2.1, we see that all the states have period 1 so the Markov chain is aperiodic. Here is another example where the Markov chain is periodic. Consider the Markov chain:



$$\{t \geq 1 : P_{00}^t > 0\} = \{2, 4, 6, \ldots\}$$
$$\{t \geq 1 : P_{11}^t > 0\} = \{2, 4, 6, \ldots\}$$
$$\{t \geq 1 : P_{22}^t > 0\} = \{2, 4, 6, \ldots\}$$
$$\{t \geq 1 : P_{33}^t > 0\} = \{2, 4, 6, \ldots\}$$

so all the states are of period 2. ◯

2.4 Long-Run Behavior of Markov chains

In this section, we will deal with the long time behaviour of Markov chains. We will talk about stationary and limiting distribution. We are going to use the notations of the previous section. In this chapter, I used [1] and [2] as my references.

2.4.1 *Stationary Distribution*

Let $(X_t)_{t\in\mathbb{N}}$ be a Markov chain with state space $S = \{1, 2, ..., N\}$ and transition matrix P. Each X_t is a random variable. Thus, it has a probability distribution. It will be a vector of non-negative probabilities that sum up to 1 We can represent the probability distribution of X_t as an $1 \times N$ vector.

Consider X_0 and let π be a $1 \times N$ denoting the probability distribution of X_0.

$$\pi = \big(\pi_1, \pi_2, ..., \pi_N\big) = \big(\mathbb{P}(X_0 = 1), \mathbb{P}(X_0 = 2), ..., \mathbb{P}(X_0 = N)\big)$$

More specifically, the distribution of an S-valued random variable X will be represented by the vector

$$\pi = (\pi_1, \pi_2, \ldots, \pi_N)$$

where $\pi_k = \mathbb{P}(X = k)$. We write $X \sim \pi$ to indicate that X is distributed according to π.

Probability distribution after one step. Suppose $X_t \sim \pi$. Then, using the law of total probability, we have

$$\begin{aligned}\mathbb{P}(X_{t+1} = y) &= \sum_{k=1}^{N} \mathbb{P}(X_{t+1} = y | X_t = x)\, \mathbb{P}(X_t = x) \\ &= \sum_{k=1}^{N} P_{xy}\pi_x \\ &= \sum_{k=1}^{N} \pi_x P_{xy} \\ &= (\pi P)_y\end{aligned}$$

Then, we conclude that $X_{t+1} \sim \pi P$.

Probability distribution after t steps. Suppose $X_0 \sim \pi$. Then, using the law of total probability, we have

$$\begin{aligned}\mathbb{P}(X_t = y) &= \sum_{k=1}^{N} \mathbb{P}(X_t = y | X_0 = x)\, \mathbb{P}(X_0 = x) \\ &= \sum_{k=1}^{N} P^t_{xy}\pi_x \\ &= \sum_{k=1}^{N} \pi_x P^t_{xy} \\ &= (\pi P^t)_y\end{aligned}$$

Then, we conclude that $X_t \sim \pi P^t$.

So we can summarize the results as follows: If $X_0 \sim \pi$, then

$$\begin{aligned} X_1 &\sim \pi P \\ X_2 &\sim \pi P^2 \\ &\vdots \\ X_t &\sim \pi P^t \end{aligned}$$

Proposition 2.4.1. *Let $(X_t)_{t\in\mathbb{N}}$ be a Markov chain with $N \times N$ transition matrix P. If the probability distribution of X_0 is given by $1 \times N$ row vector π, then the probability distribution of X_t is given by the $1 \times N$ row vector as follows:*

$$X_0 \sim \pi \implies X_t \sim \pi P^t$$

Definition 2.4.2 (Stationary Distribution)**.** Given a Markov chain $(X_t)_{t\in\mathbb{N}}$ with state space S and transition matrix P. Let $\pi = (\pi_x)$ be a distribution on S, such that $\sum_{x\in S} \pi_x = 1$ where $\pi_x \geq 0$ for all $x \in S$. We say that π is a stationary distribution if it satisfies

$$\pi_y = \sum_{x\in S} \pi_x p_{xy} \qquad \text{for all} \qquad y \in S$$

which means that $\pi = \pi P$. π is stationary means that if the current state is distributed according to π, then the following state will also be distributed according to π.

Example 2.4.3 (Finding Stationary Distribution)**.** Let us go back to our example Figure 2.1 and find its stationary distributions.

We want to find a stationary distribution π, so let us solve $\pi = \pi P$.

$$\big(\pi_1, \pi_2, \pi_3\big) = \big(\pi_1, \pi_2, \pi_3\big) \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.25 & 0.35 & 0.4 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

So we get:

$$\begin{aligned} \pi_1 &= 0.5\pi_1 + 0.25\pi_2 + 0.1\pi_3 \\ \pi_2 &= 0.2\pi_1 + 0.35\pi_2 + 0.1\pi_3 \\ \pi_3 &= 0.3\pi_1 + 0.4\pi_2 + 0.8\pi_3 \end{aligned}$$

Knowing that π is a distribution, it satisfies the normalising condition

$$\pi_1 + \pi_2 + \pi_3 = 1$$

By calculation we get: $\pi_1 = 0.21$, $\pi_2 = 0.16$ and $\pi_3 = 0.63$.

So,

$$\pi = (\pi_1, \pi_2, \pi_3) = (0.21, 0.16, 0.63)$$

In particular, we find that this Markov chain has exactly one stationary distribution. ◯

2.4.2 *Existence and Uniqueness of Stationary Distribution*

We are going to prove the existence and uniqueness of the stationary distribution for every irreducible finite-state Markov chains.

Theorem 2.4.4 (Existence)**.** *Every finite-state Markov chain has a stationary distribution.*

Proof. Let ν be a vector satisfying $\nu P = \nu$ but here we are not assuming that it has to sum to 1.

Assume that $(X_t)_{t\in\mathbb{N}}$ is a Markov chain. First, let's try to find a stationary vector. Let z be a fixed initial state, and let ν_x be the expected number of visits to x before going back to z, which is

$$\begin{aligned} v_x &= \mathbb{E}(\text{number of visits to } x \text{ before returning to } z | X_0 = z) \\ &= \mathbb{E}\sum_{t=1}^{M_z} \mathbb{P}(X_t = x | X_0 = z) \\ &= \sum_{t=1}^{\infty} \mathbb{P}(X_t = x \text{ and } t \leq M_z | X_0 = z) \end{aligned}$$

where M_z represents the return time Definition 2.2, note that $\nu_z = 1$ since only z returns to z . ν counts the number of visits in random time to the states, it looks like that when we normalize ν it becomes a stationary distribution, that is ν represents a stationary vector.

Let us prove that $\sum_x \nu_x p_{xy} = \nu_y$, so we get:

$$\begin{aligned} \sum_{x\in S} \nu_x P_{xy} &= \sum_{x\in S}\sum_{t=1}^{\infty} \mathbb{P}(X_t = x \text{ and } t \leq M_z | X_0 = z) P_{xy} \\ &= \sum_{t=1}^{\infty}\sum_{x\in S} \mathbb{P}(X_t = x \textit{ and } X_{t+1} = y \text{ and } t \leq M_z | X_0 = z) \\ &= \sum_{t=1}^{\infty} \mathbb{P}(X_{t+1} = y \textit{ and } t \leq M_z | X_0 = z) \end{aligned}$$

Now let us change the visit to z at time M_z with the visit to z at time 0, which allows us to count the visits from 0 to $M_z - 1$ instead of 1 to M_z, so we get

$$\begin{aligned} \sum_{x\in S} \nu_x P_{xy} &= \sum_{t=0}^{\infty} \mathbb{P}(X_{t+1} = y \text{ and } t \leq M_z - 1 | X_0 = z) \\ &= \sum_{t+1=1}^{\infty} \mathbb{P}(X_{t+1} = y \text{ and } t+1 \leq M_z | X_0 = z) \\ &= \sum_{t=1}^{\infty} \mathbb{P}(X_t = y \text{ and } t \leq M_z | X_0 = z) \\ &= \nu_y \end{aligned}$$

this gives us that ν is a stationary vector.

