STOCASTIC PROCESS AND MARKOV CHAINS

Project Report Submitted to

Mahatma Gandhi University

In partial fulfillment of the requirement for the award of the degree of

MASTER OF SCIENCE IN MATHEMATICS

Submitted By

ELIZABETH GEORGE M

Reg No. 200011014779

Under The Supervision of

Ms. Alakamohan

DEPARTMENT OF MATHEMATICS

Bharata Mata College, Thrikkakara

2020-2022

ACKNOWLEDGEMENT

I sincerely thank each and everyone who has helped me for the successful completion of this project. I thank God Almighty for all his blessings showered upon me. I would like to place on record all my gratitude to Ms. Alakamohan, Department of Mathematics, Bharata Mata

College, Thrikkakara, who has provided her valuable guidance throughout this project work. Also, I thank all other teachers of the department of their encouragement. My friends were also helpful to me through their suggestions for which I am grateful towards them.

ELIZABETH GEORGE M

Place: Thrikkakara

Date:

CERTIFICATE

This is to certify that the dissertation entitled **“STOCASTIC PROCESS AND MARKOV CHAINS”** submitted by ELIZABETH GEORGE M is a record of work done by the candidate during the period of her study under my supervision and guidance.

Ms. Alakamohan

Department of Mathematics

Bharata Mata College ,

Thrikkakara

Place: Thrikkakara

Date:

DECLARATION

I hereby that the project report entitled “**STOCASTIC PROCESS AND MARKOV CHAINS** “ Submitted for the Msc. Degree is my original work done under the supervision of **Ms. ALAKAMOHAN** and the project has not formed the basis for the award of my academic

qualification fellowship or other similar title of other university or board.

Place:

Date **:**

ELIZABETH GEORGE M

CONTENTS

ABSTRACT

A Stocastic Process is a sequence of random variables with time domain T and it represents a collection of experiments. Each random variable assigns a real number to every outcome of a random experiment. The set of all outcomes is known as the 'state space' for an experiment. Stocastic Processes can be classified based on the nature of state space S and parameter set T, where the time and state space can be discrete or continuous in nature. Markov Process is a particular type of stochastic process where the state of each event depents only on the state attained in the previous event. In other words, the future event depents only on the present event and not on its past and this can be expressed in terms of probability. Markov Chains are discrete Markov Processes, where the term 'discrete' represents the countable or finite state space. The 'transition probability' or probability of transition, denoted as p_{ij} is the probability that the Markov Chain moves from state i to state j. A matrix representation of transition probabilities as entries is known as the 'transition probability matrix' or simply transition matrix. The transition probabilities are always non negative and the sum of each row of transition matrix sums up to one. When there are two non successive trials, ie state j at the nth trial and state k at the (n+m)th trial, it is an m-step transition probability denoted as $p_{jk}^{(m)}$. The order of a markov chain is determined by the number of previous states it depents where it can be called as a first order markov chain or a second order markov chain and so on. Using one step transition probabilities we can find the higher transition probabilities which are determined by Chapman Kolomogorov equations.

Markov Chains can be expressed graphically by the transition probabilities and transition matrices. There are applications on graphical markov chains in the field weather forecasting and other scientific fields. In the Gamblers ruin problem one among the two gamblers play until one of them has lost all their wealth and cannot play anymore, and the game can be represented in a state transition diagram with state representing the amount player 1 has and edge represents a transition between states. The markov chains are used in google web page ranking based on the links of a webpage with another. We consider the web as a directed graph where web pages are vertices and hyperlinks as edges. A web page with higher rank appears earlier in a google web search.

Chapter 1

Basic Concepts

Random Experiment:

Random Experiment is defined as an experiment or trial with more than one possible outcome which may be repeated any number of times under same conditions and the outcomes of which vary irregularly from repetition to repetition. It is a non deterministic experiment.

In the case of deterministic experiments, it is possible to predict the outcome of any repetition of the experiment but it is not possible in the case of random experiments. Example: Tossing a coin a number of times and observing whether it shows head or tail which are the only possible outcomes.

Random Variable:

A random variable is a rule that assigns a real number to every outcome of a random experiment. It is a function whose domain is the Sample Space (set of all possible outcomes of an experiment) and range is the set of real numbers.

Random Variables are represented in capital letters X, Y, Z etc whereas small letters x, y, z etc will denote particular values a random variable may assume. A random variable X is a mapping $\mathbf{X}:\mathbf{S}\rightarrow\mathbf{R}$, where **S** is the Sample Space and **R** is the set of real numbers.

Example1.1: Consider a fair coin tossing three times. Let X be the random variable showing the total number of heads appeared. The Sample Space is { HHH, HHT, HTT, TTT, TTH, THH, HTH, THT } if the outcomes are 3 heads, 2 heads, 1 head or no head. Then the random variable X takes the values 0, 1, 2 or 3.

Equally likely outcomes:

As the name suggests these are the events with equal chance of happening. Many events are equally likely outcomes like tossing a coin which has 1/2 probability for getting heads and 1/2 probability getting tails. Also in case of tossing a die there is 1/6 probability of getting any number on the die.

Conditional Probablity and Baye's theorem:

The probability of occurrence of any event A when another event B in relation to A has already occurred is known as conditional probability. It is depicted by P(A|B).

Bayes' theorem defines the probability of occurrence of an event associated with any condition. It is considered for the case of conditional probability.

$$P(A|B) = P(B)P(A).P(A|B)$$

where P(A) is the probability of A occurring and P(B) is the probability of B occurring. P(A|B) is the probability of A given B.

Probability distribution:

A probability distribution is a statistical function that describes all the possible values and probabilities for a random variable within a given range. For instance, if X is used to denote the outcome of a coin toss experiment, then the probability distribution of X would take the value 0.5 (or 1/2) for X = heads, and 0.5 for X = tails.

Chapter 2

Stocastic Process(Random Process)

Introduction

Changes happen randomly in frequently occurring phenomenons of nature like weather, rainfall, humidity, temperature etc. Also the customers arriving at a shop, students arriving to a college etc .shows that randomness can happen in our every life. There are mathematical and statistical studies based on these random changes using theories of probability and other methods.

