

Islamic University of Gaza
Deanery of Higher Studies
Faculty of Science
Department of Mathematics

On the Random Walk Metropolis Algorithm

Presented By

Duaa Hatem Mohamed Maima

Supervisor

Professor Mohamed I. Riffi

A DISSERTATION
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF MATHEMATICS
2015

© Copyright by Duaa H. Maima (2015)
All Rights Reserved

Dedication

To...

My Parents;

My Husband;

Rami

My Brothers;

Mohamed, Mahmoud, Alaa, Hisham.

My Sisters;

Esraa, Wala and her husband.

Contents

Introduction	1
1 Markov Chain	3
1.1 Definition and Some Properties	4
1.2 Transition Probability Matrix	6
1.3 Multistep Transition Probability	8
1.4 Important Properties of Markov Chains	9
1.4.1 Irreducible and Aperiodic Markov Chains	13
1.5 Stationary Distributions	16
1.6 Convergence Theorems	19
1.6.1 Detailed Balance and Time Reversal	20
2 Random Walk	23
2.1 Random Walk as Markov Chain	24
2.2 Transition matrix	25
2.3 Stationary distributions	26
2.4 The One-Dimensional Random Walk	26
2.5 Higher Dimensions Random Walk	29
3 Markov Chain Monte Carlo	32
3.1 Markov Chain Monte Carlo Methods	32
3.2 Metropolis Algorithm	33
3.3 Metropolis-Hastings Algorithm	34
3.4 The Random Walk Metropolis Algorithm	37

4	Simulation	41
4.1	Markov Chain Simulation	41
4.2	MCMC Simulation	42
4.2.1	Metropolis-Hasting Algorithm	42
4.2.2	Random Walk Metropolis	45
4.3	conclustion	55
	Appendix	56
.1	R code of Markov Chain Simulation	56
.2	The R code of Random Walk Metropolis Algorithm	57
.3	R code of Normal Metropolis-Hasting	59
.4	R code of Males Heights Example	63
	Bibliography	67

List of Figures

1.1	Drunkards Walk	7
1.2	Frog hopping	12
1.3	Null Recurrent	13
2.1	Random Walk	26
2.2	Gamblers Ruin	29
2.3	Two Dimensions Random Walk	30
2.4	The moving mouse	31
4.1	Chain of x	42
4.2	Target Sample with $r=.9$	43
4.3	Target Sample with $r=.99$	44
4.4	Chain 1 with starting value $=-2$	45
4.5	Chain 2 with start value $=1$	46
4.6	Chain 3 with start value $=4$	46
4.7	Rang of 3 chains	47
4.8	Histogram of males heights	49
4.9	Target sample with starting point= 63 inches	52
4.10	Target sample with starting point= 70 inches	52
4.11	Target sample with starting point= 66 inch	53
4.12	Rang of 3 chains	53

List of Tables

4.1	Real males heights	48
4.2	3 chains of males heights produced by Metropolis-Hasting	50
4.3	3 chains of males heights produced by Metropolis-Hasting	51

List of Symbols

Z	integers.
\mathbb{R}	real numbers.
N	natural number.
\mathbf{P}	transition matrix.
P	probability measure.
$p(i, j)$	probability of moving from the state i to the state j in one step.
S	state space.
p_i	probability mass function of X_0 .
f_{ii}^n	probability of first recurrence to i at the n^{th} step.
f_i	probability of recurrence to i .
π	stationary distribution.
x^*	candidate point.

List of Abbreviations

i.i.d.	independent identically distributed.
MC	Markov Chain.
MCMC	Markov Chain Monte Carlo.
gcd	greatest common divisor.
MH	Metropolis-Hasting.
RW	Random Walk.
RWM	Random Walk Metropolis.
$\text{rem}(a, b)$	The remainder of the division of a by b .
$\max(a, b)$	maximum value between (a) and (b) .
$\min(a, b)$	minimal value between (a) and (b) .

Acknowledgements

First of all, praise is to God who gives me the chance to complete my Master Degree. I am grateful to my supervisor, Professor Mohamed I. Riffi, for a substantial amount of guidance, advice and support in producing this thesis.

I would like to thank all members of Mathematics Department in the Islamic University of Gaza for their help.

Finally, I would like to express my deep thanks to my family for their encouragement and support.

Abstract

The random walk Metropolis algorithm belongs to the collection of Markov Chain Monte Carlo (MCMC) methods that are used in statistical inference.

It is one of the most common Markov Chain Monte Carlo methods in practical use today.

We would like, in this thesis, to introduce the discrete time Markov chains as stochastic processes having the Markov property . We also present some properties of the Markov chains that are needed to the random walk Metropolis algorithm and related to the Markov Chain Monte Carlo methods such as the detailed balance, irreducibility, and aperiodicity properties .And we introduce the random walk as stochastic process, and present some examples of the random walk having the Markov property. we will introduce some of the basic algorithms that belong to the Markov Chain Monte Carlo methods, and we explore the theoretical properties of the random walk Metropolis algorithm for certain kinds of target random variables.

Theoretical properties of the random walk Metropolis algorithm for certain special classes of target have been investigated extensively.

We will also describe and study some of the related derived results that have important practical implications.

We will also demonstrate the impact of the random walk Metropolis algorithm for some practical examples using the R programming language in simulation.

خوارزمية متروبوليس للحركة العشوائية تنتمي إلى مجموعة من سلسلة ماركوف التي يتم استخدامها في الاستدلال الإحصائي. وهي واحدة من سلسلة ماركوف الأكثر شيوعاً في الاستخدام العملي اليومي. ونود، في هذه الأطروحة، أن نعرض سلاسل ماركوف

للزمن المنفصل كعمليات عشوائية تتوفر لديها خاصية ماركوف. كما نقدم بعض خصائص سلاسل ماركوف التي نحتاجها في خوارزمية متروبوليس للحركة العشوائية. كذلك سندرس الحركة العشوائية ، و نقدم بعض الأمثلة على الحركة العشوائية التي تتمتع بخاصية ماركوف. وسوف نقدم بعض الخوارزميات الأساسية التي تنتمي إلى طرق مونت كارلو للسلاسل العشوائية ، ونود استكشاف الخصائص النظرية من خوارزمية متروبوليس للحركة العشوائية لأنواع معينة من المتغيرات العشوائية. وقد تم دراسة الخصائص النظرية لخوارزمية متروبوليس للحركة العشوائية لبعض الفئات الخاصة المستهدفة . كذلك أيضا وصف ودراسة بعض النتائج المستمدة ذات الصلة التي لها آثار عملية هامة وكذلك توضيح تأثير خوارزمية متروبوليس للحركة العشوائية لبعض الأمثلة العملية باستخدام لغة برمجة R في المحاكاة.

Introduction

The Markov Chain Monte Carlo methods are used to generate random samples from a potentially complicated, or indirectly specified arbitrary probability distribution π by generating a Markov chain $(X_{(1)}, (X_{(2)}), \dots$ with stationary distribution π .

They are simulation based and enable the statisticians to examine and analyze data using realistic statistical models.

In this thesis, we will introduce the discrete time Markov chains as stochastic processes having the Markov property.

Then we will introduce some important MCMC methods with special emphasis on the random walk Metropolis algorithm, where its theoretical properties are investigated and studied.

Besides this introduction, this thesis has four chapters.

Chapter (1) is preparatory. We define Markov chains as stochastic processes having the Markov property. Then we define the transition probability matrix of the Markov chain and its stationary (limiting) distribution. We also present some properties of the Markov chains that are needed to the random walk Metropolis algorithm and related to the MCMC methods such as the detailed balance, irreducibility, and aperiodicity properties.

Chapter (1) has the following sections:

- Definition and Some Properties.
- Transition Probability Matrix.
- Multistep Transition Probability.

- Important Properties of Markov chains.
- Stationary Distributions.
- Convergence Theorems.

Chapter (2) is devoted to introducing the random walk as stochastic process. We present some examples of the random walk having the Markov property. We study their transition probability matrices and stationary distributions .

Chapter (2) contains the following sections:

- Random Walk as Markov Chain.
- Transition Matrix.
- Stationary Distributions.
- The One-Dimensional Random Walk.
- Higher Dimensions Random Walk.

In Chapter (3), we will introduce some of the basic algorithms that belong to the MCMC methods and vital to our discussion such as the Metropolis-Hasting algorithm and the random walk Metropolis algorithms.

Chapter (3) contains the following sections:

- Markov Chain Monte Carlo.
- Metropolis Algorithm.
- Metropolis-Hastings Algorithm.
- The Random Walk Algorithm.

In Chapter (4), we explore some of the related derived results of the random walk Metropolis algorithm that have important practical implications. We also demonstrate the impact of the random walk Metropolis algorithm for some practical examples using the R programming language in simulation .

Chapter (4) has the following sections:

- Simulation.
- Practical Examples.

Chapter 1

Markov Chain

A Markov chain is a mathematical model of a random phenomenon evolving with time in a way that the past affects the future only through the present. The time can be discrete (i.e. the integers), continuous (i.e. the real numbers). In Mathematics, a phenomenon which evolves with time in a way that only the present affects the future is called **a dynamical system**. Markov chains were introduced by the Russian Mathematician, Andrei Andreyevich Markov (1856 - 1922), a student of Chebyshev and a graduate of Saint Petersburg University (1878). During his fruitful life, Markov did extensive research in several branches of mathematics including probability theory. He was also interested in poetry and did some work in poetic styles. The topic of Markov chains is one of his ever-lasting contributions to the world of knowledge. He invented that branch of stochastic processes in an attempt to generalize the strong law of large numbers to cases in which the random variables are not independent.

The importance of Markov chains comes from two facts:

(i) There are a large number of physical, biological, economic, and social phenomena that can be described in this way.

(ii) There is a well developed theory that allows us to do computations. We begin with a famous example, then describe the property that is the defining feature of Markov chains.

An example from arithmetic of a dynamical system in discrete time is the one which finds the greatest common divisor $\gcd(a, b)$ between two positive

integers a and b . $\gcd(a, b) = \gcd(r, b)$, where $r = \text{rem}(a, b)$ is the remainder of the division of a by b . By repeating the procedure, we end up with two numbers, one of which is 0, and the other the greatest common divisor. Formally, we let our state be a pair of integers (x_n, y_n) , where $x_n \geq y_n$, initialised by $x_0 = \max(a, b)$, $y_0 = \min(a, b)$, and evolving as $x_{n+1} = y_n$, $y_{n+1} = \text{rem}(x_n, y_n)$, $n = 0, 1, 2, \dots$. In other words, there is a function F that takes the pair $X_n = (x_n, y_n)$ into $X_{n+1} = (x_{n+1}, y_{n+1})$. The sequence of pairs thus produced, X_0, X_1, X_2, \dots has the property that, for any n , the future pairs $(X_{n+1}, X_{n+2}, \dots)$ depend only on the pair X_n .

1.1 Definition and Some Properties

Suppose we have some experiments X_1, X_2, X_3, \dots , whose results (outcomes) fulfil the following properties:

- (1) Any outcome belong to a set of outcomes $\{s_1, s_2, \dots, s_m\}$, which we will call the sample space or the state space for this system. For example: if the outcome of the experiment numbered n is s_i , then we say the system is in state s_i at step n if $X_n = s_i$.
- (2) The outcome of any experiment is depends only upon immediately previous outcome.

A **random process** is a collection of random variables indexed by some set \mathbf{I} , taking values in some set \mathbf{S} . \mathbf{I} is the index set, usually time, e.g. $(\mathbb{Z}^+, \mathbb{R}^+ \dots)$. \mathbf{S} is the state space, e.g. $(\mathbb{Z}^+, \mathbb{R}^n, (1, 2, \dots, n), (a, b, c))$.

We classify random processes according to both the index set (discrete or continuous) and the state space (finite, countable or uncountable/continuous).

Definition 1.1. [8] Let S be a countable set, (i, j, k, \dots) . Each $(i \in S)$ is called a **state** and S is called the **state-space**.

A random process (stochastic process) is called a **Markov Process** if, conditional on the current state of the process, its future is independent of its past.

Then the conditional distribution of (X_{n+1}) given (X_1, \dots, X_n) depends on (X_n)

only. The set in which the X_i takes values is called the state space of the Markov chain. Since for every couple of states (s_i, s_j) , we can find the probability $p(i, j) = P(i \rightarrow j)$ which is the probability of moving from one state s_i to another state s_j at a given time n . More precisely.

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

Such these stochastic experiments are called finite **Markov Chains**.

The conditional probabilities $P(X_{t+1} = j \mid X_t = i)$ are called **(one-step)** transition probabilities.

If for each i and j , $P(X_{t+1} = j \mid X_t = i) = P(X_1 = j \mid X_0 = i)$ for all $t = (0, 1, 2, 3, \dots)$ then the (one-step) transition probabilities are said to be **stationary** and are denoted by p_{ij} . The stationary transition probability implies that the transition probabilities do not change over time.

Stochastic processes include i.i.d. sequences of random variables, random walks, and Markov chains.