It only remains to normalize ν into stationary distribution by dividing by $\sum_x \nu_x$, which we can do it because it is finite , as we know $\sum_x \nu_x$ is the expected total number of visits to all states before going back to z, in other words it is the expected return time μ_{zz}, so $\pi = (\frac{1}{\mu_{zz}})\nu$ is a stationary distribution. □

Note that this theorem is always true in this chapter because we are only dealing with finite state Markov chains but in general it is not true and it needs additional condition which is positive recurrence. If we did not restrict ourselves to Markov chains with finite state space, then we could still prove the following:

Theorem 2.4.5. *A Markov chain which is positive recurrent has a stationary distribution.*

The proof of this theorem is done in [2, Section 10.4].
The definition of positive recurrent state stated in Appendix 4.2.

Theorem 2.4.6 (Uniqueness)**.** *A finite-state Markov chain is irreducible. then its stationary distribution is unique and it is given by* $\pi_x = \frac{1}{\mu_{xx}}$ *where* μ_{xx} *denotes the expected return time to* X *starting from* x*. (see Section 2.2)*

Proof. By the previous theorem, we know that our Markov chain has at least one stationary distribution. Let us assume that the Markov chain is irreducible and that π is one of its stationary distribution. Our aim is to show that $\pi_x = \frac{1}{\mu_{xx}}$ for all x.

We are going to use this form of μ_{zz} from Proposition 2.2.2

$$\mu_{zz} = 1 + \sum_{y \neq z} P_{zy}\delta_{yz} \tag{2.7}$$

and this form of δ_{xz} from Proposition 2.2.1 by assuming that $F = \{z\}$

$$\delta_{xz} = 1 + \sum_{y} P_{xy}\delta_{yz} \qquad \text{for all} \qquad x \neq z \tag{2.8}$$

Now let multiply by π_x and sum over all $x \neq z$ equation (2.8) so we get,

$$\sum_{x} \pi_x\delta_{xz} = \sum_{x \neq z} \pi_x + \sum_{y}\sum_{x \neq z} \pi_x p_{xy}\delta_{yz} \tag{2.9}$$

Note that on the left we are summing over x since $\delta_{xx} = 0$, now let us multiply (2.7) π_z so we get

$$\pi_z\mu_{zz} = \pi_z + \sum_{y} \pi_z p_{zy}\delta_{yz} \tag{2.10}$$

Now if we add (2.9) and (2.10), so we get

$$\sum_{x} \pi_x\delta_{xz} + \pi_z\mu_{zz} = \sum_{x \neq z} \pi_x + \sum_{y}\sum_{x \neq z} \pi_x p_{xy}\delta_{yz} + \pi_z + \sum_{y} \pi_z p_{zy}\delta_{yz}$$

$$\sum_{x} \pi_x\delta_{xz} + \pi_z\mu_{zz} = \sum_{x} \pi_x + \sum_{y}\sum_{x} \pi_x p_{xy}\delta_{yz}$$

Also we are going to use $\sum_x \pi_x p_{xy} = \pi_y$ with $\sum_x \pi_x = 1$ so we finally get

$$\sum_{x} \pi_x\delta_{xz} + \pi_z\mu_{zz} = 1 + \sum_{y} \pi_y\delta_{yz}$$

But notice that $\sum_x \pi_x\delta_{xz} = \sum_y \pi_y\delta_{yz}$, and since the Markov chain is irreducible so they are finite. so we get $\pi_z = \frac{1}{\mu_{zz}}$ for all z. □

Note that the positive recurrence is needed here also if it is the case where the Markov chain is not of finite state space.

2.4.3 *Convergence to Equilibrium*

In this part we are going to see what happens to a Markov chain $(X_t)_{t\in\mathbb{N}}$ in the long run, that means when t tends to infinity.

Proposition 2.4.7 (Equilibrium Distribution)**.** *Suppose that $(X_t)_{t\in\mathbb{N}}$ is a Markov chain with finite state space S and transition matrix P. Assume that we have a distribution $\pi^* = \pi^*_x$ on S. Assume that for every $y \in S$, the limit $\pi_y = \lim_{t\to\infty} \mathbb{P}(X_t = y)$ exists. Then $\pi = (\pi_y)_{y\in S}$ is a stationary distribution.*

Theorem 2.4.8 (Convergence theorem)**.** *Let $(X_t)_{t\in\mathbb{N}}$ be an irreducible and aperiodic finite-state Markov chain, and let π denote its unique stationary distribution. Then, as $t \to \infty$, the distribution of X_t converges to π. In other words, for every state y, we have $\lim_{t\to\infty} \mathbb{P}(X_t = y) = \pi_y$.*

Proof. To prove this theorem we are going to use the coupling method, where during this method we take two random random variables X and Y that are dependent to each other.

Let $(X_t)_{t\in\mathbb{N}}$ be an irreducible and aperiodic Markov chain with transition matrix P and initial distribution λ. Let $(Y_t)_{t\in\mathbb{N}}$ be a Markov chain also with transition matrix P but "in equilibrium", which is if we start from the stationary distribution π, then we will be in this distribution forever.

Pick a state $s \in S$, and assume that $X_t = Y_t = s$ for the first time at T, and if this never happens, then $T = \infty$. Now let us apply the coupling method, after T, When both chains (X_t) and (Y_t) meet at s, then make (X_t) stick with (Y_t), $X_t = Y_t$ for $t \geq T$. Since the Markov chain is memory less, $(X_{T+t}) = (Y_{T+t})$ remains a Markov chain with the same transition probabilities from that time on. By using the coupling, as we last after T, (X_t) will have distribution π always, and this is going to be so useful in this proof.

Define $(Z_t)_{t\in\mathbb{N}}$ as $Z_t = (X_t, Y_t)$. So $(Z_t)_{t\in\mathbb{N}}$ is a Markov chain with state space $S \times S$, and T is the expected hitting time of (Z_t) to state $(s, s) \in S \times S$. The transition probabilities of (Z_t) are $\underline{P} = (P_{xy}P_{zl}$. This is the probability that the chain goes from $Z_t = (X_t, Y_t) = (x, z)$ to $Z_{t+1} = (X_{t+1}, Y_{t+1}) = (y, l)$. Since the given Markov chain is irreducible and aperiodic, this gives us that $P_{xy}, P_{zl} > 0$, this allows us to say that (Z_t) is irreducible. In addition, (Z_t) has a stationary distribution $\underline{\pi} = \pi_x\pi_z$. Thus, T is finite with probability 1.
Now we are ready to prove the limit theorem. We want to show that $\mathbb{P}(X_t = x) \to \pi_x$. The difference between them is

$$\begin{aligned}
|\mathbb{P}(X_t = x) - \pi_x| &= \mathbb{P}(t \leq T) \times |\mathbb{P}(X_t = x)|t \leq T) - \pi_x| + \mathbb{P}(t > T) \times |\pi_x - \pi_x| \\
&= \mathbb{P}(t \leq T) \times |\mathbb{P}(X_t = x)|t \leq T) - \pi_x| \\
&\leq \mathbb{P}(t \leq T).
\end{aligned}$$

