Spring 4-28-2021

# MARKOV CHAINS AND THEIR APPLICATIONS

Fariha Mahfuz
*University of Texas at Tyler*

# MARKOV CHAINS AND THEIR APPLICATIONS

by

FARIHA MAHFUZ

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
Department of Mathematics

Kassie Archer, Ph.D., Committee Chair

College of Arts and Sciences

The University of Texas at Tyler
April 2021

The University of Texas at Tyler
Tyler, Texas

This is to certify that the Master's Thesis of

FARIHA MAHFUZ

has been approved for the thesis requirement on

March 23, 2021

for the Master of Science degree.

Approvals:

Thesis Chair: Kassie Archer, Ph.D.

Member: Christina Graves, Ph.D.

Member: Scott LaLonde, Ph.D.

Chair, Department of Mathematics

Dean, College of Arts and Sciences

Acknowledgments

Throughout the writing of this thesis I have received a great deal of support and assistance. I would like to express sincere gratitude to my supervisor, Dr. Kassie Archer, for your encouragement and guidance. Without your invaluable advice and continuous support this thesis would not have been possible. I would like to thank my committee members Dr. Christy Graves and Dr. Scott LaLonde for their critique and correction which shaped the final version of this thesis significantly. I would also like to thank Dr. Nathan Smith for helping me throughout my masters degree. He showed me kindness and compassion when I was going through troubling time. I would also like to thank my mother who worked all her life for a better future for me. I thank my husband for his support.

# Table of Contents

# List of Figures

Abstract

MARKOV CHAINS AND THEIR APPLICATIONS

Fariha Mahfuz

Thesis chair: Kassie Archer, Ph.D.

The University of Texas at Tyler
April 2021

Markov chain is a stochastic model that is used to predict future events. Markov chain is relatively simple since it only requires the information of the present state to predict the future states. In this paper we will go over the basic concepts of Markov Chain and several of its applications including Google PageRank algorithm, weather prediction and gamblers ruin.

We examine on how the Google PageRank algorithm works efficiently to provide PageRank for a Google search result. We also show how can we use Markov chain to predict weather by creating a model from real life data.

# Chapter 1

# Introduction

The concept of a Markov chain was developed by a Russian Mathematician Andrei A. Markov (1856-1922). In 1907, A. A. Markov began the study of an important new type of chance process. A Markov chain is a stochastic model that describes a sequence of possible events or transitions from one state to another of a system. The probability of transitions from state to state only depends on the current state of the system. A Markov Chain is used to unravel predictions about future states of a stochastic process using only knowledge of the present state. This property of "forgetting" past states is known as the memoryless property.

To model a stochastic process with a Markov chain, we need a set of states, i.e. possible outcomes and probabilities of moving from one state to another. The state space, or set of all possible events, can be letters, numbers, weather conditions, baseball scores, or stock performances. Moving from one state to another is called a transition. Transition probabilities are the probabilities of transitioning from one state to another in a single step. We use a transition matrix to list the transition probabilities. The transition matrix will be an $n \times n$ matrix when the chain has $n$ possible states. The $(i, j)$ entry is the probability of transitioning from state $i$ to state $j$. We will also use a directed graph along with the matrix to define the transitions of a Markov Chain. A directed graph consists of vertices and edges connecting to each vertex. Each edge $(m, n)$ of the graph has a weight which denotes the probability to reach $n$ from $m$ in one step. A random walk on a directed graph consists of sequence of vertices which starts at a vertex and traversing an edge to a new vertex.

In Chapter 2, we will discuss ergodic and regular Markov Chain and define associated theorems. We will also discuss at long term behavior or Markov chain.

In Chapter 3, we will explore a Markov Chain in which it is impossible to exit a specific state once it is reached. This type of Markov Chain is known as absorbing Markov Chain.

In Chapter 4, we will discuss how Markov chain derived from random walk on a graph was applied to Google PageRank algorithm.

Finally, in Chapter 5, we provide a few other applications of Markov chains, including Gambler's Ruin and predicting weather demonstrating methods from Chapter 2.

# Chapter 2

# Markov Chain and Stationary distribution

In this Chapter, we provide necessary definitions and examples that we use throughout this paper. Definitions and theorems are borrowed from Chapter 11 of the book *Introduction to Probability* by Charles Miller Grinstead and James Laurie Snell [3].

## 2.1 Definitions and examples

Let's start with the definition of Markov chain.

**Definition 2.1.** A **Markov chain** is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.[1]

Suppose, we have a set of states, $S = (j, i_{n-1}, i_{n-2}, ..., i_0)$. Markov chains follow Markov property which can be written as:

For any $n \geq 1$,

$$P(X_n = j | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) = P(X_n = j | X_{n-1} = i_{n-1}).$$

In other words, a Markov chain model is a model in which the likelihood of an event depends on what happened last.

The **transition matrix** for a Markov chain is a stochastic matrix whose $(i, j)$ entry gives the probability that an element moves from the state $s_i$ to the state $s_j$ during the next step of the process. The probability is denoted by $p_{ij}$, and this probability does not depend upon which states the chain was in before the current state. The probabilities $p_{ij}$ are called **transition probabilities**.

**Definition 2.2.** A **probability vector** with non-negative entries is a row vector whose entries are nonnegative and sum to 1.

Let us demonstrate these definitions with an example.