Definition 1.2. [10] A Markov chain is a sequence of random variables

$$\{X_0, X_1, X_2, \dots\} : \Omega \longrightarrow S$$

defined on a probability space and mapping to a finite state space $S = \{s_1, \dots, s_m\}$ with the property (Markov property) the conditional probability distribution of the next future state X_{n+1} given the present and past states is a function of the present state X_n alone. i.e.,

$$\begin{aligned} P(X_{n+1} = s_j \mid X_0 = s_0, X_1 = s_1, \dots, X_n = s_i) \\ = P(X_{n+1} = s_j \mid X_n = s_i) = p(i, j) \end{aligned}$$

for all steps n , all states $s_0, s_1, \dots, s_i, s_j$.

A typical example for markov chain is a random walk (in two dimensions, the drunkards walk).

Definition 1.3. [4] A Markov chain (X_n) is said to be **time-homogeneous** if $P(X(s+t) = j \mid X(s) = i)$ is independent of s . When this holds, putting $(s = 0)$ gives,

$$P(X(s+t) = j \mid X(s) = i) = P(X(t) = j \mid X(0) = i).$$

Probabilities depend on elapsed time, not absolute time.

If the transition probabilities are fixed for all n , then the chain is considered homogeneous; meaning they don't change over time, so the probability of going from state s_i at time $(n + 1)$ to state s_j at time $(n + k + 1)$ is the same as the probability of going from state s_i at time n to state s_j at time $n + k$.

Definition 1.4. [3] If the transition probabilities above do not depend on t then we say that the Markov chain is **stationary**.

1.2 Transition Probability Matrix

Definition 1.5. [10] A $(k \times k)$ -matrix \mathbf{P} with elements $\{p(i, j) : i, j = 1, \dots, m\}$ is a **transition matrix** for a Markov chain with finite state space $S = \{s_1, \dots, s_m\}$ if $P(X_{n+1} = s_j \mid X_0 = s_0, X_1 = s_1, \dots, X_n = s_i) = P(X_{n+1} = s_j \mid X_n = s_i)$ for all $n = 0, 1, \dots$, all states $s_0, s_1, \dots, s_i, s_j \in S$. The elements of the transition matrix \mathbf{P} are called transition probabilities (**transition kernel**). The transition probability $p(i, j)$ is the conditional probability of being at state s_j "tomorrow" given that we are in state s_i "today".

$$\text{Suppose we have transition matrix } \mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

The elements of the first row represent the probabilities of moving to the different states if the current state is 1. Therefore, p_{11} represents the probability that state $X_{t+1} = 1$ if the current state is also 1; p_{12} represents the probability that state $X_{t+1} = 2$ if the current state is 1, etc...

Note that: every transition matrix is a stochastic square matrix with zero or positive elements less than or equal to one such that the summation of elements in each row is unity. i.e.,

$$p(i, j) \geq 0 \quad \text{and} \quad \sum_{j=1}^m p(i, j) = 1 \quad \text{for all } i, j \in \{1, \dots, m\}, \quad (1.1)$$

A stochastic matrix \mathbf{P} is **regular** if some matrix power \mathbf{P}^k contains only strictly positive entries.

Example 1.1. (A weather model)

The probabilities of weather conditions, given the weather on the preceding day, can be represented by a transition matrix

$$\mathbf{P} = \begin{matrix} & \begin{matrix} R & N & S \end{matrix} \\ \begin{matrix} R \\ N \\ S \end{matrix} & \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \end{matrix}$$

The matrix \mathbf{P} represents the weather model in which a rainy day (**R**) is %50 likely to be followed by another rainy day (**R**), a nice day (**N**) is %0 likely to be followed by another nice day (**N**), and a snowy (**S**) day is %50 likely to be followed by another snowy day (**S**).

$p(x, y)$ is the probability that, if a given day is of type x , it will be followed by a day of type y .

For Example: (p_{12}) is the probability that given a day is rainy, it will be followed by a nice day.

Example 1.2. Simple random walk (drunkards walk) in a city

Suppose that the drunkard is allowed to move in a city whose streets are laid down in a square pattern:

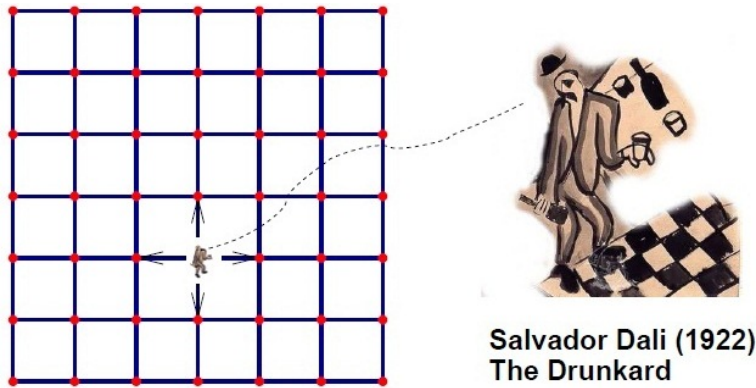


Figure 1.1: Drunkards Walk

Suppose that the drunkard moves from corner to corner. So he can move east, west, north or south, and lets say hes completely drunk, so we assign

probability ($\frac{1}{4}$) for each move. We may again want to ask similar questions. But, observe, that since there are more freedom now, there is clearly a higher chance that our man will be lost. These things are, for now, mathematically imprecise, but they will become more precise in the sequel.

Example 1.3. A frog hopping on 3 rocks. Put $S = (1, 2, 3)$.

$$\mathbf{P} = \begin{matrix} & \begin{matrix} R_1 & R_2 & R_3 \end{matrix} \\ \begin{matrix} R_1 \\ R_2 \\ R_3 \end{matrix} & \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{5}{8} & \frac{1}{8} & \frac{2}{8} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix} \end{matrix}$$

This example shows a frog movement on three rocks(R_1, R_2, R_3).

(p_{13}) is the probability that frog hopping from the rock 1(R_1) to the rock 3 (R_3) equal to %50.

Definition 1.6. [10]

In mathematics and statistics, a probability vector or a stochastic vector is a vector with non-negative entries that add up to one. Here are examples of probability vectors

$$\mathbf{P} = \begin{pmatrix} 0.25 \\ 0.5 \\ 0.25 \end{pmatrix} \text{ or } \mathbf{P} = \begin{pmatrix} 0.25 \\ 0.75 \end{pmatrix}.$$

1.3 Multistep Transition Probability

Definition 1.7. [11] The "**one-step**" transition probability is defined as $\mathbf{P}(X_{n+1} = s_j \mid X_n = s_i) = p(i, j)$, which is known as the "**short term**" behavior of a Markov chain. Suppose now that the probability that the system is in state s_i at an arbitrary time (step) n is $p_i^{(n)} = \mathbf{P}(X_n = s_i)$ which is known as the "**long term**" behavior of the Markov chain, and if this probability is a probability vector like $P^{(n)}$, then the initial probability distribution at time zero is $p_i^{(0)} = \mathbf{P}(X_0 = s_i)$ represented as a row vector of the state space probabilities at step zero given by

$$\begin{aligned}\mathbf{P}^{(0)} &= (p_1^{(0)}, p_2^{(0)}, \dots, p_m^{(0)}) \\ &= (\mathbf{P}(X_0 = s_1), \mathbf{P}(X_0 = s_2), \dots, \mathbf{P}(X_0 = s_m)).\end{aligned}\tag{1.2}$$

and $\sum_{i=1}^m p_i^{(0)} = 1$.

Often all elements of $\mathbf{P}^{(0)}$ are zero except for a single element of one, corresponding to the process starting in that particular state.

Similarly, the probability distribution at the first time (step) is

$$\mathbf{P}^{(1)} = (p_1^{(1)}, p_2^{(1)}, \dots, p_m^{(1)})$$

And the probability distribution of the Markov chain at time n is

$$\begin{aligned}\mathbf{P}^{(n)} &= (p_1^{(n)}, p_2^{(n)}, \dots, p_m^{(n)}) \\ &= (\mathbf{P}(X_n = s_1), \mathbf{P}(X_n = s_2), \dots, \mathbf{P}(X_n = s_m)).\end{aligned}$$

Suppose we know the transition matrix and the distribution of the initial state ($X_0 = x_i$), then we can compute $p^n(x_i, x_j) = \mathbf{P}(X_n = x_j | X_0 = x_i)$, the probability for moving from state x_i to state x_j at n steps exactly.

1.4 Important Properties of Markov Chains

In this section we will summarize some of the most common properties of Markov chains that are used in the context of MCMC. We always refer to a Markov chain $\{X_0, X_1, X_2, \dots\}$ with transition matrix \mathbf{P} on a finite state space $S = \{s_1, \dots, s_m\}$.

Suppose we know the transition matrix and the distribution of the initial state, then we can compute the probability for moving from state s_i to state s_j at n steps exactly. Let us show that the transition probabilities $p^n(i, j)$ satisfy the **Kolmogorov Chapman** Equation using the formula for total probability and the Markov property, let us denote that $p_i^{(n)} = \mathbf{P}(X_n = s_i)$ and $P^{(n)}$ be the distribution of the chain at time n with entries $p_i^{(n)}$.

Theorem 1.1. *[7]Kolmogorov Chapman Equation*

Let \mathbf{P} be the transition matrix of a Markov chain with initial distribution $P^{(0)}$, we have for any n that the distribution $P^{(n)}$ at time n satisfies

$$p^{(n)} = P^{(0)} \mathbf{P}^n$$

Proof. : we use induction.

Consider first the case $n = 1$

For $j = 1, \dots, m$, we have

$$\begin{aligned} p_j^{(1)} &= \mathbf{P}(X_1 = s_j) \\ &= \sum_{i=1}^m \mathbf{P}(X_0 = s_i, X_1 = s_j) \\ &= \sum_{i=1}^m \mathbf{P}(X_0 = s_i) \cdot \mathbf{P}(X_1 = s_j \mid X_0 = s_i) \\ &= \sum_{i=1}^m p_i^{(0)} p(i, j) \\ &= (p^{(0)} p)_j. \end{aligned}$$

where $(P^{(0)} P)_j$ denotes the j^{th} element of the row vector $P^{(0)} \mathbf{P}$. Hence $P^{(1)} = P^{(0)} \mathbf{P}$ Fix t and suppose that

$$P^{(t)} = P^{(0)} \mathbf{P}^t \quad \text{holds} \quad (1.3)$$

For $n = t + 1$, we have

$$\begin{aligned} P^{(t+1)}(i, j) &= \mathbf{P}(X_{t+1} = s_j \mid X_0 = s_i) \\ &= \sum_{k=1}^m \mathbf{P}(X_{t+1} = s_j, X_t = s_k \mid X_0 = s_i) \\ &= \sum_{k=1}^m \mathbf{P}(X_{t+1} = s_j \mid X_t = s_k, X_0 = s_i) \cdot \mathbf{P}(X_t = s_k, X_0 = s_i) \\ &= \sum_{k=1}^m \mathbf{P}(X_{t+1} = s_j \mid X_t = s_k) \cdot \mathbf{P}(X_t = s_k \mid X_0 = s_i) \\ &= \sum_{k=1}^m P^{(t)}(i, k) P(k, j) \\ &= (P^t P)_j. \end{aligned} \quad (1.4)$$

so that

$$\begin{aligned}
P^{(t+1)} &= P^{(t)}\mathbf{P} \\
&= P^{(0)}\mathbf{P}^t\mathbf{P} \quad \text{because of (1.3)} \\
&= P^{(0)}\mathbf{P}^{t+1}.
\end{aligned}$$

□

Equation (1.4) can be written in the matrix form as follows

$$P^{(n+1)} = P^{(n)}\mathbf{P}$$

since the rows sum to one, as $\sum_j P(i, j) = 1$ in the probability transition matrix \mathbf{P} . Now, we immediately see how to quickly iterate the Kolmogorov -Chapman equation, as

$$P^{(n)} = P^{(n-1)}\mathbf{P} = (P^{(n-2)}\mathbf{P})\mathbf{P} = P^{(n-2)}\mathbf{P}^2$$

Continuing in this fashion shows that

$$P^{(n)} = P^{(0)}\mathbf{P}^n.$$

In general, we define the n -step transition probabilities $P^{(n)}(i, j)$ by

$$P^{(n)}(i, j) = \mathbf{P}(X_{n+k} = s_j \mid X_k = s_i),$$

which is just the i, j^{th} element of the matrix \mathbf{P}^n , where \mathbf{P}^n is the n^{th} power of the matrix \mathbf{P} .

As we saw before, if the state space is finite, the transition probability distribution can be represented as a matrix, called the transition matrix, with the i, j^{th} element equal to $P(i, j) = \mathbf{P}(X_{n+1} = s_j \mid X_n = s_i)$. But the transition probabilities can be described by a (directed) graph whose vertices are the states, and an arrow from state s_i to state s_j with the number $P_{i,j}$ over it indicates that it is possible to pass from point i to point j with probability $P_{i,j}$. When $P_{i,j} = 0$, the corresponding arrow is omitted.

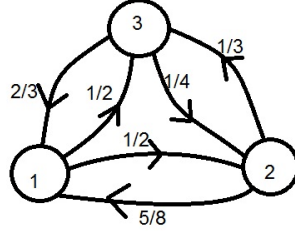


Figure 1.2: Frog hopping

by Example (1.3) we can represent the transition matrix as Figure (1.2).