Here, the equality on the first line is because (X_t) follows the stationary distribution exactly once it sticks to (Y_t) after time T, and the inequality on the third line is because the absolute difference between two probabilities is between 0 and 1. But we proved that T is finite with probability 1, thus $\mathbb{P}(t \leq T) = \mathbb{P}(t \geq T) \to 0$. □

Example 2.4.9 (Limiting behavior)**.** Let us examine the long-term behavior of the Markov chain in Figure 2.1. Recall that in Example 2.4.3 we found that the unique

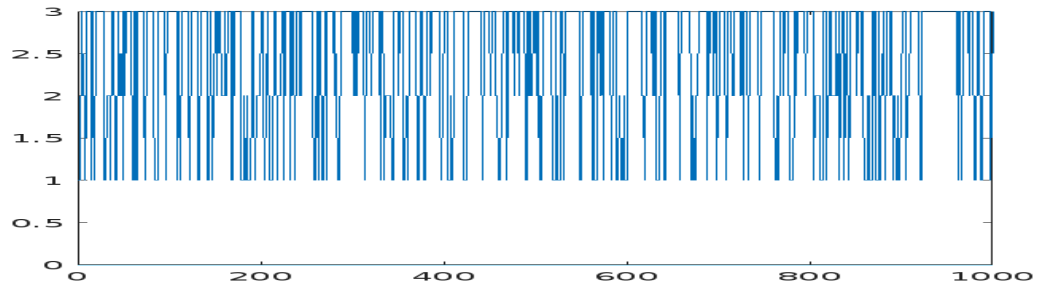
stationary distribution of this Markov chain is $\pi = (0.21, 0.16, 0.63)$. The Markov chain represented by Figure 2.1 is irreducible and aperiodic. So, the above Theorem 2.4.8 is satisfied. Thus, the t-step transition probabilities tend to π. I used the following Matlab code to simulate our Markov chain.

```
function K = markov(M,m)
    n = length(M(1,:));
    state = randi([1, n]); % initial state
    state = 1;
    y = zeros(m);
    x = 1:m;
    y(1)=state;

    for j=2:m
        prob = rand(1);
        for i=1:n
            if prob <= sum(M(state, 1:i))
                y(j)= i;
                state = i;
                break;
            end
        end
    end
    K = y(m);
    plot(x,y);
end
```

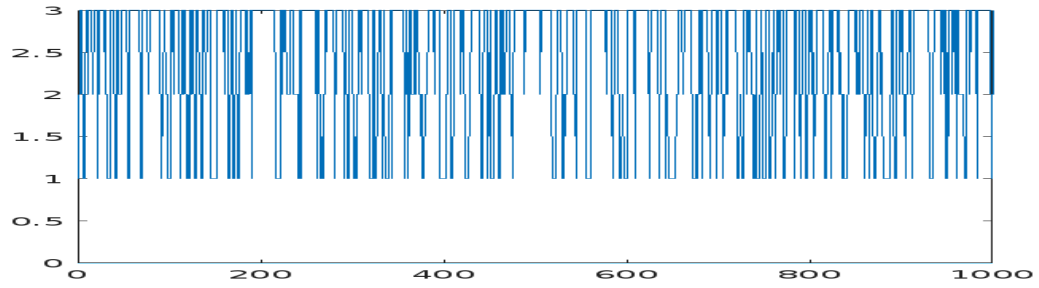
Figure 2.2 shows 5 sample trajectories of the Markov chains during 1000 steps. In addition, I wrote the following code using Matlab to give us the empirical distribution of the Markov chain after a fixed number of steps.

```
function v = results(M,m,n)
    v = zeros(n,1);
    for i=1:n
        k = markov(M,m);
        v(i)=k;
    end

    x = unique(v);
    N = numel(x);
    count = zeros(N,1);
    for f = 1:N
        count(f) = sum(v==x(f));
    end
    disp([ x(:) count ]);
end
```

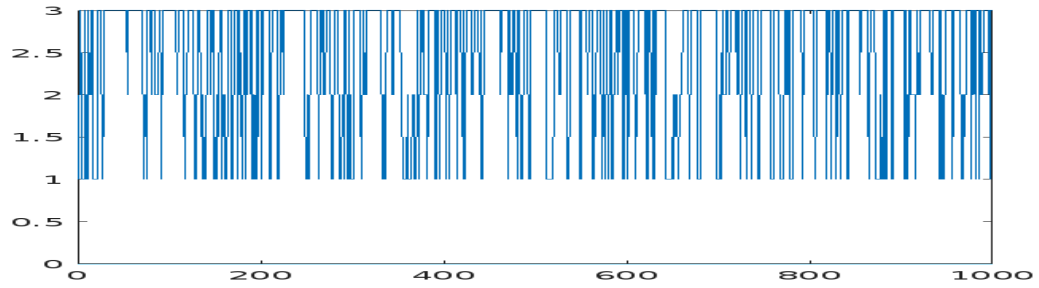
Based on the first run, the distribution at time 1000 step for the Markov chain in



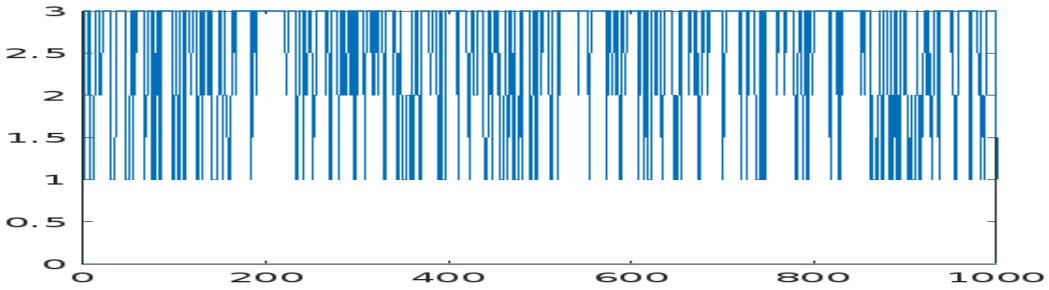
(a) Graph 1



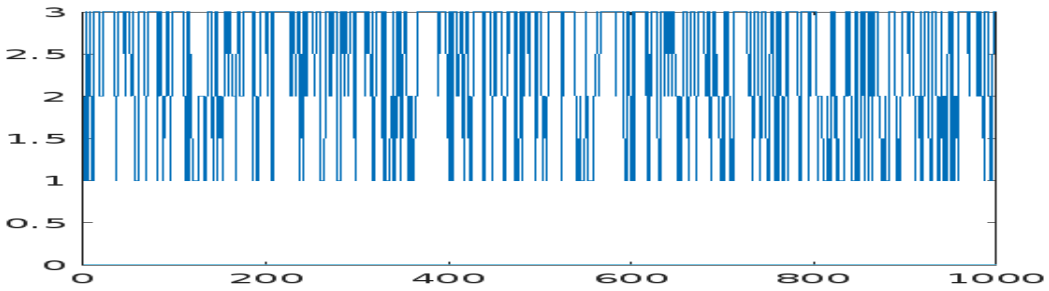
(b) Graph 2



(c) Graph 3



(d) Graph 4



(e) Graph 5

Figure 2.2: Graph of our three-state Markov chain.