**Example 2.3.** Suppose Toyota controls 40% of the car market. They decide to hire a research team to see the effects after an ad campaign. Here each time step is a week. The team finds that:

- Someone using Toyota will stay with Toyota with 85% probability,
- Someone not using Toyota will switch to Toyota with 60% probability.

Let $T$ be the state "use Toyota" and $T'$ be the state "Does not use Toyota". The Initial Distribution matrix, $A_0$ indicates the initial probabilities for each state. In this example $A_0$ is given by

$$A_0 = \begin{array}{cc} T & T' \\ \end{array}$$
$$A_0 = \begin{bmatrix} .40 & .60 \end{bmatrix}$$

Recall that the transition matrix, $P$, is the matrix of transition probabilities. In this example, since the probability of $T \to T$ is .85, we must have that $T \to T'$ is $1 - .85 = .15$. Similarly $T' \to T$ is .60 . Thus $T' \to T'$ is $1 - .60 = .40$. Therefore, we get

$$P = \begin{array}{c} \begin{array}{cc} T & T' \end{array} \\ \begin{bmatrix} .85 & .15 \\ .60 & .40 \end{bmatrix} \end{array} \begin{array}{c} \\ T \\ T' \end{array} .$$

There are several reasonable questions to ask.

- What is the market share after 1 week?
- After 2 weeks? After $n$ weeks?
- Long term, how does the market share change?

Theorem 2.4 mentioned below will allow us to compute these.

**Theorem 2.4.** Let P be the transition matrix of a Markov chain. The (i, j) entry $p_{ij}^n$ of the matrix $P^n$ gives the probability that the Markov chain, starting in state $s_i$, will be in state $s_j$ after n steps.

**Example 2.5.** Continuing Example 2.4, after 2 weeks we get $P^2$.

$$P^2 = \begin{bmatrix} 0.8125 & 0.1875 \\ 0.75 & 0.25 \end{bmatrix}$$

The probability that someone using Toyota will stay with Toyota in 2 weeks is 81.25%. After 10 weeks,

$$P^{10} \approx \begin{bmatrix} 0.80 & 0.20 \\ 0.80 & 0.20 \end{bmatrix}$$

After 100 weeks,

$$P^{100} \approx \begin{bmatrix} 0.80 & 0.20 \\ 0.80 & 0.20 \end{bmatrix}$$

In fact $P^n \approx \begin{bmatrix} 0.80 & 0.20 \\ 0.80 & 0.20 \end{bmatrix}$ for all $n$ large enough.

**Theorem 2.6.** Let P be the transition matrix of a Markov chain, and let $A_o$ be the probability vector which represents the starting distribution. Then the probability that the chain is in state $s_i$ after n steps is the ith entry in the vector

$$A_n = A_{n-1}P$$

**Example 2.7.** To see what the market share is after one week, we compute $A_1$ as

$$A_1 = A_o \times P = \begin{bmatrix} .4 & .6 \end{bmatrix} \begin{bmatrix} .85 & .15 \\ .60 & .40 \end{bmatrix} = \begin{bmatrix} .7 & .3 \end{bmatrix}.$$

This means after one week, we expect to Toyota to have 70% of the market share. Similarly the market share after two weeks, $A_2$ can be calculated as follows:

$$A_2 = A_o \times P^2 = \begin{bmatrix} .7 & .3 \end{bmatrix} \begin{bmatrix} .85 & .15 \\ .60 & .40 \end{bmatrix} = \begin{bmatrix} .775 & .225 \end{bmatrix}.$$

After two weeks, we expect to Toyota to have 77.5% of the market share. Now lets find the the market share after 10 weeks:

$$A_{10} = A_o \times P^{10} = \begin{bmatrix} .4 & .6 \end{bmatrix} \begin{bmatrix} .85 & .15 \\ .60 & .40 \end{bmatrix}^{10} = \begin{bmatrix} .799999 & .200001 \end{bmatrix}.$$

After ten weeks, we expect Toyota to have almost 80% of the market share. Continuing this, we would see that market share starts to stabilize as time passes.

We will see this is indeed the case in Section 2.3, where we demonstrate the existence of the Stationary distribution.

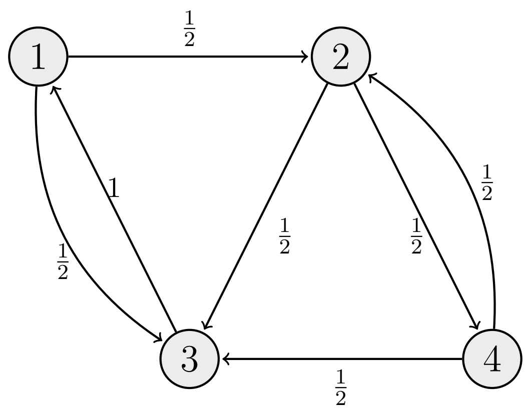
## 2.2 Random Walk

Before we talk about long term behavior let us see another important example.

**Definition 2.8.** A **finite Markov chain** can be described as a random walk on a directed graph. The directed graph has its edges labeled with transition probabilities, in a way so that the law of total probability holds (i.e., for each vertex, the sums of its outgoing edge labels is exactly 1). Each vertex in this graph is a state of the Markov chain.

Random walks in the Markov chain start from moving from vertex to neighboring vertex. If it is in state $x$, the next state $y$ is selected randomly with probability $p_{xy}$. A random walk moves right or left by at most one step on each move.