A Stocastic Process is a sequence of experiments which are random. It is a mathematical model that evolves over time in a probabilistic manner. A Stocastic Process is also known as Random Process. Families of random variables which are functions of time are known as Stocastic Process or Random Process (or random functions).

Definition and examples

A Stocastic Process is defined as a collection of random variables $\{ X(t) : t \in T\}$ that are functions of a real variable, namely time t and $t \in T$ which is the parameter set or index set . t is the parameter running over some index set T called time domain. Here are a few examples of stochastic processes.

1. Consider that there are r cells and an infinitely large number of identical balls. The balls are thrown at random, one by one into the cells, the balls thrown being equally likely to go into any one of the cells. Suppose that X(n) is the number of occupied cells after n number of throws. Then $\{ X(n),\ n \geq 1\}$ constitutes a stochastic process.

2. Consider a random event occurring in time such as the number of telephone calls received at a switchboard. Suppose that X(t) is the random variable which represents the number of incoming calls in an interval (0, t) of duration t units. Then one unit of time is a random variable X(1). So the number of calls within a fixed time interval of specified duration, there is a family $\{ X(t),\ t \in T \}$ constitutes a stochastic process.

State Spaces

The set of possible values or outcomes of a single random variable X(n) of a stochastic process $\{ X(n), n \geq 1 \}$ is known as its State Space. The State Space can be discrete or continuous.

A discrete state space denotes a set of distinct possible outcomes, which can be finite or countably infinite. For example, S = { Head, Tail } is the state space for a single coin flip is said to be a discrete state space.

A continuous state space denotes an uncountably infinite continuum of gradually varying outcomes. For example, the non negative real line $S = R+ = \{ x \in R : x \geq 0 \}$ is the state space for the amount of rainfall in a given day.

Classification of Stocastic Processes

Depending on the continuous or discrete nature of the state space S and parameter set T, a stochastic process can be classified into four types.

1..Discrete Time, Discrete State Space:

If both T and S are discrete, the stochastic process is called a **discrete random sequence.**

For example, if Xn represents the outcome of the nth toss of a fair dice, then $\{ X_n, n \geq 1\}$ is a discrete random sequence, since $T = \{ 1, 2, 3, \ldots\ldots\}$ and $S = \{ 1, 2, 3, 4, 5, 6 \}$.

2.Discrete Time, Continuous State Space:

If T is discrete and S is continuous, the stochastic process is called a **continuous random sequence.**

For example, if Xn represents the temperature at the end of the nth hour of a day, then { Xn , 1 ≤.n ≤ 24 } is a continuous random sequence, since temperature can take any value in an interval . Hence it is continuous.

3.Continuous Time, Discrete State Space:

If T is continuous and S is discrete, the random process is called a **discrete random process.**

For example, if Xt represents the number of telephone calls received in the interval (0 , t) then { Xt } is a discrete random process, since S = { 0, 1, 2, 3…….. }.

4. Continuous Time, Continuous State Space:

If both T and S are continuous, the random process is called a **continuous random process.**

For example, if Xt represents the maximum temperature at a place in the interval (0, t) then { Xt } is a continuous random process.

In the names given above, the word 'discrete' or 'continuous' is used to refer to the nature of the state space S and the word 'sequence' or 'process' is used to refer to the nature of time T.

Types of Stocastic or Random Processes

There are various types of stochastic processes . The **Bernoulli Process** is one of the simplest stochastic process . It is a sequence of independent and identically distributed

random variables , where each random variable has a probability of one or zero , say one with probability ` p' and zero with a probability 1 – p . This process is happened in the case of flipping a coin , where the probability of getting a head is p and the random variable is given the value 1 , and the probability of getting a tail is 1-p which is denoted with the random variable of value 0. ie

$$P(X = 1) = P \quad \text{and}$$

$$P(X=0) = 1 - p$$

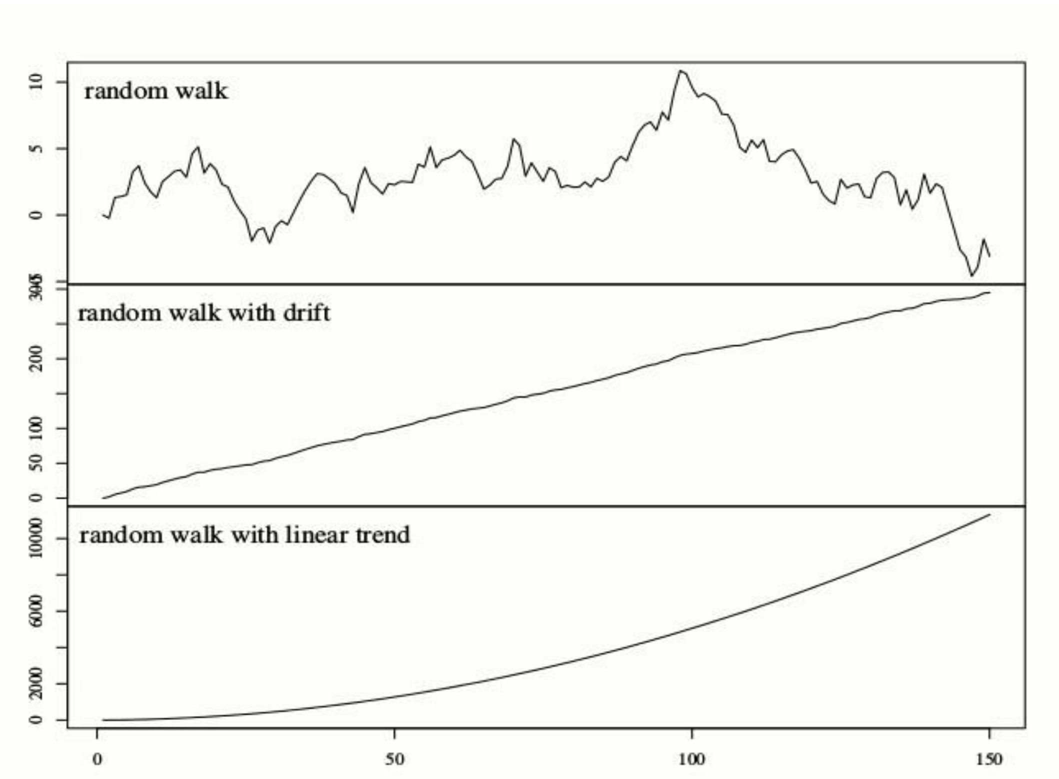
In other words , a Bernoulli Process is a series of independent and identically distributed random variables , with each coin flip representing a Bernoulli Trial .