Figure 2.1 is approximately this

$$(0.208, 0.168, 0.624)$$

And based on the second run, the distribution at time 1000 is approximately this

$$(0.196, 0.17, 0.634)$$

The results we got is approximately the same as the stationary distribution. ◯

Definition 2.4.10 (Ergodicity)**.** Let P be the transition matrix of a finite-state Markov chain $(X_t)_{t\in\mathbb{N}}$ with state space S. We say that P is ergodic if

1. P has a unique stationary distribution π.
2. For every probability distribution μ on S and every state $x \in S$, we have $(\mu P^t)_x \to \pi_x$ as $t \to \infty$.

Observe that Theorem 2.4.8 can be reformulated by saying that every irreducible and aperiodic finite-state Markov chain is ergodic.

The convergence theorem showed the limit of $\mathbb{P}(X_t = y)$, that is the probability of the Markov chain being in state y at time t at some specific point in time t a long time in the future. Now we will look at the long-run amount spent in state y. This amount is the averaging the behaviour over a long time period.

Consider

$$V_y(T) := \#\{t < T : X_t = y\}$$

for the total number of visits to state y up to time T. Then the proportion of time steps up to time t spent in y is $V_y(t)/t$. The limiting value of the latter (if it exists) is the long-run proportion of time steps spent in state y.

Theorem 2.4.11 (Empirical Distribution)**.** *Assume that we have an irreducible finite-state Markov chain $(X_t)_{t\in\mathbb{N}}$. Let π be the unique stationary distribution of the Markov chain. For any initial distribution λ, we have $V_y(t)/t \to \pi_y$ almost surely as $t \to \infty$.*

Note that almost sure convergence means that $\mathbb{P}(V_y(t)/t \to \pi_y) = 1$.

Proof. The proof of this theorem is done in [2, Section 11.4]. □

Chapter 3

CELLULAR AUTOMATA

In this chapter, I will first start by giving intuition about cellular automata. Then I will give preliminaries needed to define a cellular automaton. I will distinguish between two types of cellular automata: The deterministic cellular automata (DCA) and the probabilistic one (PCA). Then I will introduce the notions of ergodicity and uniform ergodicity. I will end the chapter with some equivalent definitions of uniform ergodicity.

3.1 Informal Description

A cellular automaton consists of a collection of entities, called cells, that are arranged on a lattice. Each cell has a certain state chosen from a finite set. At every discrete time step, the states are updated based on some rule that takes into account the states of a few neighboring cells.

Example 3.1.1 (A deterministic cellular automaton)**.** Let us consider a cellular automaton with a one-dimensional lattice whose cells have two possible states: 0 and 1. A possible configuration of the cellular automaton is the following:

$$...00101010100110101... \tag{3.1}$$

At every time step, the state of each cell is updated according to the following rules:

1. If a cell has state 0 and its right neighbor has state 0, then the cell remains in state 0.
2. If a cell has state 0 and its right neighbor has state 1, then the cell changes to state 1.
3. If a cell has state 1 and its right neighbor has state 0, then the cell remains in state 1.
4. If a cell has state 1 and its right neighbor has state 1, then the cell changes to state 0.

And If we applied the rule to the above configuration, we get:

$$\cdots 00101010100110101 \cdots$$
$$\cdots 0111111110101111 \ \cdots$$

The dynamical system obtained by repeatedly updating the cells, at discrete time steps, according to the above rules is an example of a deterministic cellular automaton. ◯

In the next example we are going to illustrate the idea of probabilistic cellular automata (PCA). A PCA is similar to a DCA except that the update rule is probabilistic.

Example 3.1.2 (A probabilistic cellular automaton)**.** Let us modify the previous example to turn it into a probabilistic cellular automaton. The update rules are as follows:

1. If the current state is 0 and its right neighbor has state 0, then with probability 0.75 the cell remains in state 0 and with probability 0.25, it changes to state 1.

2. If the current state is 0 and its right neighbor has state 1, then with probability 0.3 the cell remains in state 0 and with probability 0.7, it changes to state 1.

3. If the current state is 1 and its right neighbor has state 0, then with probability 0.6 the cell changes to state 0 and with probability 0.4 it remains in state 1.

4. If the current state is 1 and its right neighbor has state 1, then with probability 0.9 the cell changes to state 0 and with probability 0.1 it remains in state 1.

Different cells are updated independently of one another.

This probabilistic update rule can be summarized in a stochastic matrix as in Table 3.1.

	0	1
00	0.75	0.25
01	0.3	0.7
10	0.6	0.4
11	0.9	0.1

Table 3.1: Probabilistic update rule for the PCA in Example 3.1

We can simulate a probabilistic rule by drawing, for example, a number at random from the interval $[0, 1]$.For example, the following seven numbers were drawn uniformly at random from the interval $[0, 1]$.

$$\begin{pmatrix} 0.2412 \\ 0.2676 \\ 0.0353 \\ 0.2706 \\ 0.0444 \\ 0.2840 \\ 0.2152 \end{pmatrix}$$

Let us look at the third cell of the configuration (3.1). To update this cell based on the above rule, we generate a random number from a uniform distribution on $[0, 1]$. Suppose that we would get the following number 0.2412 from such a random draw. Since this number is less than 0.4 then it will remain in state 1.However, if the random number happens to be greater than 0.6 the cell changes to 0. ◯

Note that every DCA can be viewed as a PCA in which the transition probabilities are either 0 or 1. For instance, viewed as a PCA, the stochastic matrix corresponding to the DCA in Example 3.1.1 will be as in Table 3.2.

	0	1
00	1	0
01	0	1
10	0	1
11	1	0

Table 3.2: Probabilistic update rule for the DCA in Example 3.1

3.2 Preliminaries

I will start by giving some definitions needed to give mathematical and formal definitions of cellular automata.

Definition 3.2.1 (Configuration)**.** Assume that we have a finite alphabet with 2 elements or more and call it Σ, and let $d \geq 1$ be an integer. A *configuration* of the d-dimensional *lattice* $\mathbb{Z}^d$ is a map $x\colon \mathbb{Z}^d \to \Sigma$ which assigns a symbol x_i from Σ to each *site* i of $\mathbb{Z}^d$.

The symbol x_i is referred to as the *cell* of cell i. The restriction of a configuration x to a set $A \subseteq \mathbb{Z}^d$ is denoted by x_A.

Definition 3.2.2 (Pattern)**.** A *pattern* is a partial configuration $p\colon D \to \Sigma$ where $D \subseteq \mathbb{Z}^d$ is finite. The set D is called the *shape* of the pattern p.

Definition 3.2.3 (Shift Operator)**.** Given $k \in \mathbb{Z}^d$, the *shift-by-k* operator is the map $\sigma^k\colon \mathcal{X} \to \mathcal{X}$ defined by $\sigma^k(x)_i := x_{k+i}$ for $i \in \mathbb{Z}^d$.

Definition 3.2.4 (Neighborhood)**.** Let $N \Subset \mathbb{Z}^d$. Given $A \subseteq \mathbb{Z}^d$, we define $N(A) := A + N = \{a + n : a \in A \text{ and } n \in N\}$. If $k \in \mathbb{Z}^d$, we write $N(k)$ instead of $N(\{k\})$.

Definition 3.2.5 (Cylinder set)**.** A *cylinder* set $p\colon D \to \Sigma$ is a set of the form $[p] := \{x \in \mathcal{X} : x_D = p\}$ where p is a pattern with shape D.