**Example 2.9.** The following is an example of a Markov Chain on directed graph with 4 vertices.



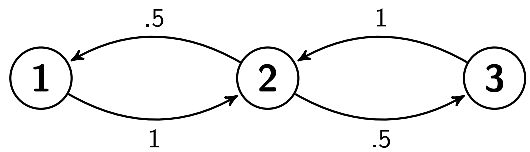

The corresponding transition matrix is, $P = \begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \end{array}.$

## 2.3 Ergodic and Regular Markov Chain

**Definition 2.10.** A Markov chain is called an **ergodic chain** if it is possible to go from every state to every state (not necessarily in one move).

For example, see the Markov chain below.

**Example 2.11.**




The figure above is the Markov chain with the transition matrix ,

$$P = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix} \end{array}$$

We can see that it is possible to move from any state to any state, so the chain is ergodic.

**Definition 2.12.** A Markov chain is called a **regular chain** if some power of the transition matrix has only positive elements (i.e. strictly greater than zero).

**Example 2.13.** From our last example we know that the chain described by the transition matrix $P$ is ergodic. However if $n$ is even, then it is not possible to move from state 1 to state 2 in $n$ steps and if $n$ is odd it is not possible to move from state 1 to state 3 in $n$ steps. So the chain is not regular.

**Example 2.14.**

$$T = \begin{bmatrix} 0 & 1 \\ .3 & .7 \end{bmatrix}$$

In this example the transition matrix $T$ does not have all positive entries. But it is a regular Markov chain because

$$T^2 = \begin{bmatrix} .3 & .7 \\ .21 & .79 \end{bmatrix}$$

has only positive entries.

**Example 2.15.** Looking back at the Example 2.4 the transition matrix

$$\begin{array}{c c} & \begin{array}{cc} T & T' \end{array} \\ P = & \begin{bmatrix} .85 & .15 \\ .60 & .40 \end{bmatrix} \begin{array}{c} T \\ T' \end{array} \end{array}$$

is clearly regular. Some powers of $P$ are

$$P^2 = \begin{bmatrix} 0.8125 & 0.1875 \\ 0.75 & 0.25 \end{bmatrix}$$

$$P^4 = \begin{bmatrix} 0.8007 & 0.199 \\ 0.796 & 0.203 \end{bmatrix}$$

$$P^{10} \approx \begin{bmatrix} 0.80 & 0.20 \\ 0.80 & 0.20 \end{bmatrix}$$

from which it appears that $P^n$ approaches a positive matrix with identical rows. This is true for all regular Markov chains and we can see it from the theorem below.

**Theorem 2.16.** Let $P$ be the transition matrix for a regular chain. Then, as $n \to \infty$ , the powers $P^n$ approach a limiting matrix $W$ with all rows the same vector w. The vector $w$ is a strictly positive probability vector (i.e., the components are all positive and they sum to one).

**Theorem 2.17.** Let P be a regular transition matrix, let

$$W = \lim_{n \to \infty} P^n,$$

let w be the common row of W, and let c be the column vector all of whose components are 1. Then

- $wP = w$, and any row vector v such that $vP = v$ is a constant multiple of w.
- $Pc = c$,and any column vector x such that $Px = x$ is a multiple of c.

Here is another important theorem related to the topic.

**Theorem 2.18.** Let P be the transition matrix for a regular chain and v an arbitrary probability vector. Then

$$w = \lim_{n \to \infty} vP^n,$$

where w is the unique fixed probability vector for P.

We also obtain a new interpretation for w. After sufficiently long time or for large enough n probability of being in the various states is given by $wP^n = w$, remains unchanged. This process is called "stationary."

**Definition 2.19.** A **stationary distribution** $w$ is a (row) vector, whose entries are non-negative and sum to 1, is unchanged by the operation of transition matrix P on it and so it satisfies,

$$wP = w.$$

Here $w$ is a normalized ($\sum_i w_i = 1$) left eigenvector of the transition matrix $P$ with an eigenvalue of 1. [1]

We can find the stationary distribution of the following transition matrix using this theorem.

**Example 2.20.**

$$P = \begin{bmatrix} .85 & .15 \\ .60 & .40 \end{bmatrix}$$

Since we know that $wP = w$, we can write this as a series of equations:

$$.85w_1 + .6w_2 = w_1$$

$$.15w_1 + .4w_2 = w_2$$

$$w_1 + w_2 = 1$$

Solving this system of equation we can see that, $w_1 = .8$ and $w_2 = .2$. Thus $w$=$\begin{bmatrix} .8 & .2 \end{bmatrix}$.
From Example 2.8 we can see that $A_{10}$ is really close to $w$ and as $n \to \infty$, $A_n \to w$.

Long term, 80% of people using Toyota will stay with Toyota with 85% probability and 20% of people not using Toyota will switch to Toyota.

# Chapter 3

# Absorbing Markov Chain

In this Chapter, we will discuss a special type of Markov chain called an absorbing Markov chain. We start off with the definition.

## 3.1 Definitions and Examples

**Definition 3.1.** A state $s_i$ of a Markov chain is called **absorbing** if it is impossible to leave it (i.e., $p_{ii} = 1$). A Markov chain is **absorbing** if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state (not necessarily in one step).