The other types of stochastic processes are **Weiner Process**, which is based on Brownian motions , where Brownian motions are the random movement of particles in a fluid . It is a real valued continuous time stochastic process. **Poisson process** is another type of random process. It is a model for a series of discrete events where the average time between events is known , but the exact timing of events is random. There are examples for this process which usually happens in the case of radioactive decay of atoms , customers calling to a call center , or visitors to a website.

Gaussian Processes are the stochastic processes where every finite collection of random variables has a multivariate normal distribution ie every linear combination of them is normally distributed ie around its mean value, this probability distribution is symmetrical.

Random walks is a stochastic process where we determine the probable location of a point subject to random motions. **Markov Processes** are the one we are going to discuss in more detail.

Many stochastic processes of theoretical and applied interest posses the property that, given the present state of the process, the past history does not affect conditional probabilities of events defined in terms of the future. Such processes are known as Markov Processes.



Chapter 3

Markov Chains

A **Markov Process** is a stochastic model describing the sequence of possible events in which the probability of each event depents only on the state attained in the previous event.

Andrei A. Markov (1856-1922)

Markov Chains are discrete parameter Markov Processes whose state space is finite or countably infinite. The Markov Property in terms of probablity is defined as

P[Future | Present, Past] = P[Future | Present)

ie, the future probablity of an event or experiment is only affected by the probablity of the present. In other words, Markov Property says that the occurrence of any event bothers only what just happened before it. It is a random process in which the future is independent of the past, given the present. (this property is also called `memoryless' property).

Markov chains are named after the Russian Mathematician Andrei Andreevich Markov (1856 -1922) . Markov's motivation in writing the Markov chain papers was to show that the two classical theorems of probability theory, the Weak Law of Large Numbers and the Central Limit Theorem, could be extended to sums of dependent random variables.

Markov Process and Markov Chains

A **Markov Chain** is a 'discrete Markov Process'. Here `discrete' refers that the state space

of possible values is finite or countable. A Markov process can be continuous or discrete.

In fact, a Markov process is more general than a Markov chain.
If a discrete time Markov process has a discrete state space, it is definitely a Markov Chain. If a continuous time Markov process has an uncountable state space, it is definitely not a Markov chain. Most often, a Markov process is called a Markov chain if the set of states is discrete. For example, if the state space is the set of all integers.

Markov processes are classified according to the nature of the time parameter and the nature of the state space. With respect to state space, a Markov process can be either a 'discretestate Markov process' or 'continuous-state Markov process'. A discrete-state Markov process is called a Markov chain. Similarly, with respect to time, a Markov process can be either a discrete-time Markov process or a continuous-time Markov process. Thus, there are four basic types of Markov processes:

1. Discrete-time Markov chain (or discrete-time discrete-state Markov process)

2. Continuous-time Markov chain (or continuous-time discrete-state Markov process)

3. Discrete-time Markov process (or discrete-time continuous-state Markov process)

4. Continuous-time Markov process (or continuous-time continuous-state Markov process)

This classification of Markov processes is illustrated in the below figure.

		State Space	
		Discrete	Continuous
Time	Discrete	Discrete-time Markov Chain	Discrete-time Markov Process
	Continuous	Continuous-time Markov Chain	Continuous-time Markov Process

Definition: A sequence of random variables $X_0, X_1,\ldots$ with values in a countable set S is a 'Markov chain' if at any time n, the future states (or values) $X_{n+1}, X_{n+2},\ldots$ depend on the history $X_0,\ldots,X_n$ only through the present state X_n.

Mathematically, we can define a markov chain as,

A stochastic process $X = \{Xn : n \geq 0\}$ on a countable set S is a Markov Chain if, for any i, $j \in S$ and $n \geq 0$,

$$\mathbf{P\{X_{n+1}=j \mid X_0,\ldots,Xn\} = P\{X_{n+1}=j \mid Xn\}} \qquad (1)$$

We take $$\mathbf{P\{X_{n+1}=j \mid Xn=i\} = pij\ .} \qquad (2)$$

The pij is the transition probability that the Markov chain jumps from state i to state j . Condition (1), called the Markov property, says that, at any time n, the next state X_{n+1} is conditionally independent of the past $X_0,\ldots,X_{n-1}$ given the present state Xn. In other words, the next state is dependent on the past and present only through the present state.

Markov chain in two states

Consider a machine that at the start of any particular day , it is either broken down or in operating condition. Assume that if the machine is broken down at the start of the nth day, the probability is p that it will be successfully repaired and in the operating condition at the start of the (n+1)th day.

Assume also that if the machine is in operating condition at the start of the nth day, the probability is q that it will have a failure causing it to be broken down at the start of the (n+1)th day. Let $\boldsymbol{\pi_0(0)}$ denote the probability that the machine is broken down initially, ie at the start of 0^{th} day.

Let the state 0 denote to the machine being broken down initially and let the state 1 correspond to the machine being in operating condition. Let Xn be the random variable denoting the state of the machine at time n. According to the above description

$$P(X_{n+1} = 1 \mid X_n = 0) = p,$$

$$P(X_{n+1} = 0 \mid X_n = 1) = q,$$

and

$$P(X_0 = 0) = \pi_0(0).$$

Since there are only two states 0 and 1, it follows that

$$P(X_{n+1} = 0 \mid X_n = 0) = 1 - p,$$

$$P(X_{n+1} = 1 \mid X_n = 1) = 1 - q,$$

And the probability $\boldsymbol{\pi_0(1)}$ of being initially in state is given by

$$\pi_0(1) = P(X_0 = 1) = 1 - \pi_0(0).$$

From this information we can easily compute $P(X_n = 0)$ and $P(X_n = 1)$

We observe that

$$\begin{aligned} P(X_{n+1} = 0) &= P(X_n = 0 \text{ and } X_{n+1} = 0) + P(X_n = 1 \text{ and } X_{n+1} = 0) \\ &= P(X_n = 0)P(X_{n+1} = 0 \mid X_n = 0) \\ &\quad + P(X_n = 1)P(X_{n+1} = 0 \mid X_n = 1) \\ &= (1 - p)P(X_n = 0) + qP(X_n = 1) \\ &= (1 - p)P(X_n = 0) + q(1 - P(X_n = 0)) \\ &= (1 - p - q)P(X_n = 0) + q. \end{aligned}$$