Definition 1.8. [6] For a Markov chain with transition matrix \mathbf{P} , giving that $X_0 = s_i$, let

$$f^n(i, i) = \mathbf{P}(X_n = s_i, X_l \neq s_i, 1 \leq l \leq n-1)$$

, be the probability of first return to state s_i at time n , and for $i \neq j$,

$$f^n(i, j) = \mathbf{P}(X_n = s_j, X_l \neq s_j, 1 \leq l \leq n-1)$$

, is the probability of first arrival at state s_j at time n .

For each $s_i \in S$ let

$$f(i, i) = \sum_{n=1}^{\infty} f^n(i, i)$$

, be the probability that leaves state s_i sooner or later, return to that state. In other words, $f(i, i) = \mathbf{P}(\sigma_i < \infty)$, where $\sigma_i = \inf(n \geq 1 : X_n = s_i)$.

A state, s_i , is a **recurrent state** if $f(i, i) = 1$. It is called transient if $f(i, i) < 1$. Thus, s_i is a transient state if given that we start in state s_i , there is a non-zero probability that we will never return back to s_i .

Lemma 1.1. [6]

(a) The state is recurrent if and only if $\sum_{n=1}^{\infty} P^n(i, i) = \infty$.

(b) If state is transient then $\sum_{n=1}^{\infty} P^n(i, j) < \infty$.

Example 1.4. Assume the states $(1, 2, 3, 4)$ and transition matrix:

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

in this example every state is recurrent.

Definition 1.9. [6] A state, s_i , is a **positive recurrent** if

$$\left(\sum_{n=1}^{\infty} n f^n(i, i) \right)^{-1} > 0$$

a recurrent state is called a positive recurrent if the expected time between the visit of the state is finite, and a null recurrent otherwise.

The following is an example of a Markov chain with states that are recurrent null.

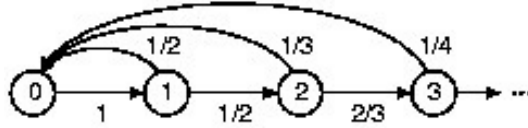


Figure 1.3: Null Recurrent

1.4.1 Irreducible and Aperiodic Markov Chains

We will discuss two such conditions on Markov chains: irreducibility and aperiodicity. These conditions are of the central importance in Markov theory, and they play a key role in the study of stationary distributions.

Suppose we have a Markov chain with state space S , and transition matrix \mathbf{P} . Two states, s_i and s_j , communicate writing $s_i \rightarrow s_j$, if there exist finite m, n such that $\mathbf{P}^m(i, j) > 0$ and $\mathbf{P}^n(j, i) > 0$.

Definition 1.10. [13] A Markov chain with state space S and a transition matrix \mathbf{P} is connected or irreducible if all states $s_i, s_j \in S$ communicate, that is

there exist a number $n \leq |S|$ such that $P^n(i, j) > 0$.

In an irreducible Markov chain there is a positive probability of going from every state to any other state in a finite number of steps, that is every state s_j is eventually reachable from any start state s_i ,

$$[\mathbf{P}(X_n = s_j \mid X_0 = s_i) > 0, \text{ for some } n \geq 0].$$

All states commute, as one can always go from any state to any other state although it may take more than one step.

Example 1.5. Consider the chain on states $(1, 2, 3)$ and

$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

As $1 \leftrightarrow 2$ and $2 \leftrightarrow 3$, this is an irreducible chain.

Example 1.6. A reducible Markov Chain

Consider a Markov chain with state space $\{(1, 2, 3, 4)\}$ and transition matrix:

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \\ 0 & 0 & 0.8 & 0.2 \end{pmatrix}$$

We see that if the chain starts in state 1 or 2, then it is restricted to state 1 or 2 forever. Similarly, if it starts in state 3 or 4, then it can never leave the subset $\{3, 4\}$ of the state space. Hence, the chain is reducible.

Note that: In example (1.15) every state is accessible from any other state, so this chain is irreducible.

Definition 1.11. A state s_i is said to have period d if $P^n(i, i) = 0$ whenever n is not divisible by d . The period $d(s_i)$ of a state $s_i \in S$ is defined as

$$d(s_i) := \gcd\{n \geq 1 : P^n(i, i) > 0, \forall i\}.$$

That is

$$\gcd\{n : \mathbf{P}(X_n = s_i \mid X_0 = s_i) > 0\}.$$

i.e. a return to state s_i after n transitions has positive probability only if n is a multiple of d . If $d(s_i) = 1$, then the state s_i is aperiodic, that is a Markov chain is said to be **aperiodic** if all its states have period 1, otherwise the chain is said to be periodic.

Example 1.7. Assume that there are only two kinds of weather: rain and sunshine, and the weather forms a Markov chain with state space $S = \{r, s\}$ and transition matrix:

$$\mathbf{P} = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

The Markov chain in this example is aperiodic since $P^n(i, i) > 0$ for all n and all states s_i .

Example 1.8. Consider the Markov chain with state space $\{0, 1, 2, 3, 4\}$ and transition matrix:

$$\mathbf{P} = \begin{pmatrix} 0 & \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}$$

The state 4 is aperiodic since $P(4, 4) > 0$, and since states 3, 4 are communicate, then state 3 is aperiodic too. States 0, 1, 2 have period 3.

Definition 1.12. A recurrent state, s_i , is a periodic state if there exist an integer $c > 1$, such that $p^{(r)}(i, i) = 0$ for all values for r other than $c, 2c, 3c, \dots$

Sometimes the terms indecomposable, a cyclic, and persistent are used as synonyms for "irreducible", "aperiodic", and "recurrent" respectively. One reason for the usefulness of aperiodicity is the following theorem.

Theorem 1.2. Suppose that we have an aperiodic Markov chain $\{X_0, X_1, \dots\}$ with state space $S = \{s_1, \dots, s_m\}$ and transition matrix \mathbf{P} . Then there exist an $N < \infty$ such that:

$$P^n(i, i) > 0$$

, for all $i \in \{1, \dots, m\}$ and all $n \geq N$.

By combining aperiodicity and irreducibility, we get the following important corollary, which will be used to prove the so-called Markov chain convergence theorem.

Corollary 1.1. *Now suppose our Markov chain is aperiodic and irreducible. Then there exist an $M < \infty$ such that*

$$P^n(i, j) > 0$$

for all $i, j \in \{1, \dots, m\}$ and all $n \geq M$.

In the context of MCMC a question of particular interest is the question of the long-term behavior of a Markov chain. Given certain conditions, does the distribution of the chain converge to a well defined and unique limit?. The concept of irreducibility and aperiodicity will provide an answer.

1.5 Stationary Distributions

We consider one of the central issues in Markov theory: asymptotic for the long-term behavior of Markov chains.

Example 1.9. Suppose the state space are (Rain, Sunny, Cloudy) and weather follows a Markov chain, that is, the probability of tomorrow's weather simply depends on today's weather, and not any other previous days.

Suppose the probability transition given today is rainy are

$$\mathbf{P}(\text{Rain tomorrow} \mid \text{Rain today}) = \frac{1}{2}$$

$$\mathbf{P}(\text{Sunny tomorrow} \mid \text{Rain today}) = \frac{1}{4}$$

$$\mathbf{P}(\text{Cloudy tomorrow} \mid \text{Rain today}) = \frac{1}{4}$$

The first row of the transition probability matrix thus become $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.

Suppose the rest of the transition matrix is given by:

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

This Markov chain is irreducible, as all states communicate with each other.

Suppose today is cloudy. What is the expected weather two days from now
? seven days?

Here $\mu^{(0)} = (0 \ 0 \ 1)$. Hence

$$\begin{aligned}\mu^{(2)} &= \mu^{(0)} P^2 = (001) \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{7}{16} & \frac{3}{16} & \frac{3}{8} \\ \frac{3}{8} & \frac{1}{4} & \frac{3}{8} \\ \frac{3}{8} & \frac{3}{16} & \frac{7}{16} \end{pmatrix} \\ &= \begin{pmatrix} \frac{3}{8} & \frac{3}{16} & \frac{7}{16} \end{pmatrix} \\ &= \begin{pmatrix} 0.375 & 0.1875 & 0.4375 \end{pmatrix}.\end{aligned}$$

and

$$\begin{aligned}\mu^{(7)} &= \mu^{(0)} P^7 = \begin{pmatrix} \frac{819}{2048} & \frac{819}{4096} & \frac{1639}{4096} \end{pmatrix} \\ &= \begin{pmatrix} 0.3999 & 0.1999 & 0.4001 \end{pmatrix} \\ &= \begin{pmatrix} 0.4 & 0.2 & 0.4 \end{pmatrix}.\end{aligned}$$

Conversely, suppose today is rainy, so that $\mu^{(0)} = (1 \ 0 \ 0)$.

The expected weather becomes

$$\begin{aligned}\mu^{(2)} &= \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{7}{16} & \frac{3}{16} & \frac{3}{8} \\ \frac{3}{8} & \frac{1}{4} & \frac{3}{8} \\ \frac{3}{8} & \frac{3}{16} & \frac{7}{16} \end{pmatrix} \\ &= \begin{pmatrix} \frac{7}{16} & \frac{3}{16} & \frac{3}{8} \end{pmatrix} \\ &= \begin{pmatrix} 0.4375 & 0.1875 & 0.375 \end{pmatrix}.\end{aligned}$$

and $\mu^{(7)} = (0.4 \ 0.2 \ 0.4)$

After a sufficient amount of time, the expected weather is independent of the starting value. In other words, the chain has reached a stationary distribution, where the probability values are independent of the actual starting value. The previous example illustrates, a Markov chain may reach a stationary distribution

π , where the vector of probabilities of being in any particular given state is independent of the initial condition.

We have this definition:

Definition 1.13. [13] Let $\{X_0, X_1, \dots\}$ be a Markov chain with state space $S = \{s_1, \dots, s_m\}$ and transition matrix \mathbf{P} .

A row vector $\pi = (\pi(1), \dots, \pi(m))$ is said to be a stationary distribution for the Markov chain, if it satisfies:

1. $\pi(i) \geq 0$ for all i and $\sum_{i=1}^m \pi(i) = 1$.
2. $\pi P = \pi$, that is $\sum_{i=1}^m \pi(i) P(i, j) = \pi(j)$ for all j .

Property (1) means that π should describe a probability distribution on S , and property (2) implies that if the initial distribution $\mu^{(0)}$ equal π , then the distribution $\mu^{(1)}$ of the chain at time 1 satisfies $\mu^{(1)} = \mu^{(0)} P = \pi P = \pi$, and by iterating we see that $\mu^{(n)} = \pi$ for every n .

Definition 1.14. A Markov chain is ergodic if it is both irreducible and aperiodic. i.e. the following conditions are equivalent to ergodicity

* Irreducible: $\forall i, j \in \{1, \dots, m\}, \exists n$ such that $P^n(i, j) > 0$.

* Aperiodic: $\forall i \in \{1, \dots, m\}, \gcd\{n : P^n(i, i) > 0\} = 1$.

Definition 1.15. [9] A Markov chain $\{X_0, X_1, X_2, \dots\}$ is called ergodic if the limit

$$\pi(j) = \lim_{n \rightarrow \infty} p^{(n)}(i, j)$$

- (1) exists for all $j \in \{1, \dots, m\}$.
- (2) is positive and does not depend on i .
- (3) $\pi = (\pi(1), \dots, \pi(m))$ is a probability distribution on S .

Ergodic Markov chain are useful for convergence theorem:

1.6 Convergence Theorems

Using the above concepts, we can formulate important convergence theorems.

Theorem 1.3. [12] *(Existence of stationary distribution)*

For any irreducible and aperiodic Markov chain, there exists at least one stationary distribution.

To achieve our final goal-a Markov chain convergence theorem-we introduce a metric on probability distributions.

Definition 1.16. Let $\mu = (\mu_1, \dots, \mu_m)$ and $\nu = (\nu_1, \dots, \nu_m)$ be two probability distribution on the state space $S = \{s_1, \dots, s_m\}$. The total variation distance between μ and ν is defined as

Equation 1.1. $d_{TV}(\mu, \nu) = \frac{1}{2} \sum_{i=1}^m |\mu_i - \nu_i|.$

A sequence of probability distribution $\pi^{(i)}$ converges in total variation to a distribution π if

$$\lim d_{TV}(\pi^{(i)}, \pi) = 0.$$

Shorthand we write $\pi^{(i)} \xrightarrow{TV} \pi$

The constant $\frac{1}{2}$ in (1.1) is designed to make the total variation distance d_{TV} take values between 0 and 1.

We are now ready to state the main result about convergence to stationarity.

Theorem 1.4. *(Markov chain convergence theorem) Consider an irreducible aperiodic Markov chain $\{X_0, X_1, \dots\}$. If we denote the chains distribution after the n^{th} transition by $\mu^{(n)}$, we have for any initial distribution $\mu^{(0)}$ and a stationary distribution π :*

$$\mu^{(n)} \xrightarrow{TV} \pi$$

In words: If we run the Markov chain for a long time, its distribution will be very close to the stationary distribution π . This is often referred to as the Markov chain approaching equilibrium as $n \rightarrow \infty$.