**Example 3.2.**

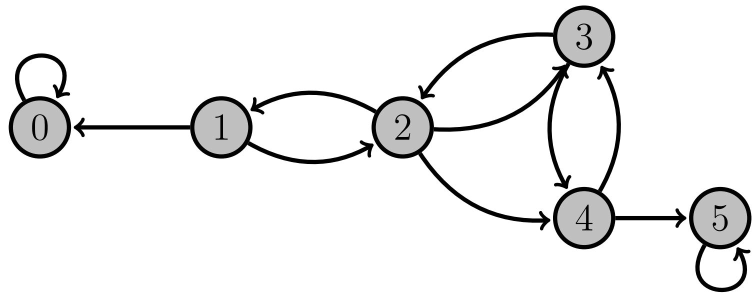
$$P = \begin{array}{c} \begin{array}{ccc} A_1 & A_2 & A_3 \end{array} \\ \begin{bmatrix} .2 & .3 & .5 \\ 0 & 1 & 0 \\ .4 & .2 & .4 \end{bmatrix} \end{array} \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} .$$

The state $A_2$ is an absorbing state since the probability of moving from state $A_2$ to $A_2$ is 1.

**Definition 3.3.** In an absorbing Markov chain, a state which is not absorbing is called **transient**.

In Example 3.2, $A_1$ and $A_3$ are transient states.

**Example 3.4.** Here is an example of a variation of Drunkard's walk which explains absorbing Markov chains. A man walks along a path with a loop through a park. If he is at corner 1, 3 or 4 then he walks to the left or right with equal probability. If he is in corner 2 he can either walk back to corner 1 or 3 or 4 . He continues until he reaches corner 5 which is a bar, or corner 0, which is his home. If he reaches either home or the bar, he stays there. We form a Markov chain with states 0, 1, 2, 3, 4 and 5. States 0 and 5 are absorbing states.




The transition matrix is then, P= $\begin{array}{c c} & \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$

The states 1, 2, 3 and 4 are transient states, and from any of these it is possible to reach the absorbing states 0 and 5. Hence the chain is an absorbing chain. When a process reaches an absorbing state, we shall say that it is absorbed.

**Definition 3.5.** If the chain has $t$ transient states and $r$ absorbing states, then the transition matrix $P$ for an absorbing Markov chain in the canonical form can be written as

$$P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix}$$

where $Q$ is a $t \times t$ matrix, $R$ is $t \times s$, 0 is the $s \times t$ zero matrix and $I$ is the $s \times s$ identity matrix.

**Theorem 3.6.** In an absorbing Markov chain, the probability that the process will be absorbed is 1 (i.e., $Q^n \to 0$ as $n \to \infty$).

**Theorem 3.7.** For an absorbing Markov chain the matrix $I - Q$ has an inverse $N$ and $N = I + Q + Q^2 + \dots$ . The $(i, j)$-entry $n_{ij}$ of the matrix $N$ is the expected number of times the chain is in state $s_j$, given that it starts in state $s_i$. The initial state is counted if $i = j$.

*(Proof)* First we want to calculate the expected number of times the chain spends in the the state $s_j$ starting $s_i$.

Let $X_{i,j}^{(k)}$ be a random variable which equals 1 if the chain, with initial state $s_i$, is in state $s_j$ on $k$-th transition, and equals 0 otherwise. Then,

$$P(X_{i,j}^{(k)} = 1) = Q_{i,j}^{(k)}$$

and

$$P(X_{i,j}^{(k)} = 0) = 1 - Q_{i,j}^{(k)}$$

where $Q_{i,j}^{(k)}$ is the $ij$-th entry of $Q^{(k)}$, with the notation $Q^0 = I$.

Let

$$X_{i,j} = X_{i,j}^{(0)} + X_{i,j}^{(1)} + X_{i,j}^{(2)} + X_{i,j}^{(3)} + X_{i,j}^{(4)} + \dots .$$

Here $X_{i,j}$ gives the total number of times the chain is in state $j$ before absorption.

Then the expected number of times the chain is in state $s_j$ in the first n steps given that we start in state $s_i$,

$$\begin{aligned} n_{ij} = E[X_{i,j}] &= E[X_{i,j}^{(0)} + X_{i,j}^{(1)} + X_{i,j}^{(2)} + X_{i,j}^{(3)} + X_{i,j}^{(4)} + \dots X_{i,j}^{(n)}] \\ &= E[X_{i,j}^{(0)}] + E[X_{i,j}^{(1)}] + E[X_{i,j}^{(3)} + \dots + E[X_{i,j}^{(n)}] \\ &= P(X_{i,j}^{(0)} = 1) + P(X_{i,j}^{(1)} = 1) + \dots + P(X_{i,j}^{(n)} = 1) \\ &= Q_{i,j}^{(0)} + Q_{i,j}^{(1)} + \dots + Q_{i,j}^{(n)}. \end{aligned}$$

In matrix form,

$$N = I + Q + Q^2 + Q^3 + \dots + Q^n.$$

where the identity matrix $I$ has the same dimension as the matrix $Q$. Letting $n$ tend to infinity we have,

$$N = I + Q + Q^2 + Q^3 + \dots$$

since $Q^n \to 0$ as $n \to \infty$, Multiplying by $(I - Q)$ on both sides, we get

$$\begin{aligned} (I - Q) \cdot N &= (I - Q)(I + Q + Q^2 + Q^3 + \dots) \\ &= I. \end{aligned}$$

we see that both $N$ and $I - Q$ are nonsingular thus invertible. Thus multiplying both sides of the equation with $(I - Q)^{-1}$ yeilds,

$$N = (I - Q)^{-1}$$

□

**Definition 3.8.** For an absorbing Markov chain $P$, the matrix $N = (I - Q)^{-1}$ is called the **fundamental matrix** for $P$. The entry $n_{ij}$ of $N$ gives the expected number of times that the process is in the transient state $s_j$ if it is started in the transient state $s_i$.