Now $P(X_0=0) = \pi_0(0)$, so

$$P(X_1 = 0) = (1 - p - q)\pi_0(0) + q$$

and

$$\begin{aligned} P(X_2 = 0) &= (1 - p - q)P(X_1 = 0) + q \\ &= (1 - p - q)^2\pi_0(0) + q[1 + (1 - p - q)]. \end{aligned}$$

It is easily seen that by repeating this procedure n times,

$$P(X_n = 0) = (1 - p - q)^n\pi_0(0) + q\sum_{j=0}^{n-1}(1 - p - q)^j. \tag{A}$$

In the trivial case $p = q = 0$, it is clear that for all n

$$P(X_n = 0) = \pi_0(0) \qquad \text{and} \qquad P(X_n = 1) = \pi_0(1).$$

Suppose now that $p + q > 0$. Then by the formula for the sum of a finite geometric progression,

$$\sum_{j=0}^{n-1}(1 - p - q)^j = \frac{1 - (1 - p - q)^n}{p + q}.$$

We conclude from (A) that

$$P(X_n = 0) = \frac{q}{p+q} + (1 - p - q)^n \left(\pi_0(0) - \frac{q}{p+q}\right), \quad \textbf{(B)}$$

And consequently we get

$$P(X_n = 1) = \frac{p}{p+q} + (1 - p - q)^n \left(\pi_0(1) - \frac{p}{p+q}\right). \quad \textbf{(C)}$$

Suppose that ***p*** and ***q*** are neither both equal to zero nor both equal to1. Then $\mathbf{0 < p + q < 2}$, which implies that $\mathbf{|\ 1-}$ p - q $\mathbf{|<1.}$ In this case let $\boldsymbol{n} \to \infty$ in (B) and (C), then we conclude that

$$\lim_{n\to\infty} P(X_n = 0) = \frac{q}{p+q} \quad \text{and} \quad \lim_{n\to\infty} P(X_n = 1) = \frac{p}{p+q}. \quad \textbf{(D)}$$

We can also obtain the probabilities q /(p + q) and p /(p + q) by a different approach. Suppose we want to choose $\boldsymbol{\pi_0(0)}$ and $\boldsymbol{\pi_0(1)}$ so that P(Xn=0) and P(Xn=1) are independent of n. It is clear from (B) and (C) that to do this we choose

$$\pi_0(0) = \frac{q}{p+q} \quad \text{and} \quad \pi_0(1) = \frac{p}{p+q}.$$

Thus we see that if Xn, n ≥ 0, starts out with the initial distribution

$$P(X_0 = 0) = \frac{q}{p+q} \quad \text{and} \quad P(X_0 = 1) = \frac{p}{p+q},$$

then for all *n*

$$P(X_n = 0) = \frac{q}{p+q} \quad \text{and} \quad P(X_n = 1) = \frac{p}{p+q}.$$

The description of the machine is vague because it does not really say whether Xn, n ≥ 0, can be assumed to satisfy the Markov property. Let us suppose, however, that the Markov property does hold. We can use this added information to compute the joint distribution of X0, X1, . . .Xn.

For example, let n = 2 and let x0, x1, x2 each equal to 0 or 1. Then

$$\begin{aligned} &P(X_0 = x_0, X_1 = x_1, \text{and } X_2 = x_2) \\ &\quad = P(X_0 = x_0 \text{ and } X_1 = x_1)P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1) \\ &\quad = P(X_0 = x_0)P(X_1 = x_1 \mid X_0 = x_0)P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1). \end{aligned}$$

Now P(X0=x0) and P(X1=x1 | X0=x0) are determined by p,q and $\pi_0(0)$, but without Markov property, we cannot evaluate P(X2=x2 | X0=x0 and X1=x1) in terms of p, q and $\pi_0(0)$. If the Markov property is satisfied, then

$$P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1) = P(X_2 = x_2 \mid X_1 = x_1),$$

which is determined by p and q. In this case

$$\begin{aligned} &P(X_0 = x_0, X_1 = x_1, \text{and } X_2 = x_2) \\ &\qquad = P(X_0 = x_0)P(X_1 = x_1 \mid X_0 = x_0)P(X_2 = x_2 \mid X_1 = x_1). \end{aligned}$$

For example,

$$\begin{aligned} &P(X_0 = 0, X_1 = 1, \text{and } X_2 = 0) \\ &\qquad = P(X_0 = 0)P(X_1 = 1 \mid X_0 = 0)P(X_2 = 0 \mid X_1 = 1) \\ &\qquad = \pi_0(0)pq. \end{aligned}$$

The remaining entries are given in the following table. It gives the joint distribution of X0, X1, and X2.

x_0	x_1	x_2	$P(X_0 = x_0, X_1 = x_1, \text{and } X_2 = x_2)$
0	0	0	$\pi_0(0)(1 - p)^2$
0	0	1	$\pi_0(0)(1 - p)p$
0	1	0	$\pi_0(0)pq$
0	1	1	$\pi_0(0)p(1 - q)$
1	0	0	$(1 - \pi_0(0))q(1 - p)$
1	0	1	$(1 - \pi_0(0))qp$
1	1	0	$(1 - \pi_0(0))(1 - q)q$
1	1	1	$(1 - \pi_0(0))(1 - q)^2$

Transition probabilities

The outcomes obtained in a markov chain are called 'states'. If Xn has the outcome j, ie Xn = j, the process is said to be at state j at the nth trial. To a pair of states (j, k) at the two successive trials (say nth and (n + 1)st trials) there is an associated conditional probability p_{jk}. It is the probability of transition from the state j at nth trial to the state k at (n + 1)st trial. The transition probabilities p_{jk} are basic to the study of the structure of Markov chain.