Theorem 1.5. [12] *(Uniqueness of the stationary distribution) Any irreducible and aperiodic Markov chain has exactly one stationary distribution.*

Remark 1.1. Even if a Markov chain has stationary distribution π , it may still fail to converge to stationarity.

Example 1.10. A Markov chain on a state space $S = \{1, 2, 3\}$, and transition matrix:

$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

with $\pi(1) = \pi(2) = \pi(3) = \frac{1}{3}$. Then π is a stationary distribution. However, let $X_0 = 1$, then $X_n \in \{1, 2\}$ for all n , this implies that $\mathbf{P}(X_n = 3) = 0$ for all n , so $\mathbf{P}(X_n = 3) \not\rightarrow \pi(3)$; and the distribution of X_n does not converge to π

Remark 1.2. The Markov chain in the above example is reducible, since the chain can never get from state 1 to state 3, in any number of steps, so the stationary distribution is not unique, and the distribution of X_n converges to a different stationary distribution defined by $\pi = (\frac{1}{2}, \frac{1}{2}, 0)$.

1.6.1 Detailed Balance and Time Reversal

We are interested in constructing Markov chains for which the distribution we wish to sample from, given by π , is invariant (stationary), so we need to find transition matrices \mathbf{P} that satisfies $\pi = \pi \mathbf{P}$.

That is if $P(i, j)$ denotes the probability of a transition from $s_i \in S$ to $s_j \in S$ under \mathbf{P} , we require that $\sum_i \pi(i)P(i, j) = \pi(j)$ for all j . We introduced a special class of Markov chains known as the reversible ones.

Definition 1.17. A probability distribution π on the state space $S = \{s_1, \dots, s_m\}$ is reversible for the Markov chain $\{X_0, X_1, \dots\}$ with transition matrix \mathbf{P} if for all $i, j \in \{1, \dots, m\}$ we have

$$\pi(i)P(i, j) = \pi(j)P(j, i)$$

which is also known as detailed balance equation.

The next simple theorem will help us with constructing Markov Chain Monte Carlo MCMC algorithms that (approximately) sample from a given distribution π .

Example 1.11. In the random walk Example , if we denote the number of neighbours of a vertex v_i by d_i , then the elements of the transition matrix are given by;

$$\textbf{Equation 1.2. } P(i, j) = \begin{cases} \frac{1}{d_i} = \frac{1}{2} & \text{if } v_i \text{ and } v_j \text{ are neighbours} \\ 0 & \text{otherwise} \end{cases}$$

which is a reversible Markov chain, with reversible distribution π given by

$$\pi = \left(\frac{d_1}{d}, \dots, \frac{d_4}{d}\right) = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$$

where $d = \sum_{i=1}^4 d_i$ since it satisfies detailed balance equation. To see that, we calculate

$$\textbf{Equation 1.3. } \pi(i)P(i, j) = \begin{cases} \frac{d_i}{d} \frac{1}{d_i} = \frac{1}{d} = \frac{1}{8} = \frac{d_j}{d} \frac{1}{d_j} = \pi(j)P(j, i) & \text{if } v_i \text{ and } v_j \text{ are neighbours} \\ 0 = \pi(j)P(j, i) & \text{otherwise} \end{cases}$$

It is possible for a distribution to be stationary without detailed balance holding.

Example 1.12. (A non reversible Markov chain).

The uniform distribution $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ on the state space $\{1, 2, 3\}$ is stationary (invariant) with respect to the homogeneous Markov chain with transition matrix:

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

but detailed balance does not hold, to see this: let $i = 1, j = 2$, we get

$$\pi(1)P(1, 2) = \frac{1}{3} \cdot 1 = \frac{1}{3} > 0 = \frac{1}{3} \cdot 0 = \pi(2)P(2, 1)$$

so that $\pi(1)P(1, 2) \neq \pi(2)P(2, 1)$ and reversibility fails.

Also, if we let $X_0 = 1$, then $X_n = 1$ whenever $n = 0, 3, 6, 9, \dots$ (a multiple of 3), thus $\mathbf{P}(X_n = 1) = 0$ or 1 , so $\mathbf{P}(X_n = 1) \not\rightarrow \pi(3)$, and there is again no convergence to π .

Theorem 1.6. (*Detailed Balance Test*) *If the probability distribution π is reversible for a Markov chain, then it is also a stationary distribution for the chain.*

Proof. Since π is reversible, we have $\forall i, j. \pi(i)P(i, j) = \pi(j)P(j, i)$. For fixed $i \in \{1, \dots, m\}$, we sum this equation with respect to $j \in \{1, \dots, m\}$ to get

$$\sum_j \pi(i)P(i, j) = \sum_j \pi(j)P(j, i)$$

. But the left hand side equals $\pi(i) \sum_j P(i, j) = \pi(i)$, since rows sum to one. Thus, $\forall i, j$.

$$\pi(i) = \sum_j \pi(j)P(j, i)$$

and this implies $\pi = \pi \mathbf{P}$, which makes π a stationary distribution. \square

If a Markov chain satisfies the detailed balance condition, then it is time reversible, i.e. one could not tell whether a sequence of samples is being acquired forward or backward in time. That is at stationarity, the probability of transition being from state s_i to state s_j is the same as the probability of it being from state s_j to state s_i .

Chapter 2

Random Walk

A **random walk** is a mathematical formalization of a path that consists of a succession of random steps. For example, the path traced by a molecule as it travels in a liquid or a gas, the search path of a foraging animal, the price of a fluctuating stock and the financial status of a gambler can all be modeled as random walks, although they may not be truly random in reality. The term random walk was first introduced by Karl Pearson in 1905. Random walks have been used in many fields: ecology, economics, psychology, computer science, physics, chemistry, and biology. Random walks explain the observed behaviors of many processes in these fields, and thus serve as a fundamental model for the recorded stochastic activity.

Various different types of random walks are of interest. Often, random walks are assumed to be Markov chains or Markov processes, but other, more complicated walks are also of interest. Some random walks are on graphs, others on the line, in the plane, in higher dimensions, or even curved surfaces, while some random walks are on groups. Random walks also vary with regard to the time parameter. Often, the walk is in discrete time, and indexed by the natural numbers, as in X_0, X_1, X_2, \dots . However, some walks take their steps at random times, and in that case the position is defined for the continuum of times $t \geq 0$. Random walks are related to the diffusion models and are a fundamental topic in discussions of Markov processes. Several properties of random walks, including dispersal distributions, first-passage times and encounter rates, have been extensively

studied.

The words, Random Walk, in their simplest incarnation, refer to this situation: The proverbial drunk is clinging to the lamppost. He decides to start walking. The road runs east and west. In his inebriated state he is as likely to take a step east (forward) as west (backward). (Probability 50% in either of the two directions.) From each new position he is again as likely to go forward as backward. Each of his steps are of the same length but of random direction east or west.

After having taken n steps (n can be one step, ten steps, fifty-three..) the walker is to be found standing at some position from which he makes a step to one of his neighboring positions. One can plot position (or displacement) against the number of steps taken for any particular random walk.

2.1 Random Walk as Markov Chain

Definition 2.1. [14] Let $(X_k)_{k=1}^{\infty}$ be a sequence of independent, identically distributed discrete random variables. For each positive integer n , we let S_n denote the sum $X_1 + X_2 + \cdots + X_n$.

The sequence $(S_n)_{n=1}^{\infty}$ is called a **random walk**. If the common range of the X_k 's is R^m , then we say that S_n is a random walk in R^m .

We view the sequence of X_k 's as being the outcomes of independent experiments. Since the X_k 's are independent, the probability of any particular (finite) sequence of outcomes can be obtained by multiplying the probabilities that each X_k takes on the specified value in the sequence. Of course, these individual probabilities are given by the common distribution of the X_k 's. We will typically be interested in finding probabilities for events involving the related sequence of S_n 's. Such events can be described in terms of the X_k 's, so their probabilities can be calculated using the above idea. There are several ways to visualize a random walk. One can imagine that a particle is placed at the origin in R^m at time $n = 0$. The sum S_n represents the position of the particle at the end of n seconds. Thus, in the time interval $[n - 1; n]$, the particle moves (or jumps) from position $[S_{n-1} - S_n]$. The vector representing this motion is

just $[S_n - S_n - 1]$, which equals X_n . This means that in a random walk, the jumps are independent and identically distributed. If $m = 1$, for example, then one can imagine a particle on the real line that starts at the origin, and at the end of each second, jumps one unit to the right or the left, with probabilities given by the distribution of the X_k 's. If $m = 2$, one can visualize the process as taking place in a city in which the streets form square city blocks. A person starts at one corner (i.e., at an intersection of two streets) and goes in one of the four possible directions according to the distribution of the X_k 's. If $m = 3$, one might imagine being in a jungle gym, where one is free to move in any one of six directions (left, right, forward, backward, up, and down). Once again, the probabilities of these movements are given by the distribution of the X_k 's. Another model of a random walk (used mostly in the case where the range is (R^1)) is a game, involving two people, which consists of a sequence of independent, identically distributed moves. The sum S_n represents the score of the first person, say, after n moves, with the assumption that the score of the second person is S_n . For example, two people might be flipping coins, with a match or non-match representing $+1$ or -1 , respectively, for the first player. Or, perhaps one coin is being flipped, with a head or tail representing $+1$ or -1 , respectively, for the first player.

Definition 2.2. [2] $S_0 = 0$ and $S_n = X_1 + \dots + X_n$ for $n \in N$ is called the position of the random walk at time n .

2.2 Transition matrix

A random walk (or Markov chain), is most conveniently represented by its transition matrix \mathbf{P} .

\mathbf{P} is a square matrix denoting the probability of transitioning from any vertex in the graph to any other vertex. Formally, $P_{ij} = P[\text{going from } i \text{ to } j, \text{ given that we are at } i]$. Thus for a random walk, $P_{ij} = \frac{1}{d_i}$, if $(i, j) \in E$, and 0 otherwise (where d_i is the degree of i). Thus for the example graph given below the transition matrix would be as:

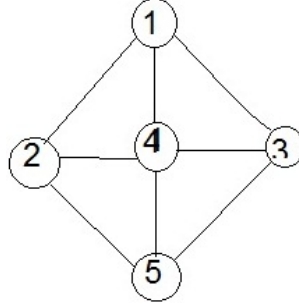


Figure 2.1: Random Walk

Assume that the states are $(1, 2, 3, 4, 5)$ and transition matrix is:

$$\mathbf{P} = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

2.3 Stationary distributions

Thus if we have a distribution π over the nodes, we can obtain the distribution after one step by computing $\pi = \pi \cdot \mathbf{P}$. Using this definition, we can define a stationary distribution π_j as a distribution with the property that $\pi_j \cdot \mathbf{P} = \pi_j$. Stationary distributions are not always unique, but under certain conditions, they are.

2.4 The One-Dimensional Random Walk

Definition 2.3. The simple random walk[16]

A sequence $\{X_n\}_{n \in \mathbb{N}_0}$ of random variables is called a **simple random walk** (with parameter $p \in (0, 1)$) if:

- $X_0 = 0$.

- $X_{n+1} - X_n$ is independent of (X_0, X_1, \dots, X_n) for all $(n \in N)$,
- usual, $q = 1 - p$.
- If $p = \frac{1}{2}$, the random walk is called symmetric.

The one-dimensional random walk is constructed as follows:

- (1) You walk along a line, each pace being the same length.
- (2) Before each step, you flip a coin.
- (3) If its heads, you take one step forward.
- (4) If its tails, you take one step back.

The coin is unbiased, so the chances of heads or tails are equal.

The problem is to find the probability of landing at a given spot after a given number of steps, and, in particular, to find how far away you are on average from where you started.

An elementary example of a random walk is the random walk on the integer number line, which starts at 0 and at each step moves $+1$ or -1 with equal probability.

This walk can be illustrated as follows. A marker is placed at zero on the number line and a fair coin is flipped. If it lands on heads, the marker is moved one unit to the right. If it lands on tails, the marker is moved one unit to the left. After five flips, the marker could now be on $(1, 1, 3, 3, 5, \text{ or } 5)$. With five flips, three heads and two tails, in any order, will land on 1. There are 10 ways of landing on 1 (by flipping three heads and two tails), 10 ways of landing on -1 (by flipping three tails and two heads), 5 ways of landing on 3 (by flipping four heads and one tail), 5 ways of landing on -3 (by flipping four tails and one head), 1 way of landing on 5 (by flipping five heads), and 1 way of landing on -5 (by flipping five tails).

To define this walk formally, take independent random variables

(X_1, X_2, X_3, \dots) , where each variable is either 1 or -1 , let time be discrete, i.e. $(t = 0, 1, \dots)$ Consider the following stochastic process S_n :

- (i) $S_0 = 0$.

(ii) at each time step it moves to 1 with equal probability $\frac{1}{2}$.

In other words, at each time step we flip a fair coin. If the outcome is heads, we move one unit to the right. If the outcome is tails, we move one unit to the left. Alternatively, we can think of the random walk as a sum of independent random variables:

$$[S_n = \sum_{j=1}^n X_j, \text{ where } X_j \in \{1, -1\} \text{ with } [P(X_j = 1) = \frac{1}{2}] .$$

The series S_n is called the simple random walk. This series (the sum of the sequence of -1 s and 1 s) gives the distance walked, if each part of the walk is of length one. The expectation of is zero. That is, the mean of all coin flips approaches zero as the number of flips increases.