**Example 3.9.** The transition matrix in example 3.3 in canonical form is

$$P= \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 0 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 0 & 5 \end{array} \\ \left[\begin{array}{cccc|cc} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \end{array}$$

Here $Q = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 \end{bmatrix}$

and $I - Q = \begin{bmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{3} & 1 & -\frac{1}{3} & -\frac{1}{3} \\ 0 & -\frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & 1 \end{bmatrix}$

Finally, $N = (I - Q)^{-1} = \begin{bmatrix} 1.34 & 1 & 0.67 & 0.67 \\ 0.67 & 2 & 1.34 & 1.34 \\ 0.45 & 1.34 & 2.23 & 1.56 \\ 0.23 & .67 & 1.12 & 1.78 \end{bmatrix}$

According to the matrix $N$, the entry 1 in the row 2, column 4 position means if the drunkard starts from state 2 the expected number of steps in state 4 before being absorbed is 1.

## 3.2 Time to Absorption

**Theorem 3.10.** Let $t_i$ be the expected number of steps before the chain is absorbed, given that the chain starts in state $s_i$, and let t be the column vector whose ith entry is $t_i$. Then

$$t = Nc,$$

where c is a column vector all of whose entries are 1.

*(Proof)* Here, $c$ is a column vector where all the entries are 1. Thus multiplying $c$ with the

$i^{th}$ row of matrix $N$ will give us the sum of the entries of that row, which is the total expected number of times in the other states starting from the state $s_i$ before being absorbed.

But, $t_i$ be the expected number of steps before the chain is absorbed. So multiplying the $i^{th}$ row of matrix $N$ with $c$ gives us $t_i$, which completes the proof. □

**Example 3.11.** We can find the time for absorption using the theorem above.

$$\begin{aligned} t &= Nc \\ &= \begin{bmatrix} 1.34 & 1 & 0.67 & 0.67 \\ 0.67 & 2 & 1.34 & 1.34 \\ 0.45 & 1.34 & 2.23 & 1.56 \\ 0.23 & .67 & 1.12 & 1.78 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 3.67 \\ 5.34 \\ 5.56 \\ 3.78 \end{bmatrix} \end{aligned} \tag{3.1}$$

Thus, starting in states 1, 2, 3 and 4 the expected times or number of steps to absorption are 3.67,5.34, 5.56 3.78 respectively.

## 3.3 Absorption Probability

**Theorem 3.12.** Let $b_{ij}$ be the probability that an absorbing chain will be absorbed in the absorbing state $s_j$ if it starts in the transient state $s_i$. Let B be the matrix with entries $b_{ij}$. Then B is an t-by-r matrix, and

$$B = NR,$$

where N is the fundamental matrix and R is as in the canonical form.

**Example 3.13.** From the canonical form, we have,

$$R = \begin{bmatrix} .5 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & .5 \end{bmatrix}$$

Thus,

$$
\begin{aligned}
B &= NR \\
&= \begin{bmatrix} 1.34 & 1 & 0.67 & 0.67 \\ 0.67 & 2 & 1.34 & 1.34 \\ 0.45 & 1.34 & 2.23 & 1.56 \\ 0.23 & .67 & 1.12 & 1.78 \end{bmatrix} \begin{bmatrix} .5 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & .5 \end{bmatrix} \\
&= \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 5 \end{array} \\ \begin{bmatrix} .66 & .34 \\ .34 & .67 \\ .23 & .78 \\ .12 & .89 \end{bmatrix} \end{array}
\end{aligned}
$$

The first row interprets the drunkard has probability .66 of making it home and probability .34 of ending up in the bar from intersection 1.

# Chapter 4

## Google PageRank

Google had been using the algorithm called PageRank to rank the webpages of a search result until September 24, 2019. The algorithm was developed by Page and Brin and "PageRank" was named after Larry Page. PageRank is one of the methods Google uses to determine a page's relevance or importance. To apply the PageRank algorithm, we consider the web as directed graph, where web pages are nodes and hyperlinks are edges. PageRank ranks pages based on the number of backlinks pointing to that webpage. A page that is linked to many pages receives a high rank itself

To find out the page score one must consider that the surfer can select any page. However it is not always the case that they select the pages sequentially. Most of the time, a surfer will follow links from a page sequentially, i.e. from a page $i$ the surfer will follow the outgoing links and move on to one of the neighbors of $i$. But this may not happen always. A smaller but positive percentage of the time, the surfer will dump the current page and choose arbitrarily a different page from the web and "teleport" there. To account for such a situation Page and Brin introduced a factor called as the damping factor $d$, that reflects the probability that the surfer drops the current page and "teleports" to a new one. Since he/she can teleport to any web page, each page has $1/n$ probability to be chosen.[2] We develop the corrected algorithm later in this chapter.