Definition: Let {X0, X1, X2, ….} be a Markov chain with state space S. The transition probabilities of the Markov chain are

$$p_{ij} = P(X_{t+1}=j \mid X_t=1) \text{ for } i, j \in S, t = 0, 1, 2, \ldots$$

Transition Matrices

The Matrix representation of transition probabilities of a markov chain is known as a **transition matrix.**

The matrix describing the Markov chain is called the transition matrix. It is the most important tool for analysing Markov chains. The matrix is called state transition matrix or transition probability matrix and is usually denoted by P. Assuming that the states are { 1,2,3,……… } and the transition probabilities are p_{jk} where j, k = 1,2,3,…… the transition matrix is given as

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

These are the probabilities of transition from state j to state k. Note that the transition probabilities p_{jk} satisfy the given conditions ie

$$p_{jk} \geq 0$$

and

$$\sum_k p_{jk} = 1 \text{ for all } j.$$

We see here that the entries in each row adds up to 1 ie the sum of transition probabilities of each row is 1 where each of the transition probabilitiy is greater than or equal to 0.

P is a Stocastic matrix ie a square matrix with non negative elements and unit row sums used to describe the markov chain. A stochastic matrix is also known as markov matrix.

The transition probability may or may not be independent if n (n is the number of trials involved in a markov chain). If the transition probability p_{jk} is independent of n, the Markov chain is said to

be `homogenous' or to have `stationary transition probabilities'. If it is dependent on n, the chain is said to be non-homogenous.

The transition probability p_{jk} refers to the states (j, k) at two successive trials (say nth and (n+1)st trial, the transition is one-step and p_{jk} is called `one-step or unit step transition probability'. In the more general case, we are concerned with the pair of states (j, k) at two non successive trials, ie state j at the nth trial and state k at the (n+m)th trial. The corresponding transition probability is then called `m-step transition probability' and is denoted by $p_{jk}^{(m)}$ ie,

$$p_{jk}^{(m)} = \Pr\{X_{n+m} = k \mid X_n = j\}.$$

Example 3.1

Suppose a coin is tossed indefinitely. The probability for obtaining a head is p. Let X_n be the outcome of the nth trial, ie k where k = 0,1,2……….n. It denotes that there is a run of k successes, ie the length of the uninterrupted block of heads is k. $\{ X_n , n \geq 0 \}$ constitutes a Markov chain, with unit step transition probabilities

$$p_{jk} = \Pr\{X_n=k / X_{n-1}=j\} = p, \quad k= j+1$$

$$= q, \quad k= 0$$

$$= 0, \quad \text{otherwise}$$

The transition matrix is given by

$$\begin{array}{lcc}
 & & \text{States of } X_n \\
\text{States of } X_{n-1} & & \begin{array}{ccccccc} 0 & 1 & 2 & \cdots & k & k+1 & \cdots \end{array} \\
 & \begin{array}{c} 0 \\ 1 \\ 2 \\ \cdot \\ \cdot \\ \cdot \\ k \\ \cdot \\ \cdot \\ \cdot \end{array} & \begin{bmatrix}
q & p & 0 & \cdots & 0 & 0 & \cdots \\
q & 0 & p & \cdots & 0 & 0 & \cdots \\
q & 0 & 0 & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
q & 0 & 0 & \cdots & 0 & p & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots
\end{bmatrix}
\end{array}$$

Example 3.2

Let $\{x_n , n \geq 0\}$ be a Markov chain with three states 0,1,2 and with transition matrix

$$\begin{pmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 3/4 & 1/4 \end{pmatrix}$$ and

the initial distribution $\Pr\{X_0 = i\} = 1/3$, $i = 0,1,2$

We have the following probability distributions;

$\Pr\{X_1=1 \mid X_0=2\} = 3/4$

$\Pr\{X_2=2 \mid X_1=1\} = 1/4$

$\Pr\{X_2=2, X_1=1 \mid X_0=2\} = \Pr\{X_2=2, X_1=1\} \Pr\{X_1=1 \mid X_0 =2\} = 1/4 . 3/4 = 3/16$
$\Pr\{X_2=2, X_1=1, X_0=2\} = \Pr\{X_2=2, X_1=1 \mid X_0=2\} \Pr\{X_0=2\} = 3/16. 1/3 = 1/16$

$\Pr\{X_3=1, X_2=2, X_1=1, X_0=2\} = \Pr\{X_3=1 \mid X_2=2, X_1=1, X_0=2\} .\Pr\{X_2=2, X_1=1, X_0=2\}$

$= \Pr\{X_3=1 \mid X_2=2\}.1/16 = 3/4 . 1/16 = 3/64$

[By the Markov property, $\Pr\{X_3=1 \mid X_2=2, X_1=1, X_0=2\}$ becomes $\Pr\{X_3=1 \mid X_2=2\}$]

Order of a Markov chain

Definition: A Markov chain $\{X_n\}$ is of order s where s = 1,2,3……, if for all n

$$\Pr\{X_n = k \mid X_{n-1} = j, X_{n-2} = j_1, \ldots, X_{n-s} = j_{s-1}, \ldots\}$$
$$= \Pr\{X_n = k \mid X_{n-1} = j, \ldots, X_{n-s} = j_{s-1}\}.$$ whenever

the LHS is defined.

A Markov chain is $\{X_n\}$ is said to be of order one (or simply a Markov chain) if

$$\Pr\{X_n = k \mid X_{n-1} = j, X_{n-2} = j_1, \ldots\} = \Pr\{X_n = k \mid X_{n-1} = j\}$$
$$= p_{jk}$$

whenever $\Pr\{X_{n-1}=j, X_{n-2}=j_1, \ldots\ldots\} > 0$

A chain is said to be of order zero if $p_{jk} = p_k$ for all j. This implies independence of X_n and X_{n-1}.

Order means the dependence on the history. First order Markov property says that the future state depends only on the current. Second order Markov property says that the future depends on current state and the previous state. Mathematically,

First order: $P[S_{t+1} \mid S_t,\ldots,S_2,S_1] = P[S_{t+1} \mid S_t]$ $P[S_{t+1} \mid S_t,\ldots,S_2,S_1] = P[S_{t+1} \mid S_t]$

Second order: $P[S_{t+1} \mid S_t,\ldots,S_2,S_1] = P[S_{t+1} \mid S_t,S_{t-1}]$ $P[S_{t+1} \mid S_t,\ldots,S_2,S_1] = P[S_{t+1} \mid S_t,S_{t-1}]$ Similarly, the nth order Markov is following:

nth order: $P[S_{t+1} \mid S_t,\ldots,S_2,S_1] = P[S_{t+1} \mid S_t,S_{t-1},\ldots,S_{t-(n-1)}]$

Chapman Kolmogorov equation

Chapman Kolomogorov equations helps us to find the higher step transition probabilities using one step transition probabilities. Then the 1-step transition probabilities determine the n-step transition probabilities, for any n. This fact is contained in what are known as the ChapmanKolmogorov Equations.