As a Markov chain A one-dimensional random walk can also be looked at as a Markov chain whose state space is given by the integers for some number q satisfying $(0 < q < 1)$, the transition probabilities (the probability $P_{i,j}$ of moving from state i to state j) are given by $[P_{i,i+1} = q = 1 - P_{i,i-1}]$.

Example 2.1. (Gambler's Ruin) As an example of a more applied problem, consider the following.

Suppose a game is being played in which a gambler flips a coin, and gains 1 dollar if it lands on heads, and loses 1 dollar if it lands on tails. Furthermore, suppose that the gambler wins the game if he reaches n dollars, and loses the game if he reaches 0 dollars. This game is represented by a random walk on non - negative integer, with the fortune of the gambler at time t given by S_t . we know the game must terminate, since either point n or 0 is eventually reached.

Gamblers: A, B have a total of N dollars.

Game: Toss Coin

If

- (**H**) A receives 1 from B.
- (**T**) B receives 1 from A.
- $P(H) = p, P(T) = q = 1 - p$.

X_n = Amount of money A has after n plays.

$$P(X_{n+1} = X_n + 1 | X_n) = p.$$

$$P(X_{n+1} = X_n - 1 | X_n) = q.$$

Game ends if $X_n = 0$ or $X_n = N$.

State space = $\{0, 1, 2, \dots, N\}$

The Transition Matrix is

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & \dots & N-2 & N-1 & N \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ . \\ . \\ N-1 \\ N \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ q & 0 & p & 0 & \dots & 0 & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & q & 0 & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Transition Diagram for Gamblers Ruin

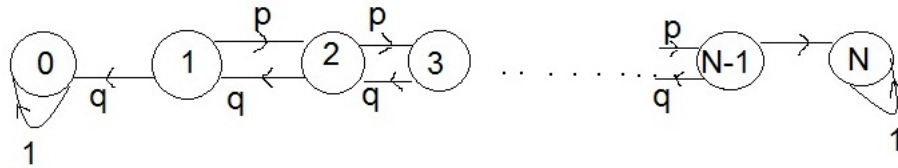


Figure 2.2: Gamblers Ruin

2.5 Higher Dimensions Random Walk

Imagine now a drunkard walking randomly in an idealized city. The city is effectively infinite and arranged in a square grid, and at every intersection, the drunkard chooses one of the four possible routes (including the one he came from) with equal probability. Formally, this is a random walk on the set of all points in the plane with integer coordinates. Will the drunkard ever get back

to his home from the bar? This is the 2-dimensional equivalent of the level crossing problem discussed above. It turns out that he almost surely will in a 2-dimensional random walk, but for 3 dimensions or higher, the probability of returning to the origin decreases as the number of dimensions increases. In 3 dimensions, the probability decreases to roughly 34%.

The trajectory of a random walk is the collection of sites it visited, considered as a set with disregard to when the walk arrived at the point. In one dimension, the trajectory is simply all points between the minimum height the walk achieved and the maximum.

Simple symmetric random walk on Z^2 (Two dimensions random walk)

The simple random walk on Z^2 has diagram

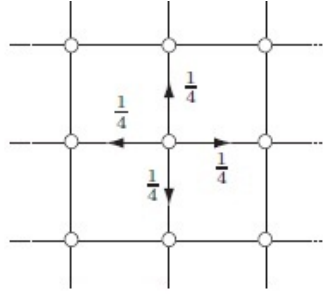


Figure 2.3: Two Dimensions Random Walk

and **transition probabilities**

$$p_{ij} = \begin{cases} \frac{1}{4}, & \text{if } |i - j| = 1 \\ 0, & \text{otherwise.} \end{cases}$$

Example 2.2. Suppose that a mouse is moving inside the maze shown in Figure (2.4), from one cell to another, in search of food. When at a cell, the mouse will move to one of the adjoining cells randomly. For $n \geq 0$, let X_n be the cell number the mouse will visit after having changed cells n times. Then $\{X_n : n = 0, 1, \dots\}$ is a Markov chain with state space $\{1, 2, \dots, 9\}$ and transition probability matrix:

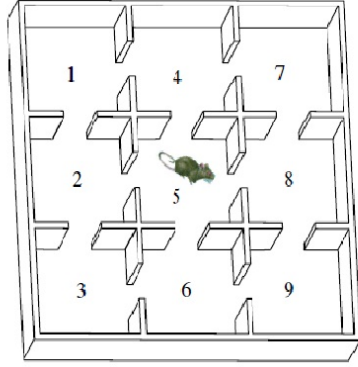


Figure 2.4: The moving mouse

$$\mathbf{P} = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

Chapter 3

Markov Chain Monte Carlo

3.1 Markov Chain Monte Carlo Methods

It is often difficult to sample directly from the distribution π ; the reason is that they are defined as $\pi(x) = \frac{f(x)}{z}$, where $f(x)$ is an easily computable density weight function, and z is unknown normalizing constant that is often very difficult to compute.

In such circumstance we can use algorithms (Markov Chain Monte Carlo MCMC simulation)

The problem to be solved:

The aims of Monte Carlo methods are to solve one or both the following problems:

- (1) To generate samples $\{X_n\}_{n \geq 1}$ from a given probability distribution.
- (2) To estimate expectations of functions under this distribution.

Consider the general problem of trying to calculate characteristics of a complicated multivariate probability distribution $f(X)$ on X such that X is a vector of dimension m .

For example, suppose we want to calculate $\iint x.f(x,y)dxdy$ where

$$f(x,y) \propto (1+x^2)^{-1}y^{-n} \exp(\frac{-1}{2y^2} \sum_{i=1}^n (\lambda_i - x)^2 - y),$$

$\lambda_1, \dots, \lambda_n$ are fixed constants, which is analytically difficult to compute.

Markov Chain Monte Carlo MCMC methods are class of algorithms for sampling from probability distributions based on constructing a Markov chain having state space S and has the desired (target) distribution as its stationary distribution π . The Markov chain is designed to be ergodic, i.e., the probability distribution over S converges asymptotically to π , regardless of the initial state.

Now suppose we want to draw a sample from a density distribution π . The main idea behind MCMC is to start from an arbitrary state in S , run a Markov chain, whose stationary distribution equals to π , for a long time, say N iterations, and note what state the chain is in after these N iterations, and then take the result "output" as an approximate sample from π .

The ergodicity means that, by taking N large enough, we can ensure that the distribution of the output state is arbitrarily close to the desired distribution π . So if the Markov chain has reached equilibrium, the output is distributed according to the stationary distribution, as desired.

MCMC techniques provide a powerful tool for statistical analysis for data. In this chapter we present a common techniques which have ignited MCMC: Metropolis-Hastings algorithms.

3.2 Metropolis Algorithm

This algorithm has been applied extensively to problems in statistical physics. The Metropolis algorithm, dating back to 1953, uses a symmetric candidate transition (a proposal transition matrix $Q=q(i,j)$) on S for which $q(i,j) = q(j,i)$. Suppose our goal is to draw samples from some distribution $\pi(x)$ where $\pi(x) = \frac{f(x)}{z}$, such that the normalizing constant z may not be known, and very difficult to compute.

Define a Markov chain with the following process:

- (1) Start with any initial value x_0 satisfying $f(x_0) > 0$.
- (2) Using current value $x = i$, sample a candidate point $x^* = j$ from a symmetric proposal distribution $q(i,j)$ which is the probability of returning a value of j given a previous value of i .

- (3) Calculate the ratio r of the density at the candidate $x^* = j$ and current point $x = i$ (this is easy since the normalization constant z cancels)

$$r = \frac{\pi(x^*)}{\pi(x)} = \frac{\frac{f(x^*)}{z}}{\frac{f(x)}{z}}.$$

- (4) If the above calculation gives ($r > 1$), accept the candidate point, and return to step 2 .
- (5) With probability $\min(r, 1)$ transition to $x^* = j$, otherwise stay in the same state x .

This generates a Markov chain $\{x_0, x_1, \dots, x_k, \dots\}$ as the transition probabilities from x_n to x_{n+1} depends only on x_n and not on $\{x_0, \dots, x_{n-1}\}$.

Following a sufficient burn-in period (of k steps), the chain approaches its stationary distribution and samples from $\{x_{k+1}, \dots, x_{k+m}\}$ are samples from $\pi(x)$.

Example 3.1. Consider the scalar inverse χ^2 distribution,

$$f(x) = C.x^{-\frac{n}{2}} \exp\left(\frac{-a}{2x}\right)$$

and suppose we wish to simulate draws from the distribution with ($n = 5$) degrees of freedom, and scaling factor $a = 4$ using the Metropolis algorithm.

Suppose that a uniform distribution on $(0, 100)$ is the candidate distribution.

Take $x = 1$ as starting (initial) value, and suppose the uniform returns a candidate value of $x^* = 39.82$. Here

$$\alpha = \min(r, 1) = \min\left(\frac{f(x^*)}{f(x)}, 1\right) = \min\left(\frac{(39.82)^{-2.5} \exp\left(\frac{-2}{39.82}\right)}{(1)^{-2.5} \exp\left(\frac{-2}{1}\right)}, 1\right) = 0.0007$$

$\alpha = 0.0007 < 1$, implies x^* is accepted with probability 0.0007 .

Thus, we randomly draw U from a uniform $(0, 1)$ and accept x^* if $U \leq \alpha$. In this case, $u = 0.1885523$ then, the candidate is rejected.

3.3 Metropolis-Hastings Algorithm

Suppose we would like to sample from a potentially difficult distribution π that is in a sample space of high dimension. If we could find a Markov chain whose

unique stationary distribution is π , we could run this chain long enough and then take the result as an approximate sample from π . A very general way to construct such a Markov chain is the Metropolis Hastings algorithm. Since the symmetry requirement of the Metropolis proposal distribution can be hard to satisfy, Hastings in (1970) extended the Metropolis algorithm to a more general candidate transition which converges to $\pi(x)$ by using an arbitrary transition probability function $q(i, j) = p(i \rightarrow j)$ (i.e. a Markov transition kernel).

The algorithm proposes a new point (state) on the Markov chain which is either accepted or rejected.

- If the state is accepted, the Markov chain moves to the new state.
- If the state is rejected, the Markov chain remains in the same state.

By choosing the acceptance probability correctly, we create a Markov chain which has π as a stationary distribution.

We begin with a state space S and a probability distribution π on S . Then we choose a candidate distribution $q(i, j) : i, j \in \{1, \dots, m\}$ with $q(i, j) \geq 0$ and $\sum_j q(i, j) = 1$ for each i . Given that $X_n = i$, the algorithm works as follows to compute X_{n+1} .

- (1) Sample a candidate point "state" $Y_{n+1} = j$ from the candidate distribution $q(i, j) = p(Y_{n+1} = j \mid X_n = i)$ (Markov chain) which is not necessarily symmetric.
- (2) Compute the ratio $r = \frac{\pi(j)q(j, i)}{\pi(i)q(i, j)}$
- (3) With probability $\alpha(i, j) = \min(r, 1)$, let $X_{n+1} = Y_{n+1} = j$ (transition to X_{n+1} "accept the candidate"). Otherwise, let $X_{n+1} = X_n = i$ (stay in the same state X_n "reject the candidate"), where

$$\alpha(i, j) = \begin{cases} \min(r, 1) & \text{if } \pi(i)q(i, j) > 0 \\ 1 & \text{if } \pi(i)q(i, j) = 0 \end{cases} \quad (3.1)$$

In detail, this can be done by generating a random number, U , from the uniform distribution on $(0, 1)$, and then setting the next state as follows:

$$X_{n+1} = \begin{cases} j & \text{if } U < \alpha(i, j) \\ i & \text{otherwise} \end{cases} \quad (3.2)$$

(4) Discard initial "burn in" values.

The initial value is arbitrarily selected, which means Metropolis-Hastings typically has "burn in" period at the start of a simulation.

Now for reversible chains it is impossible to distinguish between forward and backward running of the chain. So, to construct the transition probabilities $p(i, j)$, we set

$$p(i, j) = q(i, j)\alpha(i, j), \quad (3.3)$$

where $q(i, j)$'s are the transition probabilities for another Markov chain fulfilling

$$q(i, j) > 0 \implies q(j, i) > 0 \quad \forall \quad i, j \in \{1, \dots, m\}. \quad (3.4)$$

Finally, to see why the Metropolis-Hastings algorithm works (generates a Markov chain whose equilibrium density is that candidate density $f(x)$), it is sufficient to show that the implied transition kernel $q(i, j)$ of any Metropolis-Hastings algorithm satisfies the detailed balance equation; i.e., we want to show that the Markov chain resulting from the Metropolis-Hastings algorithm is reversible with respect to π .

Theorem 3.1. [6] *The Metropolis-Hastings algorithm produces a Markov chain $\{X_0, X_1, \dots\}$ which is reversible with respect to π*

Proof. We must show that $\pi(i)P(i, j) = \pi(j)P(j, i)$.