The space $\Sigma^{\mathbb{Z}^d}$ of all configurations can be given the product topology. The product topology is the topology generated by the cylinder sets. This is a compact and metrizable topology. The proof of being compact and metrizable is in this book [3].

Now, assume that $\mathcal{X}$ is a compact metric space. Let us equip $\mathcal{X}$ with the Borel σ-algebra $\mathscr{F}$ and denote the set of all probability measures on $(\mathcal{X}, \mathscr{F})$ by $\mathscr{P}(\mathcal{X})$. The set $\mathscr{P}(\mathcal{X})$ can be made into a topological space by equipping it with a weak topology.

Definition 3.2.6 (Weak topology)**.** Recall that every continuous real-valued function on a compact metric space is bounded. Recall also that continuous functions are measurable with respect to the Borel σ-algebra. $C(\mathcal{X})$ is a Banach space of continuous real valued function on $\mathcal{X}$ with uniform norm which is the supreme norm.

Define a topology on $\mathscr{P}(\mathcal{X})$ (smallest topology) that makes this map $\mu \in \mathscr{P}(\mathcal{X}) \mapsto \int f d\mu$ continuous for every continuous function f. The weak topology on $\mathscr{P}(\mathcal{X})$ is defined as the smallest topology that makes the map $\mu \in \mathscr{P}(\mathcal{X}) \mapsto \int f d\mu$ continuous for every $f \in C(\mathcal{X})$. The weak topology on $\mathscr{P}(\mathcal{X})$ is compact and metrizable. The proof of being compact and metrizable is in section 6.1 in this book [4].

Note that convergence in the weak topology is known as weak convergence.

Definition 3.2.7 (Transition Kernels)**.** Let $(\mathcal{X}, \mathscr{F})$ be an arbitrary measurable space. A function $\Phi : \mathcal{X} \times \mathscr{F} \to [0,1]$ is said to be a transition kernel on $\mathcal{X}$ if it has the following properties:

1. $\Phi(x, \cdot)$ is a probability measure on $\mathcal{X}$ for each $x \in \mathcal{X}$.
2. $\Phi(\cdot, E)$ is measurable for each $E \in \mathscr{F}$.

Transition kernels play a similar role as transition matrices. In other words, in the same fashion that the law of transitions of a finite-state Markov chain can be described by its transition matrix, the law of transitions of a Markov process on a general measurable space can be described by a transition kernel.

Definition 3.2.8 (Markov Process with Transition Kernel Φ)**.** Let Φ be a transition kernel on a measurable space $(\mathcal{X}, \mathscr{F})$. A Markov process with transition kernel Φ is a sequence $(X_t)_{t\in\mathbb{N}}$ of random variables with values in $\mathcal{X}$, such that for every $E \in \mathscr{F}$ and $t \in \mathbb{N}$, we have

$$\mathbb{P}(X_{t+1} \in E | X_0, X_1, \ldots, X_t) = \Phi(X_t, E)$$

almost surely.

This generalizes the notion of a (time-homogeneous) Markov chain.
We are going to give some notations in order to make our work easier later on.

1. Given a probability measure μ on $\mathcal{X}$, we denote by $\mu\Phi$ the probability measure on $\mathcal{X}$ defined by $(\mu\Phi)(E) = \int \Phi(x, E)\mu(dx)$ for all $E \in \mathscr{F}$.

   The meaning of this notation can be understood as follows: let X be the current state of the process and Y be the state of the process after one step. If X has distribution μ, then Y has distribution $\mu\Phi$.

2. Given a bounded measurable function $f : \mathcal{X} \to \mathbb{R}$, we denote by Φf the (bounded measurable) function defined by $(\Phi f)(x) = \int f(y)\Phi(x, dy)$ for every $x \in \mathcal{X}$.

   The meaning of this notation can be understood as follows: let X be the current state of the process and Y be the state of the process after one step. Then, $(\Phi f)(x)$ is the expected value of $f(Y)$ conditioned on $X = x$.

Intuitively, if Φ is a transition kernel describing the 1-step transitions of a Markov process, then Φ^n is the transition kernel describing the n-step transitions of the same Markov process. More precisely, if Φ is a transition kernel on a measurable space $(\mathcal{X}, \mathscr{F})$, then for $n \in \mathbb{N}$, we can define Φ^n inductively as follows.

1. $\Phi^0(x, E) = \delta_x(E)$.
2. $\Phi^{n+1}(x, E) = \int \Phi(y, E)\Phi^n(x, dy)$.

Definition 3.2.9 (Feller Property)**.** Suppose that $\mathcal{X}$ is a compact metric space with the Borel σ-algebra, and let Φ be a transition kernel on $\mathcal{X}$. We say that Φ has the Feller property (or is a Feller kernel) if one of the following equivalent conditions is satisfied:

- The map $\mu \to \mu\Phi$ on $\mathscr{P}(\mathcal{X})$ is continuous.
- If $f \in C(\mathcal{X})$ then $\Phi f \in C(\mathcal{X})$.

Definition 3.2.10 (Full support)**.** We say that a Borel probability measure μ on $\mathcal{X}$ has full support if $\mu(B) > 0$ for all open $B \neq \emptyset$. We say that a transition kernel Φ on $\mathcal{X}$ has full support if the measure $\Phi(x, \cdot)$ has full support for every $x \in \mathcal{X}$.

3.3 Cellular Automata

In this section I will give the formal definitions of deterministic and probabilistic cellular automata. In the previous section, $\mathcal{X}$ was an arbitrary compact metric space. But here, $\mathcal{X}$ is supposed to be the space of configurations.

Let Σ be a finite alphabet and $\mathcal{X} = \Sigma^{\mathbb{Z}^d}$.

Definition 3.3.1 (DCA)**.** A map $F : \mathcal{X} \to \mathcal{X}$ is a cellular automaton if there exists a finite set $N \Subset \mathbb{Z}^d$ and a function $f : \Sigma^N \to \Sigma$ such that $F(x)_i = f(\sigma^i(x)_N)$ for every $x \in \mathcal{X}$ and $i \in \mathbb{Z}^d$. The map F is known as the global transition map of the cellular automaton , f is the local rule of the cellular automaton and N is the neighborhood of the cellular automaton.

Definition 3.3.2 (PCA)**.** A *probabilistic cellular automaton (PCA)* is a Markov process on $\mathcal{X}$ with a local shift-invariant transition kernel. Namely, let $N \Subset \mathbb{Z}^d$ be a finite set of sites where $d \in \mathbb{Z}^+$. Any stochastic matrix $\varphi\colon \Sigma^N \times \Sigma \to [0, 1]$ defines a transition kernel Φ on $\mathcal{X}$ by

$$\Phi(x, [y_A]) \coloneqq \prod_{k \in A} \varphi(\sigma^k(x)_N, y_k)$$

for every configuration $x \in \mathcal{X}$ and every finite pattern $y_A \in \Sigma^A$. We call φ the *local* transition kernel and Φ the *global* transition kernel of the PCA.

The following theorem is known as the Curtis–Hedlund–Lyndon theorem.

Theorem 3.3.3 (DCA)**.** *Every DCA map is continuous and commutes with the shifts (i.e., $F \circ \sigma^k = \sigma^k \circ F$ for every $k \in \mathbb{Z}^d$). Conversely, every continuous map $F : \mathcal{X} \to \mathcal{X}$ that commutes with the shifts is a CA.*

The proof of the theorem can be found in [3, Chapter 5].