The one step or unit step transition probability is the probability of X_n given X_{n-1} ie the probability of the outcome at the nth step or trial given the outcome at the previous step. P_{jk} gives the probability of unit step transition from the state j at a trial to the state k at the next following trial.

The m step transition probability is denoted by

$$\Pr\{X_{m+n} = k \mid X_n = j\} = p_{jk}^{(m)}$$

$p_{jk}^{(m)}$ gives the probability that from the state j at nth trial, the state k is reached at (m+n)th trial in m steps, ie the probability of transition from the state j to the state k in exactly m steps. The one step transition probabilities $p_{jk}^{(1)}$ are denoted by p_{jk} for simplicity. Consider

$$p_{jk}^{(2)} = \Pr\{X_{n+2} = k \mid X_n = j\}$$

The state k can be reached from the state j in two steps through some intermediate state r. Consider a fixed value of r, then we have

$$\begin{aligned}
\Pr(X_{n+2} &= k, X_{n+1} = r \mid X_n = j\} \\
&= \Pr\{X_{n+2} = k \mid X_{n+1} = r, X_n = j\} \Pr\{X_{n+1} = r \mid X_n = j\} \\
&= p_{rk}^{(1)} p_{jr}^{(1)} = p_{jr}\, p_{rk}.
\end{aligned}$$

Since the intermediate states r can assume values r = 1,2,………, we have

$$\begin{aligned}
p_{jk}^{(2)} = \Pr\{X_{n+2} = k \mid X_n = j\} &= \sum_r \Pr\{X_{n+2} = k, X_{n+1} = r \mid X_n = j\} \\
&= \sum_r p_{jr}\, p_{rk}
\end{aligned}$$

(summing over for all the intermediate states).

By induction , we have

$$\begin{aligned}
p_{jk}^{(m+1)} &= \Pr\{X_{n+m+1} = k \mid X_n = j\} \\
&= \sum_r \Pr\{X_{n+m+1} = k \mid X_{n+m} = r\} \Pr\{X_{n+m} = r \mid X_n = j\} \\
&= \sum_r p_{rk} p_{jr}^{(m)}.
\end{aligned}$$

Similarly we get,

$$p_{jk}^{(m+1)} = \sum_r p_{jr} p_{rk}^{(m)}.$$

In general, we have

$$p_{jk}^{(m+n)} = \sum_r p_{rk}^{(n)} p_{jr}^{(m)} = \sum_r p_{jr}^{(n)} p_{rk}^{(m)}.$$

These are the Chapman Kolmogorov equations, which is satisfied by the transition probabilities of a Markov chain. From the above equation, we get

$$p_{jk}^{(m+n)} \geq p_{jr}^{(m)} p_{rk}^{(n)}, \quad \text{for any } r.$$

We can put the results in terms of transition matrices as follows. Let P = (p_{jk}) denote the transition matrix of the unit step transitions and $p^{(m)}$ = ($p_{jk}^{(m)}$) denote the transition matrix of the m step transitions. For m=2, we have the matrix $P^{(2)}$ whose elements are given by the two-step transition probabilities $p_{jk}^{(2)}$. The elements of $P^{(2)}$ are the elements of the matrix obtained by multiplying the matrix P by itself, ie

$$P^{(2)} = P \cdot P = P^2.$$

Similarly,

$$P^{(m+1)} = P^m \cdot P = P \cdot P^m$$ and

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

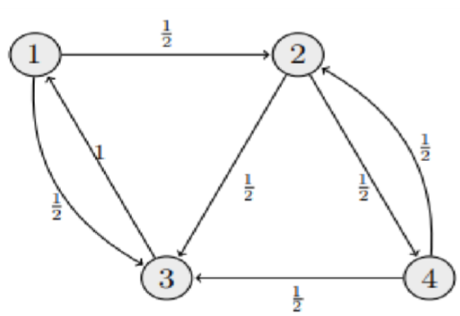
Markov chains as graphs

The Markov chain can be expressed graphically using the transition probabilities. The states of a Markov chain may be represented by the vertices (nodes) of the graph. The one step transitions between the states can be denoted by directed arcs. If i is the transition state from j (i →j), then vertices i and j are joined by a directed arc with arrow towards j, where the value of p_{ij}, is the arc weight, indicated by the directed arc.

If S = { 1,2,3,…..m} is the set of vertices corresponding to the state space of the chain and A is the set of directed arcs between these vertices, then the graph G = { S, A } is the directed graph or digraph or transition graph of the markov chain. A digraph such that its arc weights are positive and sum of the arc weights of the arcs from each node is unity is called a 'stochastic graph'.

A transition graph is of great aid in visualizing a Markov chain. It is a useful tool in studying the properties of the chain.

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



The corresponding transition matrix is

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \end{array}.$$

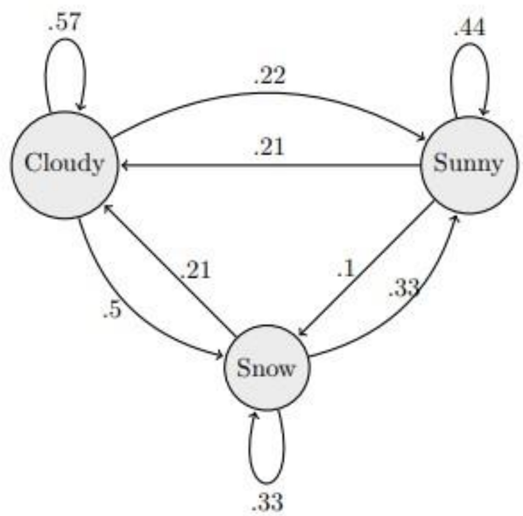
where the states are 1,2,3 and 4. The transition probabilities are $p_{11} = 0$, $p_{12} = ½$, $p_{13} = ½$, $p_{14} = 0$, $p_{21} = 0$, $p_{22} = ½$, $p_{23} = 0$, $p_{24} = ½$, $p_{31} = 1$, $p_{32} = 0$, $p_{33} = 0$, $p_{34} = 0$, $p_{41} = 0$, $p_{42} = ½$, $p_{43} = ½$, $p_{44} = 0$.