Obviously this holds if $i = j$, so we will consider $i \neq j$, then

$$\begin{aligned} \pi(i)P(i, j) &= \pi(i)q(i, j)\alpha(i, j) \\ &= \pi(i)q(i, j) \min\left\{1, \frac{\pi(j)q(j, i)}{\pi(i)q(i, j)}\right\} \\ &= \min\{\pi(i)q(i, j), \pi(j)q(j, i)\} \end{aligned} \quad (3.5)$$

Similarly, we find that

$$\pi(j)P(j, i) = \min\{\pi(i)q(i, j), \pi(j)q(j, i)\}$$

This implies that

$$\pi(i)P(i, j) = \pi(j)P(j, i).$$

Therefore, the chain is reversible. \square

It follows that the Metropolis-Hastings algorithm converges to the target distribution π which is the stationary distribution by detailed balance.

3.4 The Random Walk Metropolis Algorithm

Suppose that we want to generate random variables according to an arbitrary probability density $\pi(x)$. The Metropolis algorithm produces a "random walk" of points $\{x_n\}$ whose asymptotic probability approaches after a large number of steps.

The random walk is defined by a "transition probability" $p(x_i \rightarrow x_j)$ for one value x_i to another x_j in order that the distribution of points converges to $\pi(x)$. It can be shown that it is sufficient (but not necessary) to satisfy the "detailed balance" condition

$$\pi(x_i)p(i, j) = \pi(x_j)p(j, i)$$

The random walk Metropolis is a special case of Metropolis algorithm, utilizing asymmetric candidate transition, that is $q(x, y) = q(y, x)$ (symmetric) in R^d assumes $q(x, y) = q(y - x)$.

Then we have:

$$\alpha(i, j) = \min[1, \frac{\pi(j)}{\pi(i)}]$$

This choice can be described by the following steps. Suppose that the "random walker" is at position x_n . To generate x_{n+1} we

- (1) Choose a trial position $x_t = x_n + \delta_n$, where the δ_n is a random number in the interval $[-\delta, +\delta]$.

- (2) Calculate $r = \frac{\pi(x_t)}{\pi(x_n)}$.
- (3) If $r \geq 1$ we accept the change and let $x_{n+1} = x_t$.
- (4) If $r < 1$, generate a random number U .
- (5) If $U \leq r$, accept the change and let $x_{n+1} = x_t$.
- (6) If the trial change is not accepted, the let $x_{n+1} = x_n$.

It is necessary to sample a number of points of the random walk before the asymptotic probability is attained. How do we choose the "step size" ?

If it is too large, only a small fraction of changes will be accepted and the sampling will be inefficient. If it is too small, a large number will be accepted, but it would take too long to sample over the whole interval of interest.

Example 3.2. [5] Write the Random walk Metropolis algorithm for obtaining samples from the posterior distribution $\pi(x) = N(5, (1.5)^2)$. The posterior distribution is given by:

$$\begin{aligned}\pi(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{1.5\sqrt{2\pi}} e^{-\frac{(x-5)^2}{2(1.5^2)}} \\ &\propto e^{-2\frac{(x-5)^2}{9}}\end{aligned}$$

We can select the proposal distribution $q(x, x^*) = N(x, 1)$, which is symmetric distribution. Then we can use Random walk Metropolis algorithm for obtaining samples.

The acceptance probability is given by:

$$\begin{aligned}\alpha(i, j) &= \min\left[1, \frac{\pi(x_j)}{\pi(x_i)}\right] \\ &= \min\left[1, \frac{e^{-2\frac{(x^*-5)^2}{9}}}{e^{-2\frac{(x-5)^2}{9}}}\right]\end{aligned}$$

The Metropolis algorithm proceeds as follows:

* Select starting point $x_0 = 0$, generate $x^* = 0.2893$ from $N(x_0 = 0, 1)$, and $u = 0.571$ from uniform $(0, 1)$.

Then the acceptance probability:

$$\begin{aligned}\alpha(i, j) &= \min\left[1, \frac{e^{-2\frac{(0.2893-5)^2}{9}}}{e^{-2\frac{(0-5)^2}{9}}}\right] \\ &= \min\{1, 1.3198\} \\ &= 1\end{aligned}$$

Since $u = 0.571 < \alpha = 1$, then the Metropolis algorithm accepts $x^* = 0.2893$.

Therefore $x_1 = 0.2893$ with probability $= 1$.

* Generate $x^* = 0.4462$ from $N(x_1 = 0.2893, 1)$, and $u = 0.9801$ from uniform $(0, 1)$. Then the acceptance probability:

$$\begin{aligned}\alpha(i, j) &= \min\left[1, \frac{e^{-2\frac{(0.4462-5)^2}{9}}}{e^{-2\frac{(0.2893-5)^2}{9}}}\right] \\ &= \min\{1, 1.1544\} \\ &= 1\end{aligned}$$

Since $u = 0.9801 \leq \alpha = 1$, then the Metropolis algorithm accepts the point $x^* = 0.4462$. Therefore $x_2 = 0.4462$ with probability $= 1$.

Generate $x^ = 1.8602$ from $N(x_2 = 0.4462, 1)$, and $u = 0.211$ from uniform $(0, 1)$. Then the acceptance probability:

$$\begin{aligned}\alpha &= \min\left[1, \frac{e^{-2\frac{(1.8602-5)^2}{9}}}{e^{-2\frac{(0.4462-5)^2}{9}}}\right] \\ &= \min\{1, 0.0743\} \\ &= 0.0743\end{aligned}$$

Since $u = 0.211 > \alpha = 0.0743$, then the Metropolis algorithm rejects the point $x^* = 1.8602$

Generate $x^ = 0.6989$ from $N(x_2 = 0.4462, 1)$, and $u = 0.003$ from uniform $(0, 1)$.

$$\begin{aligned}\alpha &= \min\left[1, \frac{e^{-2\frac{(0.6989-5)^2}{9}}}{e^{-2\frac{(0.4462-5)^2}{9}}}\right] \\ &= \min\{1, 1.2473\} \\ &= 1\end{aligned}$$

Since $u = 0.003 \leq \alpha = 1$, then the Metropolis algorithm accepts the point $x^* = 0.6989$

Therefore $x_3 = 0.6989$ with probability $\alpha = 1$.

Generate $x^ = 0.8729$ from $N(x_3 = 0.6989, 1)$, and $u = 0.157$ from uniform $(0, 1)$. Then the acceptance probability:

$$\begin{aligned}
\alpha &= \min\left[1, \frac{e^{-2\frac{(0.8729-5)^2}{9}}}{e^{-2\frac{(0.6989-5)^2}{9}}}\right] \\
&= \min\{1, 0.2061\} \\
&= 0.2061
\end{aligned}$$

Since $u = 0.157 \leq \alpha = 0.2061$, then the Metropolis algorithm accepts the point $x^* = 0.8729$.

Therefore $x_4 = 0.8729$ with probability $= 0.2061$.

Therefore $x_1 = 0.2893$, $x_2 = 0.4462$, $x_3 = 0.6989$, $x_4 = 0.8729$, \dots

Chapter 4

Simulation

In this chapter we explore some of the related derived results of the random walk Metropolis algorithm that have important practical implications. We also demonstrate the impact of the random walk Metropolis algorithm for some practical examples using the R programming language in simulation.

4.1 Markov Chain Simulation

We will use R program to simulate n values of Markov chains, we using simulation when Markov chains have three states.

Example 4.1. A frog hopping

The probabilities of a frog hopping, can be represented by a transition matrix:

$$P = \begin{matrix} & \begin{matrix} R_1 & R_2 & R_3 \end{matrix} \\ \begin{matrix} R_1 \\ R_2 \\ R_3 \end{matrix} & \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{5}{8} & \frac{1}{8} & \frac{2}{8} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix} \end{matrix}$$

Suppose we want simulate 100 values of a Markov Chain having transition matrix P , starting at $x_1 = 2$ (Rock 2).

A Markov chain simulation generates 100 values of above Markov chain having transition matrix P , starting at $x_1 = 2$.

```

[1] 2 1 3 2 1 3 1 3 1 2 2 1 2 3 2 1 2 1 3 1 3 1 3 1 2 1 3 1 3
[30] 1 2 1 3 2 1 3 2 1 2 1 3 2 3 2 2 1 2 3 2 1 3 1 2 1 3 1 3 2
[59] 2 1 3 1 3 1 2 1 3 1 2 1 3 2 3 2 3 1 2 2 3 2 1 2 2 1 3 2 3
[88] 1 3 1 3 1 3 1 2 3 2 3 1 3

```

Then if the frog is in rock 2, after 100 iteration the frog is in rock 3.

Full R code of this Example is in Appendix.

4.2 MCMC Simulation

4.2.1 Metropolis-Hasting Algorithm

Suppose that we have normal distribution, we use R program to plot the target sample of Metropolis-Hasting with normal distribution $N(\mu, \sigma)$, and plot the chain of sample x .

Example 4.2. Suppose we have normal distribution with candidate = $c(1.2, 1.2)$, after 10000 iteration. We plot the chain of x as Figure (4.1):

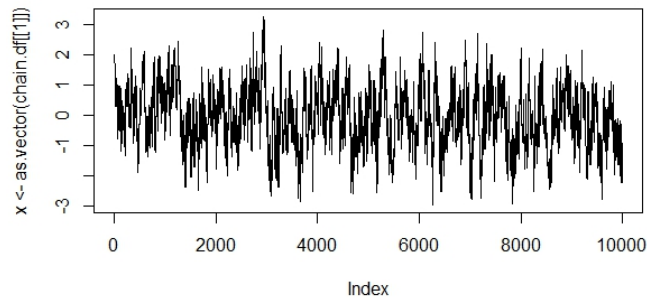


Figure 4.1: Chain of x

And we plot the target sample of two chains such that;
Chain (1) has candidate = $c(1.2, 1.2)$, $r=.9$ and after 10000 iterations we have

target sample as Figure (4.2).

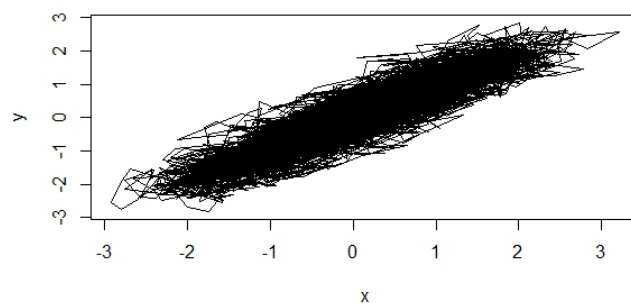


Figure 4.2: Target Sample with $r=.9$

Chain (2) has candidate = $c(1.2, 1.2)$, $r=.99$ and after 10000 iterations we have target sample as Figure (4.3).

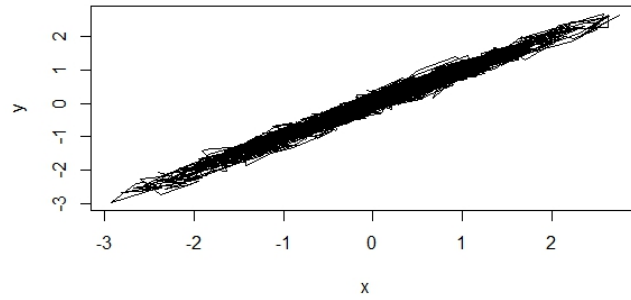


Figure 4.3: Target Sample with $r=.99$

If we take $r = .99$, we note that the target sample of random walk is more homogenous.

R code of Metropolis-Hasting simulation is in Appendix.

4.2.2 Random Walk Metropolis

Now we use symmetric normal Metropolis-Hasting to simulate chain of X_n with R program, such that if we selected any value of chain we have approached value of real data.

R code of symmetric normal Metropolis-Hasting simulation is in Appendix.

Example 4.3. Suppose we have normal distribution with candidate = $c(0, 1)$, We plot 3 chains such that;
Chain (1) has candidate = $c(0, 1)$, starting Value = -2 and after 10000 iterations, we have Figure (4.4).

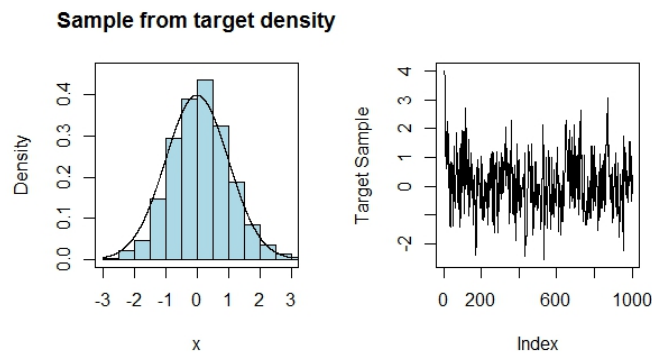


Figure 4.4: Chain 1 with starting value = -2

Chain (2) has candidate= $c(0,1)$, starting Value = 1 and after 10000 iterations, we have Figure (4.5).