First, the simplified version of the PageRank algorithm is defined below,

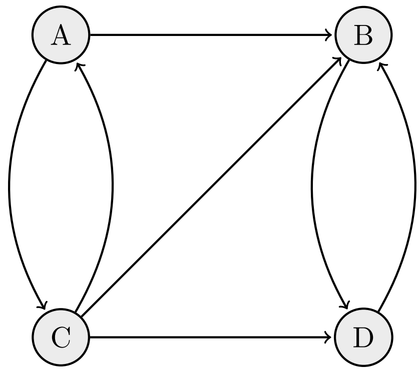
**Definition 4.1.** Suppose that page $P_j$ has $L(j)$ links. If one of those links is to page $P_i$, then $P_j$ will pass on $\frac{1}{L(j)}$ of its importance to $P_i$. The importance ranking of $P_i$ is then the sum of all the contributions made by pages linking to it. That is, if we denote the set of pages linking to $P_i$ by $B_u$, then then the PageRank of $P_i$, $PR(i)$, is given by the formula,

$$PR(i) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$$

.

To demonstrate this algorithm let's look at the example below.

**Example 4.2.** We first consider an example of the internet with only 4 webpages $A, B, C$ and $D$. In this example webpage $A$ has links to pages $C$ and $B$, page $C$ had a link to page $A, B$,and $D$, page $B$ had link to page $D$, and page $D$ had link to page $B$.
The figure below is the graph of these 4 webpages.




The matrix representation of the above graph is

$$P = \begin{array}{c c} & \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

We are going to assign each node or webpage an initial PageRank $\frac{1}{4}$. Then using the simplified PageRank algorithm, $PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$, we calculate the PageRank of each node.

For example, if we want to calculate the PageRank for page $B$,

$$PR(B) = \frac{PR(A)}{2} + \frac{PR(C)}{3} + \frac{PR(D)}{1}$$

$$= \frac{1/4}{2} + \frac{1/4}{3} + \frac{1/4}{1} = 0.45$$

Then the new pageranks are,

$$PR(A) = .083$$

$$PR(B) = .45$$

$$PR(C) = .125$$

$$PR(D) = .333$$

The associated matrix representation is,

$$\begin{bmatrix} 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/3 & 1 \\ 1/2 & 0 & 0 & 0 \\ 0 & 1 & 1/3 & 0 \end{bmatrix} \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} .083 \\ .45 \\ .125 \\ .333 \end{bmatrix}$$

We can see that after one iteretaion PageRank for page $A, B, C$ and $D$ is .083,.45,.125 and .333. Page $B$ has the highest PageRank,which means it will appear earlier in a Google search.

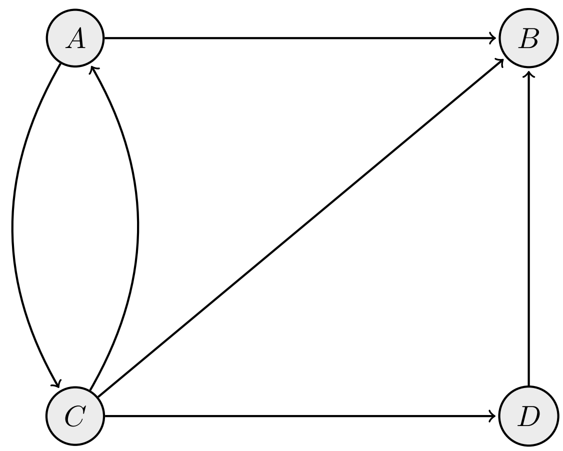
**Definition 4.3.** A node is called a **dangling node** if it does not contain any out-going link.

**Example 4.4.** For instance, node $B$ in this example is a dangling node.

Shown below is the matrix representation of the graph.

$$T = \begin{bmatrix} 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/3 & 1 \\ 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \end{bmatrix}$$

We can see the second column of the matrix is zero which causes the PageRank og page B to converge to zero. To fix this problem we replace the zero column with $\frac{1}{4}$, then our new matrix representation,

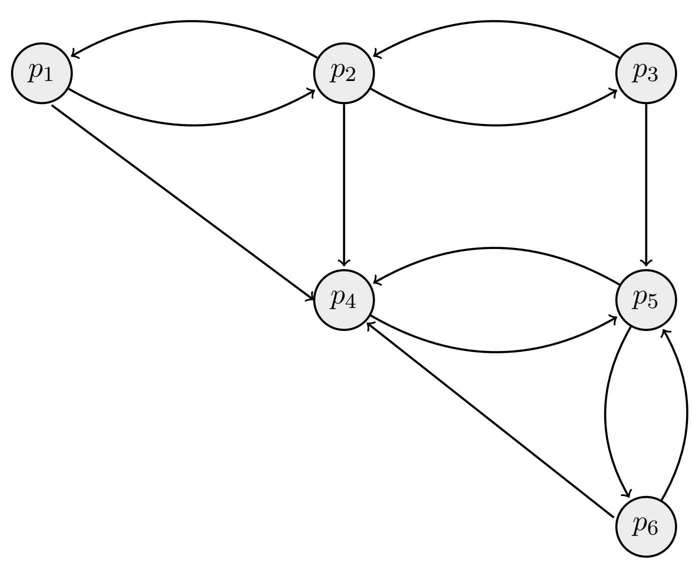



$$T = \begin{bmatrix} 0 & 1/4 & 1/3 & 0 \\ 1/2 & 1/4 & 1/3 & 1 \\ 1/2 & 1/4 & 0 & 0 \\ 0 & 1/4 & 1/3 & 0 \end{bmatrix}.$$

The $\frac{1}{4}$ corresponds to jumping randomly to any website with equal probability.