Note that here $\sum p_{ij} = 1$ for all i,j = 1,2,3,4.

Applications on Markov chain

1.Weather Prediction

Markov chains have been researched heavily for predicting weather. Studies have found out that Markov chains are useful in predicting future weather state using present weather condition. 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. Given below is a transition graph depicting the probabilities of different weather states.



The state diagram using a directed graph

The transition matrix is given as

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

We need to find out the probability that if today cloudy what is the weather condition after three days.

By observation, today's weather is cloudy. This is represented by the vector,

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

Then

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

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

This matrix gives the transition probabilities of weather states after three days, ie based on the conditional probabilities we can say

Pr(coudy after 3 days |today is cloudy) = 0.44

Pr(sunny after 3 days | today is cloudy) = 0.30

Pr(snowy after 3 days | today is cloudy) = 0.23

2. Gamblers ruin

The Gambler's Ruin Problem in its most basic form consists of two gamblers *A* and *B* who are playing a probabilistic game multiple times against each other. Every time the game is played, there is a probability p ($0 < p < 1$) that gambler *A* will win against gambler *B*. The probability that gambler *B* will win is $q = 1 - p$. Each gambler also has an initial wealth that limits how much they can bet. The total combined wealth is denoted by k and gambler *A* has an initial wealth denoted by i, which implies that gambler *B* has an initial wealth of $k - i$. Wealth is required to be positive.

The last condition we apply to this problem is that both gamblers will play indefinitely until one of them has lost all their initial wealth and thus cannot play anymore.

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 1-p & 0 & p & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1-p & 0 & p \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$$

The general transition diagram and matrix of gamblers ruin problem

Two players A and B play a game with the following rules:

- A fair coin is tossed. Player A wins on heads, and Player B wins on tails.
- The player that loses transfers one unit of money to the winner, and then the game is continued, and the coin is tossed again. ○ The game terminates when one of the players is ruined, i.e., has zero money left to play with.

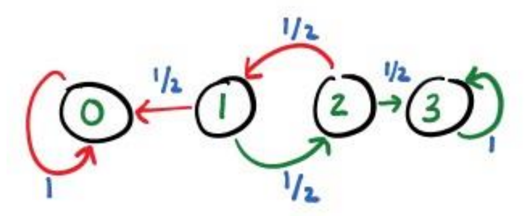
The money they start with is R_A and R_B respectively. We need to find the probability of winning of a player.

Markov chain modeling:

Let us simplify the problem by setting small, concrete values of R_A and R_B to 1 and 2 respectively. At each step of the game, R_A can increment or decrement by one, with equal probabilities. Since the total money in game is 3, there are four possibilities for R_A in the game ie {0, 1, 2, 3}. The game ends when R_A is either 0 (A loses all the money) or 3 (B loses all the money). The game continues for the other two states.

Let us denote this game as a state transition diagram. A state represents the current amount of money A has, and an edge represents a transition between states. Note that the state of player A alone is sufficient to capture the state of the game, as the total money in the game is constant, i.e, RA+RB=3. The following figure shows the state transition diagram. A node denotes the money

that player A has, i.e., RA. When the game is in state 2, player A can lose to go to state 1 or she may win to go to state 3. The probability in states 1 and 2 for winning and losing the current step is determined by tossing a fair coin and is 0.5 each. The states 0 and 3 indicate termination of the game. Once these states are attained there is no way going to another state.



The transition distribution for the game is the following:

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} \\ p_{10} & p_{11} & p_{12} & p_{13} \\ p_{20} & p_{21} & p_{22} & p_{23} \\ p_{30} & p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Let V denote the probability distribution of states, i.e., probability of the game being in any particular state. It is a 1x4 matrix, representing the probability of each state. If V_t is the probability distribution of states at time t, then V_{t+1} can be computed by the following equation:

$$V_{t+1} = V_t P$$

We are given that the initial value of P_A is 1, therefore, the initial probability distribution is

$$\nu_0 = [0, 1, 0, 0]$$

We can compute the probability distribution after one step, two steps, three steps, ten steps and hundred steps as follows:

$$\nu_1 = \nu_0 P = [0.5, 0, 0.5, 0]$$

$$\nu_2 = \nu_0 P^2 = [0.5, 0.25, 0, 0.25]$$

$$\nu_3 = \nu_0 P^3 = [0.625, 0, 0.125, 0.25]$$

$$\nu_{10} = \nu_0 P^{10} = [0.66601562, 0.00097656, 0, 0.33300781]$$

$$\nu_{100} = \nu_0 P^{100} = [\approx 0.6666, \approx 0, \approx 0, \approx 0.3333]$$

V_{100} denotes that after 100 steps the probability that the game is in losing state (state 0) is 2/3, and the game is in winning state (state 3) is 1/3. Moreover, the probability that game is still not terminated is almost zero, i.e., the probability being in states 1 and 2.

Absorbing Markov chain

A state si of a Markov chain is called absorbing if it is impossible to leave it (i.e., p_{ij}= 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 3.3:

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

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

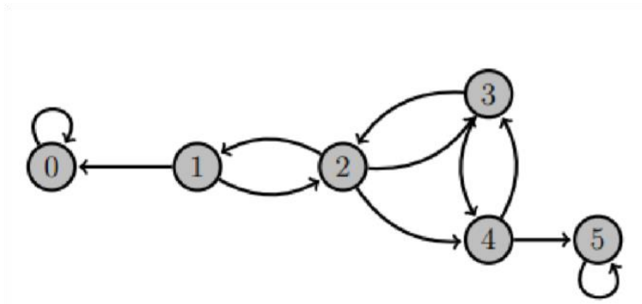
Definition:In an absorbing Markov chain, a state which is not absorbing is called **transient.**

In above example , A1 and A3 are transient states.