Theorem 3.3.4 (PCA)**.** *The global transition kernel of every PCA has the Feller property and satisfies $\Phi((\sigma^k x), E) = \Phi(x, \sigma^{-k} E)$ for every x, E and k.*

The proof of the theorem can be found in [5].

Now I am going to show how DCA can be viewed as a special type of PCA. Note that a DCA is identified by its global transition map F, whereas a PCA is identified by its global transition kernel. What is the global transition kernel Φ associated to a DCA F?

For $x \in \mathcal{X}$, denote by δ_x the Dirac measure concentrated on the configuration x. The global transition kernel of a DCA F can be defined as $\Phi(x, \cdot) = \delta_{F(x)}$. In particular, $\Phi^n(x, \cdot) = \delta_{F^n(x)}$. Moreover if f is the local rule of the DCA, then the associated local transition kernel φ is either 0 or 1.

3.4 Ergodicity and Uniform Ergodicity

Recall that in the previous chapter we defined the notions of ergodicity for time-homogeneous and finite-state Markov chain in Definition 2.4.10. Now I am going to give a more general notion of ergodicity for Feller transition kernels on compact metric spaces. But in fact, here we have two non-equivalent notions: ergodicity and uniform ergodicity.

Definition 3.4.1 (Invariant Measure)**.** Suppose that $\mathcal{X}$ is a compact metric space and Φ is a Feller transition kernel. A probability measure $\pi \in \mathscr{P}(\mathcal{X})$ is said to be invariant under Φ if $\pi\Phi = \pi$.

Definition 3.4.2 (Ergodicity)**.** Suppose that $\mathcal{X}$ is a compact metric space and Φ is a Feller transition kernel. We say that Φ is ergodic if it satisfies the following properties:

1. Φ has a unique invariant measure π.
2. For every measure $\mu \in \mathscr{P}(\mathcal{X})$, we have $\mu\Phi^n \to \pi$ as $n \to \infty$.

   That is, for every μ, and every U, where U is an open neighborhood of π, there is an $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$ we have: $\mu\Phi^n \in U$.

   In other words, if $(X_t)_{t\in\mathbb{N}}$ is a Markov process with transition kernel Φ, then irrespective of the distribution of X_0, the distribution of X_n converges to π as $n \to \infty$.

Definition 3.4.3 (Uniform Ergodicity)**.** Suppose that $\mathcal{X}$ is a compact metric space and Φ is a Feller transition kernel. We say that Φ is uniformly ergodic if it satisfies the following properties:

1. Φ has a unique invariant measure π.
2. For every measure $\mu \in \mathscr{P}(\mathcal{X})$, we have $\mu\Phi^n \to \pi$ as $n \to \infty$ uniformly.

   That is, for every U, an open neighborhood of π, there is an $n_0 \in \mathbb{N}$, such that for every μ, and all $n \geq n_0$ we have: $\mu\Phi^n \in U$.

   In other words, if $(X_t)_{t\in\mathbb{N}}$ is a Markov process with transition kernel Φ, the distribution of X_n converges to π as $n \to \infty$, uniformly in the distribution of X_0.

The definitions of ergodicity and uniform ergodicity can be rewritten as follows:

Proposition 3.4.4 (Ergodicity)**.** *Let Φ be a Feller kernel on a compact metric space $\mathcal{X}$. Then, Φ is ergodic if and only if it has an invariant measure π such that for all $x \in \mathcal{X}$, we have $\Phi^n(x, \cdot) \to \pi$ as $n \to \infty$.*

Proposition 3.4.5 (Uniform Ergodicity)**.** *Let Φ be a Feller kernel on a compact metric space $\mathcal{X}$. Then, Φ is uniform ergodic if and only if it has an invariant measure π such that for all $x \in \mathcal{X}$, we have $\Phi^n(x, \cdot) \to \pi$ uniformly as $n \to \infty$.*

Proposition 3.4.6 (Equivalence Definitions for Uniform Ergodicity)**.** *Let Φ be a Feller transition kernel on a compact metric space $\mathcal{X}$. The following conditions are equivalent:*

(i) *Φ has a unique invariant measure π, and $\mu\Phi^n \to \pi$ uniformly for $\mu \in \mathscr{P}(\mathcal{X})$.*

(ii) *$\bigcap_{n=0}^{\infty} \mathscr{P}(\mathcal{X})\Phi^n$ is singleton.*

(iii) *For each $f \in C(\mathcal{X})$, the sequence $\Phi^n f$ converges to a constant c_f.*

(iv) *Φ has a unique invariant measure π, and $\Phi^n(x, \cdot) \to \pi$ uniformly for $x \in \mathcal{X}$.*

The proof of this proposition is found in chapter 2 of the book [5].

Chapter 4

ERGODICITY VERSES UNIFORM ERGODICITY IN CELLULAR AUTOMATA

The potential distinction between ergodicity and uniform ergodicity was neglected in the literature. It is clear that uniform ergodicity of a Feller kernel implies its ergodicity, but is the converse true? The following example shows that, in general, an ergodic Feller kernel on a compact metric space is not necessarily uniformly ergodic.

Example 4.0.1 (Square mod 1)**.** Let $F\colon \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be the map given by $F(x) \coloneqq x^2$. Then, the transition kernel corresponding to F is ergodic but not uniformly ergodic. ◯

Although this implication is not valid in general, it might still be valid in the case of PCA.

Question 4.1. *Is every ergodic PCA uniformly ergodic?*

I am going to answer this question in two special cases: the fully deterministic case, and the fully probabilistic case.

4.1 Fully Deterministic Cellular Automata

In this section I will introduce the notions of nilpotency and asymptotic nilpotency. I will prove that, for fully deterministic cellular automata, uniform ergodicity is equivalent to nilpotency, and ergodicity is equivalent to asymptotic nilpotency. It has been proved that, for deterministic cellular automata, nilpotency is equivalent to asymptotic nilpotency and this gives us the proof for ergodicity is equivalent to uniform ergodic for fully deterministic cellular automata.

4.1.1 *Nilpotency.*

Definition 4.1.1 (Nilpotency)**.** A DCA $F : \mathcal{X} \to \mathcal{X}$ is called *nilpotent* if there is a configuration $z \in \mathcal{X}$ and an $n \in \mathbb{Z}^+$ such that $F^n(x) = z$ for every $x \in \mathcal{X}$.

Definition 4.1.2 (Asymptotic Nilpotency)**.** A DCA $F : \mathcal{X} \to \mathcal{X}$ is called *asymptotically nilpotent* if there is a configuration $z \in \mathcal{X}$ such that for every $x \in \mathcal{X}$, we have $F^n(x) \to z$ as $n \to \infty$.