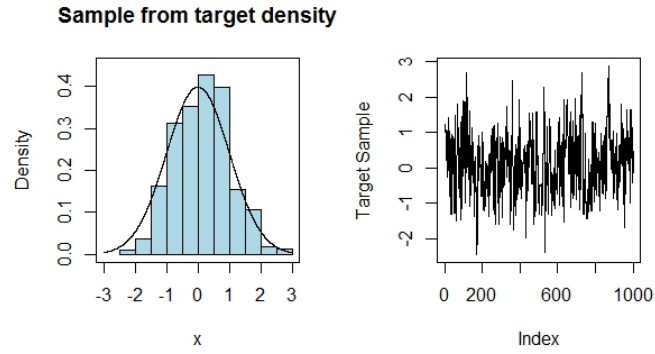


Figure 4.5: Chain 2 with start value =1

Chain (3) has candidate= $c(0,1)$, starting Value = 4 and after 10000 iterations, we have Figure (4.6).

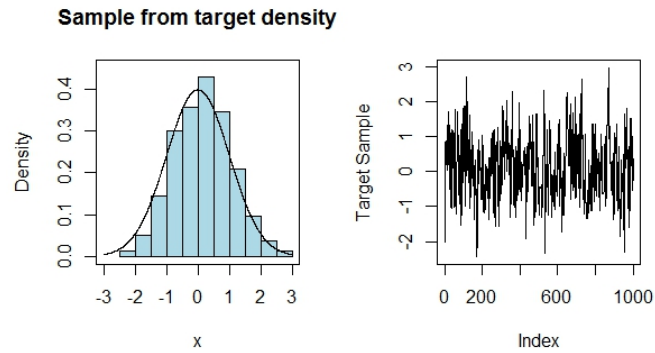


Figure 4.6: Chain 3 with start value =4

3 figures Explain us that the target sample was not affected by the starting point.

And we plot the rang of 3 chains together as Figure (4.7):

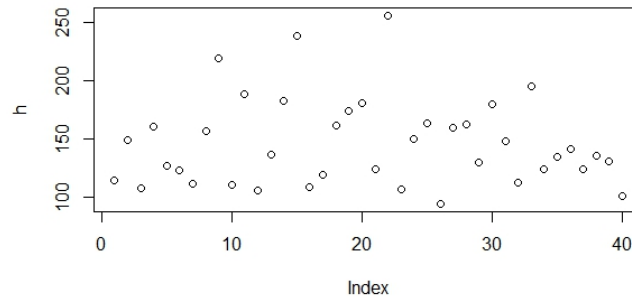


Figure 4.7: Rang of 3 chains

Example 4.4. (Males Heights)

It is know that the heights of adult males is normal distribution

$N(\mu = 70, \sigma = 4)$, we take a sample of 40 males and we measured their heights with inches.

The real height of males with inches as Table (4.1):

Num	HT	Num	HT
1	70.8	21	65.4
2	66.2	22	70.0
3	71.7	23	62.9
4	68.7	24	68.5
5	67.6	25	68.3
6	69.2	26	69.4
7	66.5	27	69.2
8	67.2	28	68.0
9	68.3	29	71.9
10	65.6	30	66.1
11	63.0	31	72.4
12	68.3	32	73.0
13	73.1	33	68.0
14	67.6	34	68.7
15	68.0	35	70.3
16	71.0	36	63.7
17	61.3	37	71.1
18	76.2	38	65.6
19	66.3	39	68.3
20	69.7	40	66.3

Table 4.1: Real males heights

We can represent to the data of males heights as histogram:

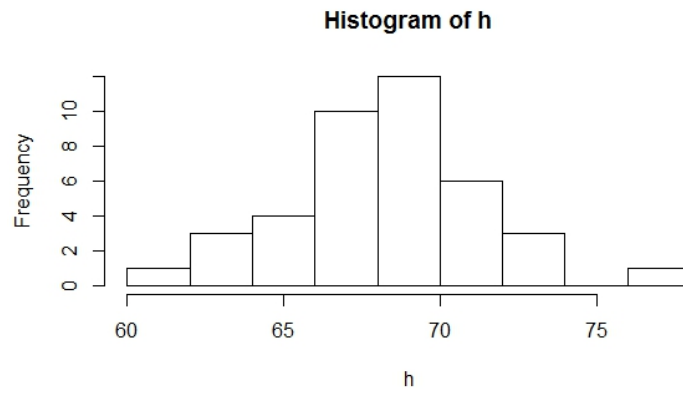


Figure 4.8: Histogram of males heights

And by using Metropolis-Hasting Algorithm to produce 3 chains, we have the following values:

Num	HT	Chain1	Chain2	Chain3
1	70.8	63.00000	70.00000	66.00000
2	66.2	62.67513	69.67513	65.67513
3	71.7	63.31462	70.31462	66.31462
4	68.7	62.36680	69.36680	65.36680
5	67.6	62.36680	69.36680	65.36680
6	69.2	65.40925	72.40925	68.40925
7	66.5	64.45876	71.45876	67.45876
8	67.2	64.14798	71.14798	67.14798
9	68.3	66.00744	73.00744	69.00744
10	65.6	66.97106	73.97106	69.97106
11	63.0	66.97106	72.85528	68.85528
12	68.3	67.24505	73.52147	69.52147
13	73.1	67.10238	72.98660	68.98660
14	67.6	66.64015	73.12927	69.12927
15	68.0	68.18419	72.52437	68.52437
16	71.0	69.60410	74.06841	70.06841
17	61.3	70.18147	75.48832	71.48832
18	76.2	70.30118	76.06569	72.06569
19	66.3	69.80165	76.18540	72.18540
20	69.7	70.37691	75.68587	71.68587

Table 4.2: 3 chains of males heights produced by Metropolis-Hasting

Num	HT	Chain1	Chain2	Chain3
21	65.4	69.77906	76.26113	72.26113
22	70.0	69.57523	75.66328	71.66328
23	62.9	70.15927	75.45946	71.45946
24	68.5	69.29630	76.04349	72.04349
25	68.3	70.01127	75.18052	71.18052
26	69.4	68.73284	75.89549	71.89549
27	69.2	68.73284	74.61706	70.61706
28	68.0	68.86689	72.25843	68.25843
29	71.9	68.76974	72.39247	68.39247
30	66.1	68.09229	72.29532	68.29532
31	72.4	67.42588	71.61787	67.61787
32	73.0	68.73284	70.95146	66.95146
33	68.0	67.97155	71.49714	67.49714
34	68.7	68.02020	71.54579	67.54579
35	70.3	67.95093	71.47651	67.47651
36	63.7	67.95093	70.88755	67.47651
37	71.1	67.96053	70.89715	67.48611
38	65.6	66.82643	69.76305	66.35201
39	68.3	67.51213	70.44875	67.03771
40	66.3	67.28118	70.21780	66.80677

Table 4.3: 3 chains of males heights produced by Metropolis-Hasting

Now we want to plot the target sample of males heights by using 3 chains. 3 chains have the same steps, but they are different with starting point. Chain (1) has candidate = $c(0, 1)$, starting value 63 and after 10000 iterations, we have Figure (4.9).

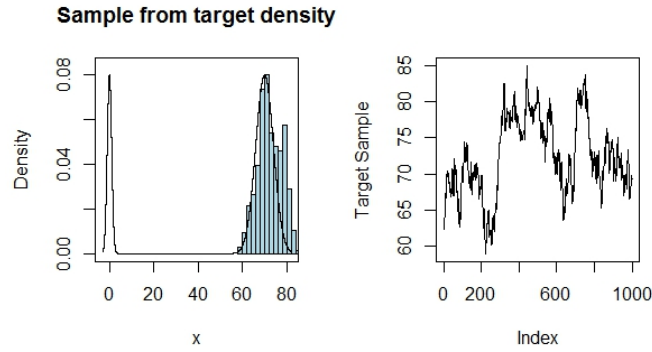


Figure 4.9: Target sample with starting point=63 inches

Chain (2) has candidate = $c(0, 1)$, starting value 70 and after 10000 iterations, we have Figure (4.10).

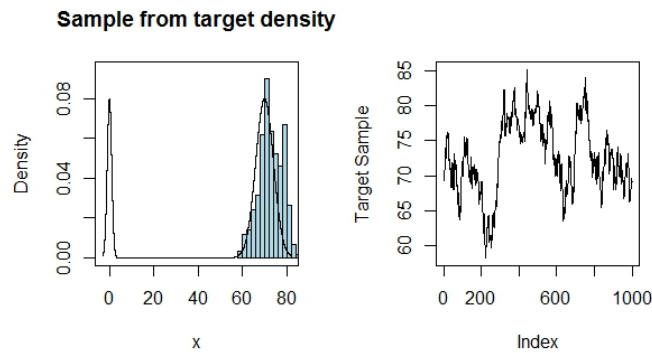


Figure 4.10: Target sample with starting point=70 inches

Chain (3) has candidate = $c(0,1)$, starting value 63 and after 10000 iterations, we have Figure (4.9).

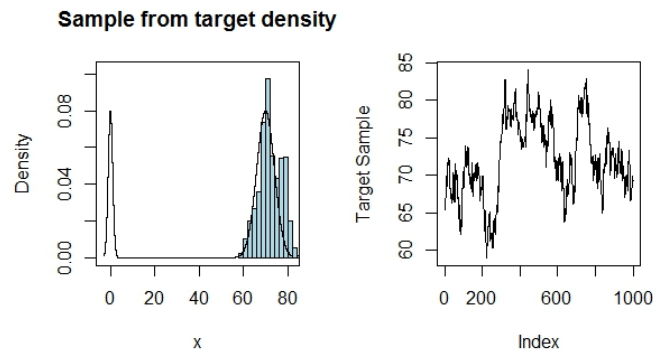


Figure 4.11: Target sample with starting point=66 inch

3 figure Explain to us that the target sample was not affected by the starting point.

And we can plot the rang of 3 chains together as follows:

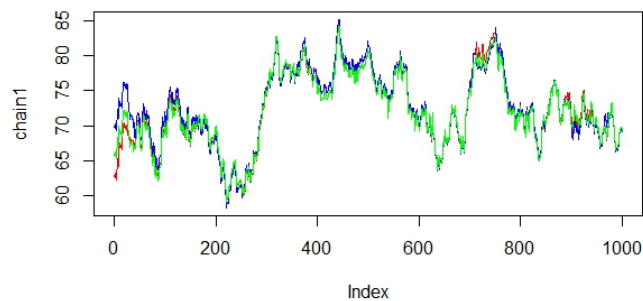


Figure 4.12: Rang of 3 chains

And if we use R program to calculate the mean of males height, we have mean for this sample ($\mu = 68.3$).

And if you calculate the standard deviation, we have ($S = 3$).

Full R code of this Example in Appendix.

4.3 conclusion

- Markov Chain is a random process that undergoes transitions from one state to another on a state space. It must possess a property that is usually characterized as "memorylessness" the probability distribution of the next state depends only on the current state and not on the sequence of events that preceded it. This specific kind of "memorylessness" is called the Markov property.
- Markov Chains have many applications as statistical models of real-world processes.
- The importance of Markov chains comes from two facts:
 - (i) there are a large number of physical, biological, economic, and social phenomena that can be described in this way.
 - (ii) There is a well developed theory that allows us to do computations.
- Markov chain Monte Carlo(MCMC) methods are a class of algorithms for sampling from a probability distribution based on constructing a Markov chain that has the desired distribution.
- The state of the chain after a number of steps is then used as a sample of the desired distribution.
- The quality of the sample improves as a function of the number of steps.
- The aims of Monte Carlo methods are to solve one or both the following problems:
 - (1) To generate samples X_n from a given probability distribution.
 - (2) To estimate expectations of functions under this distribution.
- MetropolisHastings Algorithm is a Markov chain Monte Carlo(MCMC) method for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult. This sequence can be used to approximate the distribution (i.e., to generate a histogram), or to compute an integral such as an expected value).

- The random walk Metropolis (RWM) is one of the most common Markov Chain Monte Carlo Algorithms in practical use today. Its theoretical properties have been explored for certain classes of target, and a number of results with important practical implications have been derived.
- The random walk Metropolis is special case of Metropolis Hasting Algorithm.
- We can use R program to generate samples from posterior distributions by MCMC methods.
- I would suggest that is the study of other types of MCMC and expand the simulation to include other types of probability distributions and the generation of mixed samples of distributions.