Example3.4

Drunkards walk: The example of Drunkard's walk 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,

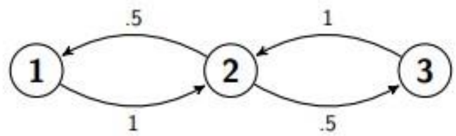
$$P= \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

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.

Ergodic and Regular Markov Chain

Definition: 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.



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: 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).

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.

Example3.5:

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

In this example the transition matrix T does not have all positive entries. But it is a regular Markov chain because T^2 has only positive entries.

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

Stationary distribution

Theorem: 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: A **stationary distribution** w is a (row) vector, whose entries are nonnegative and sum to 1, is unchanged by the operation of transition matrix P on it and so it satisfies,

$$wP = w.$$

where $\sum_i w_i = 1$

Here w is a normalized left eigenvector of the transition matrix P with an eigenvalue of 1. Now we can find the stationary distribution of the following transition matrix .

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

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, w1 = .8 and w2 = .2. Thus

$$w = \begin{bmatrix} .8 & .2 \end{bmatrix}.$$

Chapter 4

Google PageRank

Google is a widely used search engine engine to gather information. There are a lot of web pages associated with each web search that we are doing and they are placed according to their page rank by google.

Google uses the algorithm called PageRank to rank the webpages of a search result. The algorithm was developed by the founders of google, Larry Page and Sergey 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 (vertices) 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.

Page Rank and Markov chains

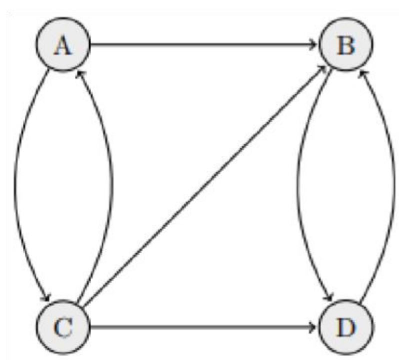
To find out the page score one must consider that the surfer can select any page. However it is not always the case that they select the pages sequentially. Most of the time, a surfer will follow links from a page sequentially, i.e. from a page i the surfer will follow the outgoing links and move on 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. The page rank algorithm is modelled as the behavior of a randomized web surfer; this model can be seen as Markov chain to predict the behavior of a system that travels from one state to another state considering only the current condition.

Definition: Suppose that page Pj has L(j) links. If one of those links is to page Pi , then Pj will pass on 1/ L(j) of its importance to Pi . The importance ranking of Pi 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 Pi by B_u, then then the PageRank of Pi , PR(i), is given by the formula

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

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,

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

We are going to assign each node or webpage an initial PageRank 1/4 . Then we are using the simplified PageRank algorithm ie,

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

Then we calculate the PageRank of each node.

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

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

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

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.

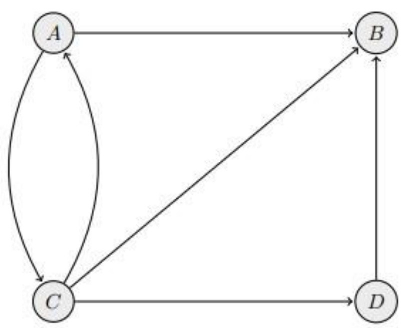
Definition: A node is called a ‘dangling node’ if it does not contain any out-going link.

Example4.1.

Shown below is the matrix representation of the graph. For instance, node B in this example is a dangling node.

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

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 1/4 , then our new matrix representation,



Then our new matrix is,

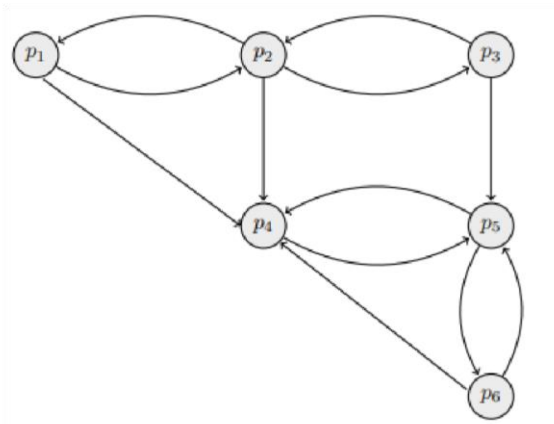
$$T = \begin{bmatrix} 0 & 1/4 & 1/3 & 0 \\ 1/2 & 1/4 & 1/3 & 1 \\ 1/2 & 1/4 & 0 & 0 \\ 0 & 1/4 & 1/3 & 0 \end{bmatrix}.$$

The 1/4 corresponds to jumping randomly to any website with equal probability.

Definition: 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.

Example4.2

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.

In order to calculate PageRanks properly for a reducible web graph, there is 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 p1, p2, ..., pN are the pages under consideration, M(pi) is the set of pages that link to pi , L(pj) is the number of outbound links on page pj , 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'.

Example4.3:

For previous example, calculating the PageRank using the corrected formula for p2 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 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 p5 has the highest rank. Page p5 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 p5 will appear early in the Google search and page then p4 and so on.

CONCLUSION

Randomness is part of daily life. The mathematical implementation on the random changes is give rise to concept of random variables and their sequences. We are interested in how a random variable changes over time. The study of how a random variable changes over time includes stochastic processes.

A Markov model is a stochastic model used to model randomly changing systems where it is assumed that future states depend only on the present state and not on the sequence of events that preceeded it. A mathematical process that is stochastic, discrete, and has the property that the probability that a particular outcome occurs depends only on the previous outcome is called a Markov chain. It can be visualized by a chain or directed graph. Also it can be analyzed by matrix theory.

Many of the artificial intelligence tools use the principle of Markov chain in some form. In simple words, Markov analysis is a technique of dealing with future occurences with currently known probabilities. Hence there are numerous applications in Business, in the case of share market analysis, bad debt prediction, machine breakdown prediction, university enrolment etc.

REFERENCES

- Stocastic Processes, J. Medhi, second edition.
- Introduction to Stocastic Processes, Hoel Port Stone.
- Textbook of Complimentary Statistics, L.R. Pillai.
- Mahfuz, Fariha, "MARKOV CHAINS AND THEIR APPLICATIONS" (2021). Math Theses. Paper 10. http://hdl.handle.net/10950/3704