Observation 4.1.3. *Every nilpotent DCA is clearly also asymptotically nilpotent.*

Proposition 4.1.4. *If F is asymptotically nilpotent, then $F(z) = z$.*

Proof. Let x be a configuration. Since F is asymptotic nilpotent, we have $F^n(x) \to z$. Similarly, $F^n(F(x)) \to z$. Since $F^n(F(x)) = F^{n+1}(x) = F(F^n(x))$, we obtain that $F(F^n(x)) \to z$. On the other hand, since $F^n(x) \to z$ and F is continuous, we have $F(F^n(x)) \to F(z)$. Thus, $F(z) = z$. □

Proposition 4.1.5. *If F is asymptotically nilpotent, then $\sigma^k(z) = z$ for every $k \in \mathbb{Z}^d$. In particular, z is a homogeneous configuration (i.e., all cells have the same state in z).*

Proof. Let x be a configuration. Since F is asymptotic nilpotent we have $F^n(x) \to z$. Similarly, σ by Theorem 3.3.3 we get $\sigma^k(F^n(x)) \to z$. On the other hand, since $F^n(x) \to z$ and σ^k is continuous, we have $\sigma^k(F^n(x)) \to \sigma^k(z)$. Thus, $\sigma^k(z) = z$. □

4.1.2 *Ergodicity is equivalent to Uniform Ergodicity for DCA*

Theorem 4.1.6. *Every ergodic DCA is uniform ergodic.*

The prove of this theorem will not be a direct proof. It will depend on the concepts of nilpotency and asymptotic nilpotency. We show that ergodicity is equivalent to nilpotency and uniform ergodicity is equivalent to asymptotic nilpotency. Then proving that nilpotent CA is equivalent to asymptotic nilpotent CA will prove Theorem 4.1.6.

The following two propositions appear in [6]

Proposition 4.1.7. *Consider a DCA F and let Φ be the transition kernel corresponding to F. Then:*

F is asymptotically nilpotent if and only if Φ is ergodic.

Proof. Assume that F is asymptotically nilpotent and let z be the configuration specified in the definition of asymptotic nilpotency. Then by Proposition 4.1.4 we get $F(z) = z$. Then, it follows that that δ_z is an invariant measure for Φ. Note that for all $x \in \mathcal{X}$, we have $\Phi^n(x, \cdot) = \delta_{F^n(x)}$. Since F is asymptotically nilpotent we have $F^n(x) \to z$, hence $\delta_{F^n(x)} \to \delta_z$. Thus that for all x we have $\Phi^n(x, \cdot) \to \delta_z$. Therefore according to Proposition 3.4.4 we conclude that Φ is ergodic.

Now, assume that Φ is ergodic. Then Φ has a unique invariant measure. Let us start with a homogeneous configuration $\underline{x}$. Observe that for every n, $F^n(\underline{x})$ is homogeneous configuration. Since there are finitely many homogeneous configurations, the sequence $F^n(\underline{x})$ eventually reaches a cycle. Let l be the length of this cycle. I am going to prove that $l = 1$. Assume that $l > 1$. There is an invariant measure $(\delta_{\underline{x}} + \delta_{F(\underline{x})} + \cdots + \delta_{F^{l-1}(\underline{x})})/l$ supported at the cycle of configurations $\underline{x}, F(\underline{x}), \ldots, F^{l-1}(\underline{x})$. Since Φ is ergodic then $\mu = (\delta_{\underline{x}} + \delta_{F(\underline{x})} + \cdots + \delta_{F^{l-1}(\underline{x})})/l$ is a unique invariant measure. However, the sequence $\delta_{\underline{x}}\Phi^n$ does not converge weakly to μ, which is a contradiction. Therefore $l = 1$.

As a result of getting $l = 1$ we get $F^{n+1}(\underline{x}) = F^n(\underline{x})$ for some n. Let $z = F^n(\underline{x})$. By definition, $F(z) = z$. This implies that δ_z is invariant for Φ. Since Φ is ergodic, for all μ, we have $\mu\Phi^n \to \delta_z$ as $n \to \infty$. In particular, if $\mu = \delta_x$ for some $x \in \mathcal{X}$, then $\delta_x\Phi^n \to \delta_z$. Recall that, $\delta_x\Phi^n = \delta_{F^n(x)}$ because Φ is the kernel associated to F. Therefore, $\delta_{F^n(x)} \to \delta_z$. This implies that, $F^n(x) \to z$. Since this is true for every x, we conclude that F is asymptotically nilpotent. □

Proposition 4.1.8. *Consider a DCA F and let Φ be the transition kernel corresponding to F. Then:*

F is nilpotent if and only if Φ is uniformly ergodic.

Proof. The proof of this proposition is similar to the proof of the previous Proposition 4.1.7 □

Theorem 4.1.9 (Nilpotency ⇔ asymptotic nilpotency)**.** *Every asymptotically nilpotent CA is nilpotent.*

The proof of this theorem in the case of one-dimensional CA is done in this paper [7]. Its extension to higher dimensions is carried out in this paper [8].

Combining Theorem 4.1.9 with Propositions 4.1.7 and 4.1.8 proves Theorem 4.1.6.

4.2 Fully Probabilistic Cellular Automata

In the previous section, we already showed a special case in which the answer to Question 4.1 is positive. Here, we are going to present another special case in which the answer is positive.

Note that the theorem will be proven in a setting which is more general than PCA.

First, I am going to give a definition that will be useful later on.

Definition 4.2.1 (Uniformly full Support)**.** A Feller kernel Φ on a compact metric space $\mathcal{X}$ is said to be uniformly full-support if for every non-empty open set $A \subseteq \mathcal{X}$, there exists $\delta > 0$ such that $\Phi(x, A) \geq \delta$ for every $x \in \mathcal{X}$.

Here is also a reminder about Baire's theorem where we used it in the proof of Lemma 4.2.5.

Theorem 4.2.2 (Baire's theorem)**.** *Let X be a complete metric space.*

1. *Let $U_1, U_2, \ldots$ be a countable family of sets that are open and dense. Then, $\bigcap_{k=1}^{\infty} U_k$ is also dense.*
2. *Let $V_1, V_2, \ldots$ be a countable family of sets that are closed and nowhere dense. Then $\bigcup_{k=1}^{\infty} V_k$ is nowhere dense.*

Theorem 4.2.3 (Ergodicity + full support $\implies$ uniform ergodicity)**.** *Suppose that $\mathcal{X}$ is a compact metric space and Φ is a full support Feller kernel on $\mathcal{X}$. Then Φ is ergodic if and only if Φ is uniformly ergodic.*

Proposition 4.2.4 (Full Support)**.** *A PCA has full support if and only if its local transition kernel is a strictly positive matrix.*

The proof of the above Theorem 4.2.3 is based on the following two lemmas:

Lemma 4.2.5. *Assume that Φ is a Feller kernel on a compact metric space $\mathcal{X}$ and that Φ is ergodic with unique invariant measure π. Let U be an open neighborhood of π. Then, there is an $n \in \mathbb{N}$ and there is a non empty open set $A \subseteq \mathcal{X}$ such that $\Phi^t(x, \cdot) \in U$ for all $x \in A$ and $t \geq n$.*

Proof. Choose an open set U such that $\pi \in U'$ and $\overline{U'} \subseteq U$. For each $n \in \mathbb{N}$, define

$$E_m = \{x \in \mathcal{X} : \Phi^t(x, \cdot) \in \overline{U'} \text{ for every } t \geq m\}$$

Note that E_m is closed since Φ has the Feller property. Since Φ is ergodic, $\bigcup_{m=0}^{\infty} E_m = \mathcal{X}$. Since $\mathcal{X}$ is not nowhere dense, by Baire's theorem, there exists n such that E_n is not nowhere dense. That is, there is a non-empty open set $A \subseteq \overline{E_n} = E_n$. For every $x \in A \subseteq E_n$ and $t \geq n$, we have $\Phi^t(x, \cdot) \in \overline{U'} \subseteq U$. □

Lemma 4.2.6 (full support $\implies$ uniform full support)**.** *Every full support Feller kernel is uniformly full support.*