**Definition 4.5.** The following web graph belongs to the category of reducible graph. In general, a graph is called **irreducible** if for any pair of distinct nodes, we can start from one of them, follow the links in the web graph and arrive at the other node, and vice versa. A graph which is not irreducible is called **reducible**.

**Example 4.6.** In the following example there is no path from $p_4$ to $p_1$, and no path from $p_5$ to $p_3$. The graph is therefore reducible.




We can also find stationary distribution for PageRank using the following condition:

$$I = IT$$

.

Here vector $I$ is an eigenvector of the matrix $T$ with eigenvalue 1
In this case, the matrix is $T$ is,

$$T = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \text{with the stationary distribution } I = \begin{bmatrix} 0 \\ 0 \\ 0 \\ .33 \\ .44 \\ .22 \end{bmatrix}$$

We can see that from the stationary distribution that the PageRanks of the first three webpages are zero even though they all have links connecting from other webpages. This occurs because the graph is reducible.

**Definition 4.7.** In order to calculate PageRanks properly for a reducible web graph, Page and Brin developed the following formula, for webpage $A$

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)},$$

where $p_1, p_2, ..., p_N$ are the pages under consideration, $M(p_i)$ is the set of pages that link to $p_i$, $L(p_j)$ is the number of outbound links on page $p_j$, and $N$ is the total number of pages. Here $d$ is a number between 0 and 1. The constant $d$ is usually called the damping factor or the damping constant. The default value of $d$ is 0.85 in Google. [4]

**Example 4.8.** For example 4.6, calculating the PageRank using the corrected formula for $p_2$ we get,

$$\begin{aligned} PR(p_2) &= \frac{1-.85}{6} + .85\left\{\frac{PR(p_1)}{2} + \frac{PR(p_3)}{2}\right\} \\ &= \frac{0.15}{6} + .85\left\{\frac{1}{6 \cdot 2} + \frac{1}{6 \cdot 2}\right\} \\ &= 0.16. \end{aligned}$$

Calculating the new PageRanks we get:

$$PR(p_1) = .072$$

$$PR(p_2) = .16$$

$$PR(p_3) = .072$$

$$PR(p_4) = .284$$

$$PR(p_5) = .308$$

$$PR(p_6) = .095$$

We can see that page $p_5$ has the highest rank. Page $p_5$ has 3 incoming links. It satisfies with the reasoning of the PageRank algorithm that a page with larger incoming links has higher importance. In this case the page $p_5$ will appear early in the Google search and page then p4 and so on.

# Chapter 5

# Other applications

In this Chapter, we will see two other uses of Markov chain. We won't go too deep in these sections. We will construct a smaller example to see the process.

## 5.1 Gamblers Ruin

Consider this gambling game where the probability of winning \$1 is .58 and probability of losing \$1 is .42.

- The gambler must have at least \$1 up to \$5 to continue playing.
- The game stops which in one of two states: you have \$6, or you have \$0.
- You bet \$1 in each round, and earn \$1 if you win, and lose your dollar if you lose the round.

Using Markov Chain the Gambler's Ruin problem can be modeled as a random walk on a finite Markov chain. We can find the probability of winning \$6 or going broke.

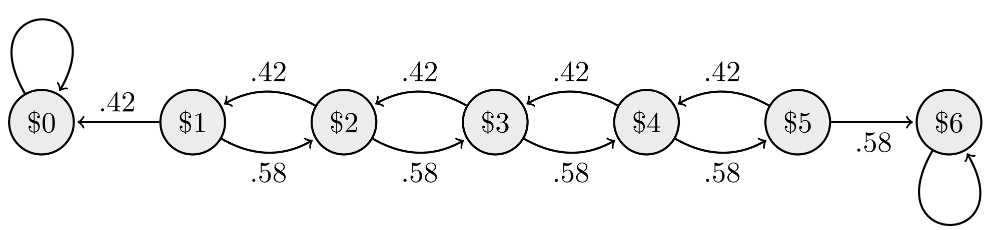



Figure 5.1: The state diagram of the Gambler's Ruin Markov chain

The associated transition matrix for the game:

$$
P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\begin{bmatrix}
1 & .42 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & .42 & 0 & 0 & 0 & 0 \\
0 & .58 & 0 & 0.42 & 0 & 0 & 0 \\
0 & 0 & .58 & 0 & .42 & 0 & 0 \\
0 & 0 & 0 & .58 & 0 & .42 & 0 \\
0 & 0 & 0 & 0 & .58 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.58 & 1
\end{bmatrix}
\end{array}
$$

If the gambler starts with \$1, the initial distribution matrix is given by,

$$A_0 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$$

After 5 rounds,

$$A_5 = A_0 P^5 = \begin{bmatrix} .57 & 0 & 0.17 & 0 & 0.19 & 0 & .065 \end{bmatrix}^T$$

After 50 rounds,

$$A_{50} \approx A_0 P^{50} = \begin{bmatrix} .68 & 0 & 0 & 0 & 0 & 0 & .32 \end{bmatrix}^T$$

Similarly the probability can be calculated if the gambler starts with \$5 , the initial distribution matrix in this case is,

$$A_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}^T$$

After 5 rounds,

$$A_5 = A_0 P^5 = \begin{bmatrix} 0.013 \\ 0 \\ 0.072 \\ 0 \\ 0.125 \\ 0 \\ 0.79 \end{bmatrix}$$

After 55 rounds,

$$A_{55} = A_0 P^{55} \approx \begin{bmatrix} 0.064 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.936 \end{bmatrix}$$

This interprets that after 55 rounds of game, the chance of being ruined is .064% and chance of winning is .936%. Comparing the results we can see that chance of winning is significantly higher with a higher bid of \$5 compared to \$1. Thus one should play this game only if the bid is high.