.1 R code of Markov Chain Simulation

```
MC.sim <- function(n,P,x1) {
  sim <- as.numeric(n)
  m <- ncol(P)
  if (missing(x1)) {
    sim[1] <- sample(1:m,1) # random start
  } else { sim[1] <- x1 }
  for (i in 2:n) {
    newstate <- sample(1:m,1,prob=P[sim[i-1],])
    sim[i] <- newstate
  }
  sim
}

options(width=55)
set.seed(12867)
#example 1
```

```
P <- matrix(c(0,.5,.5,5/8,1/8,2/8,2/3,1/3,0), nrow=3)
MC.sim(100, P, 2)
```

.2 The R code of Random Walk Metropolis Algorithm

```
bivnormMH<- function (rho, rho1 = 0.9, sigma = c(1.2, 1.2), steps = 10000,
                      type = "rw")
{
  if (rho < -1 | rho > 1) {
    stop("rho must be between -1 and 1")
  }
  if (steps < 100)
    warning("You should really do more than 100 steps")
  target = candidate = matrix(0, ncol = 2, nrow = steps)
  if (length(grep("[Rr]", type)) > 0) {
    type = "rw"
  }
  else {
    stop("Type must be rw")
  }
  x0 = c(0, 0)
  x1 = c(0, 0)
  mu = c(0, 0)
  if (type == "rw") {
    startValue = c(2, 2)
    sigma1 = 0.5
    var1 = sigma1^2
    sigma2 = 1
    var2 = sigma2^2
```

```

k = 2 * pi/1000
u = runif(steps)
z1 = rnorm(steps, 0, sigma1)
z2 = rnorm(steps, 0, sigma1)
w = 1 - rho^2
target[1, ] = startValue
x1 = target[1, ]
mu = target[1, ]
x0 = c(mu[1] + z1[1], mu[2] + z2[1])
candidate[2, ] = x0
for (n in 2:steps) {
  n1 = n - 1
  x1 = target[n1, ]
  x0 = candidate[n, ]
  canDens = exp(-1/(2 * var2 * w) * (x0[1]^2 - 2 *
                                     rho * x0[1] * x0[2] + x0[2]^2))
  curDens = exp(-1/(2 * var2 * w) * (x1[1]^2 - 2 *
                                     rho * x1[1] * x1[2] + x1[2]^2))

  if (u[n] < canDens/curDens) {
    target[n, ] = x0
  }
  else {
    target[n, ] = target[n1, ]
  }
  mu = target[n, ]
  x0 = c(mu[1] + z1[n], mu[2] + z2[n])
  if (n < steps)
    candidate[n + 1, ] = x0
}
}
target = data.frame(x = target[, 1], y = target[, 2])
plot(y ~ x, data = target, type = "l")
invisible(list(targetSample = target))

```

```

}

##Example
## random walk chain
chain.df<-bivnormMH(0.9, type = 'r')$targetSample
chain.df<-bivnormMH(0.99, type = 'r')$targetSample
##plotting scatterplot
plot(y ~ x, type = 'l', chain.df, main = 'Random Walk')

## plotting the chain
plot(x <- as.vector(chain.df[[1]]), type="l")

```

.3 R code of Normal Metropolis-Hasting

```

normMH <- function (theta0, candidate, steps = 1000, type = "rw",
                    randomSeed = NULL, startValue = NULL)
{
  if (steps < 100) {
    warning("Function should take at least 100 steps")
  }
  mu0 <- theta0[1]
  sigma0 <- theta0[2]
  mu <- candidate[1]
  sigma <- candidate[2]
  if (any(c(sigma0, sigma) <= 0))
    stop("All standard deviations must be strictly non-zero and positive")
  theta <- seq(from = mu0 - 3 * sigma0, to = mu0 + 3 * sigma0, by = 0.001)
  fx <- dnorm(theta, mu0, sigma0)
  targetSample <- rep(startValue, steps)
  if (type == "rw") {

```

```

if (!is.null(randomSeed))
  set.seed(randomSeed)
z <- rnorm(steps, mu, sigma)
u <- runif(steps)
if (is.null(startValue))
  startValue <- z[1]
targetSample[1] <- startValue
g <- rep(0, steps)
proposal <- rep(0, steps)
alpha <- rep(0, steps)
k <- 1/sigma0 * exp(-0.5 * ((targetSample[1] - mu0)/sigma0)^2)
g[1] <- k
i1 <- 1
for (n in 2:steps) {
  proposal[n] <- targetSample[i1] + z[n]
  k <- 1/sigma0 * exp(-0.5 * ((proposal[n] - mu0)/sigma0)^2)
  g[n] <- k
  k2 <- g[n]
  k3 <- g[i1]
  alpha[n] <- ifelse(k2/k3 > 1, 1, k2/k3)
  if (u[n] >= alpha[n]) {
    targetSample[n] <- targetSample[i1]
  }
  else {
    targetSample[n] <- proposal[n]
    i1 <- n
  }
}
}
else {
  if (!is.null(randomSeed))
    set.seed(randomSeed)
  z <- rnorm(steps, mu, sigma)

```



```

u <- runif(steps)
if (is.null(startValue))
  startValue <- z[1]
density0 <- dnorm(z, mu, sigma)
density1 <- dnorm(z, mu0, sigma0)
densityMix <- density1
alpha <- rep(0, steps)
targetSample[1] <- startValue
i1 <- 1
for (n in 2:steps) {
  alpha[n] <- density0[i1] * densityMix[n]/(density0[n] *
                                           densityMix[i1])

  alpha[n] <- ifelse(alpha[n] > 1, 1, alpha[n])
  if (u[n] >= alpha[n]) {
    targetSample[n] <- targetSample[i1]
  }
  else {
    targetSample[n] <- z[n]
    i1 <- n
  }
}
}

oldPar <- par(mfrow = c(1, 2), pty = "s")
h <- hist(targetSample, plot = FALSE)
ymax <- max(c(h$density, fx)) * 1.05
hist(targetSample, prob = TRUE, col = "light blue", xlim = range(theta),
      ylim = c(0, ymax), main = "Sample from target density",
      xlab = "x", ylab = "Density")
lines(theta, fx)
box()
plot(targetSample, type = "l", main = "", ylab = "Target Sample")
par(oldPar)
invisible(targetSample)

```

```

}

rseed <- floor(1e+06*runif(1)) +1

chain1 <- normMH(theta0 = c(3,2), candidate = c(0,1), steps = 10000,
                 type="rw", randomSeed = rseed, startValue =-2)
chain2 <- normMH(theta0 = c(3,2), candidate = c(0,1), steps = 10000,
                 type="rw", randomSeed = rseed, startValue =1)
chain3 <- normMH(theta0 = c(3,2), candidate = c(0,1), steps = 10000,
                 type="rw", randomSeed = rseed, startValue =4)

yRange <- range(c(chain1, chain2, chain3))
plot(chain1, type="l", col="red", ylim=yRange)
lines(chain2, col="blue")
lines(chain3, col="green")

```

.4 R code of Males Heights Example

```
normMixMH <- function (theta0, theta1, p, candidate, steps = 1000,
  type = "ind", randomSeed = NULL, startValue = NULL)
{
  if (steps < 100) {
    warning("Function should take at least 100 steps")
  }
  if (p <= 0 | p >= 1)
    stop("Mixture proportion p must be between 0 and 1")
  mu0 <- theta0[1]
  sigma0 <- theta0[2]
  mu1 <- theta1[1]
  sigma1 <- theta1[2]
  mu <- candidate[1]
  sigma <- candidate[2]
  if (any(c(sigma0, sigma1, sigma) <= 0))
    stop("All standard deviations must be strictly non-zero and positive")
  if (length(grep("[Ii]", type)) > 0) {
    type <- "ind"
  }
  else if (length(grep("[Rr]", type)) > 0) {
    type <- "rw"
  }
  else {
    stop("Type must be ind or rw")
  }
  theta <- seq(from = min(mu0 - 3 * sigma0, mu1 - 3 * sigma1),
    to = max(mu0 + 3 * sigma0, mu1 + 3 * sigma1), by = 0.001)
  fx <- p * dnorm(theta, mu0, sigma0) + (1 - p) * dnorm(theta,
    mu1, sigma1)

  targetSample <- rep(startValue, steps)
  if (type == "rw") {
```

```

if (!is.null(randomSeed))
  set.seed(randomSeed)
z <- rnorm(steps, mu, sigma)
u <- runif(steps)
if (is.null(startValue))
  startValue <- z[1]
targetSample[1] <- startValue
g <- rep(0, steps)
proposal <- rep(0, steps)
alpha <- rep(0, steps)
k1 <- p/sigma0 * exp(-0.5 * ((targetSample[1] - mu0)/sigma0)^2)
k2 <- (1 - p)/sigma1 * exp(-0.5 * ((targetSample[1] -
                                mu1)/sigma1)^2)

g[1] <- k1 + k2
i1 <- 1
for (n in 2:steps) {
  proposal[n] <- targetSample[i1] + z[n]
  k1 <- p/sigma0 * exp(-0.5 * ((proposal[n] - mu0)/sigma0)^2)
  k2 <- (1 - p)/sigma1 * exp(-0.5 * ((proposal[n] -
                                mu1)/sigma1)^2)

  g[n] <- k1 + k2
  k3 <- g[n]
  k4 <- g[i1]
  alpha[n] <- ifelse(k3/k4 > 1, 1, k3/k4)
  if (u[n] >= alpha[n]) {
    targetSample[n] <- targetSample[i1]
  }
  else {
    targetSample[n] <- proposal[n]
    i1 <- n
  }
}
}

```

```

else {
  if (!is.null(randomSeed))
    set.seed(randomSeed)
  z <- rnorm(steps, mu, sigma)
  u <- runif(steps)
  if (is.null(startValue))
    startValue <- z[1]
  density0 <- dnorm(z, mu, sigma)
  density1 <- dnorm(z, mu0, sigma0)
  density2 <- dnorm(z, mu1, sigma1)
  densityMix <- p * density1 + (1 - p) * density2
  alpha <- rep(0, steps)
  targetSample[1] <- startValue
  i1 <- 1
  for (n in 2:steps) {
    alpha[n] <- density0[i1] * densityMix[n]/(density0[n] *
                                              densityMix[i1])

    alpha[n] <- ifelse(alpha[n] > 1, 1, alpha[n])
    if (u[n] >= alpha[n]) {
      targetSample[n] <- targetSample[i1]
    }
    else {
      targetSample[n] <- z[n]
      i1 <- n
    }
  }
}

oldPar <- par(mfrow = c(1, 2), pty = "s")
h <- hist(targetSample, plot = FALSE)
ymax <- max(c(h$density, fx)) * 1.05
hist(targetSample, prob = TRUE, col = "light blue", xlim = range(theta),
      ylim = c(0, ymax), main = "Sample from target density",
      xlab = "x", ylab = "Density")

```

```

    lines(theta, fx)
    box()
    plot(targetSample, type = "l", main = "", ylab = "Target Sample")
    par(oldPar)
    invisible(targetSample)
}

rseed <- floor(1e+06*runif(1)) +1

chain1 <- normMixMH(theta0 = c(70,4), theta1 = c(0,1), p=0.5,
  candidate = c(0,1), steps = 1000, type="rw", randomSeed = rseed,
  startValue =63)
chain2 <- normMixMH(theta0 = c(70,4), theta1 = c(0,1), p=0.5,
  candidate = c(0,1), steps = 1000, type="rw", randomSeed = rseed,
  startValue =70)
chain3 <- normMixMH(theta0 = c(70,4), theta1 = c(0,1), p=0.5,
  candidate = c(0,1), steps = 1000, type="rw", randomSeed = rseed,
  startValue =66)

h <-c(70.8,66, 71.7,68.7, 67.6, 69.2, 66.5, 67.2, 68.3, 65.6, 63.0,
      68.3, 73.1, 67.6, 68.0, 71.0,61.3,76.2,66.3,69.7, 65.4, 70.0,
      62.9,68.5, 68.3, 69.4,69.2,68.0,71.9, 66.1, 72.4, 73.0, 68.0,
      68.7, 70.3, 63.7, 71.1,65.6,68.3, 66.3)

mean(h,40)
sd(h,40)
plot(h)
hist(h)
head(chain1,40)
head(chain2,40)
head(chain3,40)

yRange <- range(c(chain1, chain2, chain3))

```

```
plot(chain1, type="l", col="red", ylim=yRange)
lines(chain2, col="blue")
lines(chain3, col="green")
```

Bibliography

- [1] *Arnold O. Allen ,Probability, Statistics, And Queueing Theory: With Computer Science Applications,Academic Press Limited,London 1990*
- [2] *B. Hughes,Random Walks And Random Environments, Oxford University Press 1995*
- [3] *David M. McClendon, Notes On Random Walks Andmarkov Chains, Northwestern University March 2010.*
- [4] *David Sirl ,Markov Chains:An Introduction/Review,The University Of Queensland,Australia,April2005.*
- [5] *Dr S. K. Sahu, Lecture Notes: Bayesian Methods, Southampton Unversity (2001).*
- [6] *Evans, Michael J. And Rosenthal, J. S. Probability And Statistics, The Science Of Uncertainty, W. H. Freeman And Company, Newyork (2003).*
- [7] *Frank Beichelt And L. Paul Fatti, Stochastic Processes And Their Applications, Library Of Congress Cataloging In Puplication Data 2002.*
- [8] *James Norris, Markov Chain,Cambridge University 1997.*
- [9] *Jennifer A. Hoeting, Geof H. Givens,Computational Statistics, Colorado State University2013*
- [10] *Haggstrom.O, Finite Markov Chains And Algorithmic Applications, Cambridge University Press 2000.*

- [11] *Linda J.S Allen, An Introduction To Stochastic Processes With Applications To Biology, Taylor And Francis Group* 2010.
- [12] *Ming Liao, Applied Stochastic Processes, Taylor and Francis Group(U.S)* 2014.
- [13] *Olle Hggstrm,Finite Markov Chains and Algorithmic Applications,Cambridge University* 2002.
- [14] *Robert G. Gallager,Stochastic Processes: Theory For Applications,Cambridge University* 2013.
- [15] *Saeed Ghahramani, Fundamentals Of Probability With Stochastic Processes ,Western New England College* 2005.
- [16] *Spitzer, F. Principles Of Random Walk. Van Nostrand. Princeton*1964.