Proof. Suppose Φ is Feller kernel on a compact metric space $\mathcal{X}$ but not uniformly full-support. Then there is a non empty set A subset of $\mathcal{X}$ such that $\inf\{\Phi(x, A) : x \in \mathcal{X}\} = 0$ and hence there is a sequence $x_1, x_2, ... \in \mathcal{X}$ such that $\Phi(x_n, A) \to 0$ as $n \to \infty$. By compactness, there are $n_1 < n_2 < \cdots$ such that the sub-sequence $x_{n_1}, x_{n_2}, \ldots$ converges to a point $x \in \mathcal{X}$. Now by the Feller property, $\Phi(x_{n_k}, \cdot) \to \Phi(x, \cdot)$ weakly as $k \to \infty$. By the portmanteau theorem [4, Chapter 6], $\Phi(x, A) \leq \liminf_{k\to\infty} \Phi(x_{n_k}, A) = 0$, thus $\Phi(x, A) = 0$. This implies that ϕ is not full-support. □

Proof of Theorem 4.2.3. Assume that Φ is a full support Feller ergodic kernel on $\mathcal{X}$. We want to show that for all $\pi \in \mathscr{P}(\mathcal{X})$, we have $\pi\Phi^t \to \mu$ weakly and uniformly, which is equivalent to showing that for every $f \in C(\mathcal{X})$ we have $\pi\Phi^t f \to \mu(f)$ uniformly.

Let $f \in C(\mathcal{X})$, $\epsilon > 0$ and $\delta > 0$ (to be determined later). Set $U = \{\nu \in \mathscr{P}(\mathcal{X}) : |\nu(f) - \mu(f)| < \delta\}$. Let $\varnothing \neq A \subseteq \mathcal{X}$ and n be as in Lemma (4.2.5). And by Lemma 4.2.6, there is $\gamma > 0$ such that $\Phi(x, A) \geq \gamma$ for every $x \in \mathcal{X}$.

Let $\pi \in \mathscr{P}(\mathcal{X})$ be any measure. Let $Y_0, Y_1, ...$ be a Markov process with kernel Φ and $Y_0 \sim \pi$, thus we get that $Y_n \sim \pi\Phi^n$. Define the first hitting time for A as $T_A = \inf\{n \geq 0 : X_n \in A\}$. For $t \geq n$, we have:

$$\begin{aligned}
\pi\Phi^t f &= \mathbb{E}[f(Y_t)] \\
&= \mathbb{E}[\mathbb{E}[f(Y_t)|T_A]] \\
&= \sum_{m=0}^{t-n} \mathbb{P}(T_A = m)\mathbb{E}[f(Y_t)|T_A = m] + \sum_{m>t-n} \mathbb{P}(T_A = m)\mathbb{E}[f(Y_t)|T_A = m]
\end{aligned}$$

Now

$$\begin{aligned}
\mathbb{E}[f(Y_t)|T_A = m] &= \mathbb{E}[\mathbb{E}[f(Y_t]|T_A = m, Y_m] \\
&= \mathbb{E}[\mathbb{E}[f(Y_t]|Y_m] \quad \text{by the Markov property} \\
&= \mu(f) \pm \delta
\end{aligned}$$

Now

$$\begin{aligned}
&\pi\Phi^t f \\
&\quad = \mathbb{E}[\mathbb{1}_A(Y_m).\mathbb{E}[f(Y_t)|Y_m]] + \sum_{m>t-n} \mathbb{P}(T_A = m)\mathbb{E}[f(Y_t)|T_A = m]
\end{aligned}$$

But

$$\mathbb{P}(T_A = m) \leq (1-\gamma)^{m-1}.$$

and $f(Y_t)$ is continuous and bounded. Hence for large t we have

$$\begin{aligned} |\pi\Phi^t f - \mu(f)| &\leq \delta + \sum_{m>t-n} (1-\gamma)^{m-1}.||f|| \\ &\leq \delta + 2||f||(1-\frac{1}{\gamma})(1-\gamma)^{t-n} \end{aligned}$$

Choose δ and n such that the below inequality holds.

$$\delta + 2||f||(1-\frac{1}{\gamma})(1-\gamma)^{t-n} \leq \epsilon$$

Thus, choose $\delta < \epsilon/2$. This will give us the value of γ. Thus we get $|\pi\Phi^t f - \mu(f)| < \epsilon$ independently of π. And we get our result, that is Φ is uniformly ergodic. □

Appendix
RECURRENT MARKOV CHAINS

.0.1 *Recurrent States*

Definition .0.7 (Recurrent State)**.** Let $(X_t)_{t\in\mathbb{N}}$ be a Markov chain with state space S. A state $x \in S$ is called recurrent, if starting from state x , the chain will almost surely go back to state x within a finite number of time steps.

Mathematically,

$$\mathbb{P}(X_t = x \text{ for some } t \geq 1 | X_0 = x) = 1$$

Definition .0.8 (Positive Recurrence)**.** Let $(X_t)_{t\in\mathbb{N}}$ be a Markov chain with state space S. A recurrent state x is said to be positively recurrent if starting from x, the expected return time to x is finite, that is, $\mu_{xx} < \infty$.

Bibliography

[1] N. Privault, *Understanding Markov chains*, ser. Springer Undergraduate Mathematics Series. Springer, Singapore, 2018, pp. xvii+372, Examples and applications, Second edition of [MR3100135], ISBN: 978-981-13-0658-7; 978-981-13-0659-4. DOI: 10.1007/978-981-13-0659-4. [Online]. Available: https://doi-org.ezproxy.aub.edu.lb/10.1007/978-981-13-0659-4.

[2] M. Aldridge. "Introduction to Markov processes." (2020), [Online]. Available: https://mpaldridge.github.io/math2750/S00-about.html (visited on 02/26/2020).

[3] P. Kurka, *Topological and symbolic dynamics*, ser. Cours Spécialisés [Specialized Courses]. Société Mathématique de France, Paris, 2003, vol. 11, pp. xii+315, ISBN: 2-85629-143-0.

[4] P. Walters, *An introduction to ergodic theory*, ser. Graduate Texts in Mathematics. Springer-Verlag, New York-Berlin, 1982, vol. 79, pp. ix+250, ISBN: 0-387-90599-5.

[5] A. L. Toom, N. B. Vasilyev, O. N. Stavskaya, L. G. Mityushin, G. L. Kuryumov, and S. A. Pirogov, "Discrete local Markov systems," in *Stochastic cellular systems: ergodicity, memory, morphogenesis*, R. L. Dobrushin, V. I. Kryukov, and A. L. Toom, Eds., Manchester University Press, 1990, ISBN: 9780719022067.

[6] J. Mairesse and I. Marcovici, "Around probabilistic cellular automata," *Theoret. Comput. Sci.*, vol. 559, pp. 42–72, 2014, ISSN: 0304-3975. DOI: 10.1016/j.tcs.2014.09.009. [Online]. Available: https://doi-org.ezproxy.aub.edu.lb/10.1016/j.tcs.2014.09.009.

[7] P. Guillon and G. Richard, "Nilpotency and limit sets of cellular automata," in *Mathematical foundations of computer science 2008*, ser. Lecture Notes in Comput. Sci. Vol. 5162, Springer, Berlin, 2008, pp. 375–386. DOI: 10.1007/978-3-540-85238-4_30. [Online]. Available: https://doi-org.ezproxy.aub.edu.lb/10.1007/978-3-540-85238-4_30.

[8] V. Salo, "On nilpotency and asymptotic nilpotency of cellular automata," *Electronic Proceedings in Theoretical Computer Science*, vol. 90, pp. 86–96, 2012. DOI: 10.4204/eptcs.90.7.