Also the outcome differs when the game is fair, i.e. the probability of winning and losing is 50%.

## 5.2 Weather prediction

Markov chains have been researched heavily for predicting weather. In the paper, "Weather Forecasting Using Hidden Markov Model", Khaitani, Diksha and Ghose, Udayan, 2017 [5] trained a Markov model with weather data from past 21 years. The authors used Viterbi Algorithm and MATLAB software to calculate predicted data. The study found Markov model performed well predicting weather for the next 5 days based on current day's weather.

Jordan and Talkner in their paper [6] investigated daily weather types of the Alpine region using seasonal Markov chain models. The study compared a 1st and 2nd order Markov chain model and found the two yield similar results. The predictions by Markov models were found to coincide with actual observations for different types of time periods. However, the predictive power of the models were not found to be very high, which was due to high randomness of the data and not due to weakness of the model.

Here in our next example we have observed the weather of Bemidji, Minnesota and calculated the probabilities using Markov model. For simplification assuming the weather can only be in one of 3 possible states, "sunny","snowy" or "cloudy". In the context of Markov chains the probability of the weather being sunny, snowy or cloudy tomorrow, only depends on whether it is sunny, snowy or cloudy today.

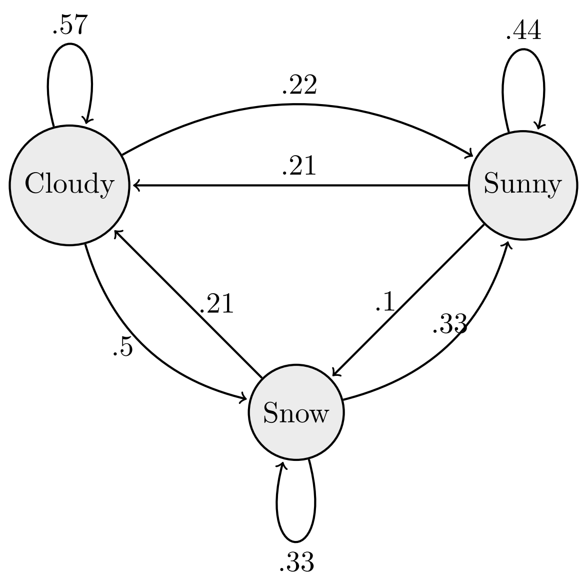



Figure 5.2: The state diagram using a directed graph

The following transition matrix was constructed by taking data from the website https://www.timeandd
at 12 p.m. time for every day in November, 2019, totalling 30 observations.

$$P = \begin{array}{c} \\ \text{cloudy} \\ \text{sunny} \\ \text{snowy} \end{array} \begin{array}{c} \begin{array}{ccc} \text{cloudy} & \text{sunny} & \text{snowy} \end{array} \\ \begin{bmatrix} .57 & .22 & .5 \\ .21 & .44 & .1 \\ .21 & .33 & .33 \end{bmatrix} \end{array}$$

By observation, today's (12 November, 2020) weather at noon is cloudy. This is represented by the vector,

$$x_0 = \begin{array}{c} \text{cloudy} \\ \text{sunny} \\ \text{snowy} \end{array} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

To predict the weather of 15 November, can be predicted the following way,

$$x_3 = P^3.x_0 = \begin{bmatrix} .57 & .22 & .5 \\ .21 & .44 & .1 \\ .21 & .33 & .33 \end{bmatrix}^3 \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.44 \\ 0.30 \\ 0.23 \end{bmatrix} \begin{array}{c} \text{cloudy} \\ \text{sunny} \\ \text{snowy} \end{array}$$

We followed up and checked that it was cloudy on November 15th.

## Bibliography


[1] https://en.wikipedia.org/wiki/PageRank

[2] Prerna Rai, Arvind Lal, *PageRank Model*, International Journal of Computer Applications,Volume 138,2016,(0975 – 8887)

[3] Charles Miller Grinstead, James Laurie Snell, *Introduction to Probability*, Second revised edition,Chapter 11, (1988)

[4] Sergey Brin, Lawrence Page, *"The Anatomy of a Large-Scale Hypertextual Web Search Engine*, Computer Networks and ISDN Systems, no. 1–7, Elsevier BV, Apr. 1998, pp. 107–17. Crossref, doi:10.1016/s0169-7552(98)00110-x

[5] D. Khiatani and U. Ghose, *Weather forecasting using Hidden Markov Model*, 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), Gurgaon, India, 2017, pp. 220-225, doi: 10.1109/IC3TSN.2017.8284480.

[6] Paul Jordan  Peter Talkner (2000),*A seasonal Markov chain model for the weather in the central Alps*, Tellus A: Dynamic Meteorology and Oceanography, 52:4, 455-469, DOI: 10.3402/tellusa.v52i4.12274