Markov Chains

Prof. Dr. Narayan Prasad Adhikari
Central Department of Physics
Tribhuvan University Kirtipur, Kathmandu, Nepal

November 8, 2022

Contents

■Random variables

- Consider an elementary event with a countable set of random outcomes, $x_1, x_2, ...x_k$ (e.g. you can consider a rolling dice OR a set of "Khodkhode", for example you can consider the outcome of two dices after rolling)
- You are data scientist so you need to consider this event occurring repeatedly say N times such that $N >>> 1$ and we count how often the outcome x_k is observed (N_k).
- The probabilities p_k for outcome x_k is

$$p_k = \lim_{N\to\infty}\left(\frac{N_k}{N}\right) \tag{1}$$

with $\sum_k p_k = 1$.
Obviously $0 \leq p_k \leq 1$
You are familiar with conditional probability $P(j/i)$, average of any outcomes of such random events x_i, its variances and so on.

■Notation

This lecture contains Markov Chain Stochastic Simulation so its about probabilities and statistics (I presume that you are familiar with probability density function).

- Distributions are identified with their density or probablity functions.
- Variables are generally treated as if they are continuous.
- Posterior densities are denoted by π and their approximations by q.
- $x, y, ...$ denote the observed quantities whereas unobserved quantities or parameters are denoted by Greek letters $\theta,\ \phi,\ ...$.

■Notation

- No distinctions between a random variable and its observed value
- Scalars and vectors both are denoted by small letters whereas capital letters represent matrices.
- The transpose of x is denoted by x'. Dimension of matrices is denoted by d.
- The component of A is $\bar{A}$.
- The probability of an event A is denoted by $Pr(A)$.
- Expectation and variance of a quantity x are $E(x)$ and $Var(x)$.
- The covariance and correlation between random quantities x and y are $Cov(x, y)$ and $Cor(x, y)$
- The number of elements of a set A is denoted by $\#A$
- . denotes approximations and they are with appropriate symbols

Markov Chains - Introduction

Figure: Andrei Andreivich Markov - Russian Mathematician (1856-1922)

- Markov dependence is a concept attributed to the Russian mathematician Andrei Andreivich Markov that at the start of the 20th century investigated the alternance of vowels and consonants in the poem *Onegin* by Poeshkin
- He developed a probabilistic model where successive results depended on all their predecessors only through the immediate predecessor. The model allowed him to obtain good estimates of the relative frequency of vowels in the poem.

■Markov Chains - Introduction

- A Markov chain is a special type of stochastic process (random and usually dependent on time), which deals with characterization of sequences of random variables. Special interest is paid to the dynamic and the limiting behaviors of the sequence. A stochastic process can be defined as a collection of random quantities $\{\theta^{(t)} : t \in T\}$ for some set T.
- The set $\{\theta^{(t)} : t \in T\}$ for some set T, is said to be a stochastic process with state space S and index (or parameter) set T. T is taken to be countable, defining a discrete time stochastic process , i.e. $T \in N$, with N the set of natural numbers. State Space is the set of all possible and known states of a system. In state-space, each unique point represents a state of the system. For example, Take a pendulum moving in to and fro motion. Then its state is represented by its angle and angular velocity. Similarly consider rolling of two dices. Then the state space is $\{2, 3, ..., 12\}$

■Random variables & PDF

- The state space will be a subset of R^d representing support of a parameter vector.
- Find a formula for the probability distribution of the total number of heads obtained in four tosses of a balanced coin.
- The sample space, probabilities and the value of the random variable (X where X is the number of heads obtained in four tosses) are given in table.
  From the table we can determine the probabilities as

$$P(X=0)=\frac{1}{16},\ P(X=1)=\frac{4}{16},\ P(X=2)=\frac{4}{16},\ P(X=3)=\frac{4}{16},\ P(X=4)=\frac{1}{16} \tag{2}$$

Notice that the denominators of the five fractions are the same and the numerators of the five fractions are $1, 4, 6, 4, 1$. The numbers in the numerators is a set of binomial coefficients.

■Random variables & PDF

$$
\begin{aligned}
\frac{1}{16} = \left({}^4C_0\right)\frac{1}{16},\ \ \frac{4}{16} = \left({}^4C_1\right)\frac{1}{16} \\
\frac{6}{16} = \left({}^4C_2\right)\frac{1}{16},\ \ \frac{4}{16} = \left({}^4C_3\right)\frac{1}{16},\ \ \frac{1}{16} = \left({}^4C_4\right)\frac{1}{16}
\end{aligned}
\tag{3}
$$

We can then write the probability mass function as

■Random variables & PDF

TABLE 1. **Probability of a Function of the Number of Heads from Tossing a Coin Four Times.**

Table R.1 Tossing a Coin Four Times		
Element of sample space	Probability	Value of random variable X (x)
HHHH	1/16	4
HHHT	1/16	3
HHTH	1/16	3
HTHH	1/16	3
THHH	1/16	3
HHTT	1/16	2
HTHT	1/16	2
HTTH	1/16	2
THHT	1/16	2
THTH	1/16	2
TTHH	1/16	2
HTTT	1/16	1
THTT	1/16	1
TTHT	1/16	1
TTTH	1/16	1
TTTT	1/16	0

■Definition and Transition Probabilities

- A Markov chain is a stochastic process where given the present state, past and future states are independent.
- This property can be formally stated through

$$Pr(\theta^{(n+1)} \in A|\theta^{(n)} = x, \theta^{(n-1)} \in A_{n-1}, ..., \theta^{(0)} \in A_0) \tag{4}$$

$$= Pr\left(\theta^{(n+1)} \in A|\theta^{(n)} = x\right) \tag{5}$$

for all sets $A_0,\ A_1, ...A_{n-1}, A \in S$. All these $\theta' s$ may be random numbers from set $\{2, 3, ..., 12\}$ in case of rolling of two dices.

Definition and Transition Probabilities

- The state space in this case is

$$S = \left\{\begin{array}{l} (1,1),(1,2),(1,3),(1,4),(1,5),(1,6), \\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6), \\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6), \\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6), \\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6), \\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \end{array}\right\}$$

■Definition and Transition Probabilities

Markovian property equation 5 can be rewritten as

$$E\left[f(\theta^{(n)}|\theta^{(m)},\theta^{(m-1)},....\theta^{(0)})\right] = E\left[f(\theta^{(n)}|\theta^{(m)})\right] \tag{6}$$

for all bounded functions f and $n > m \geq 0$.

■Definition and Transition Probabilities

Equivalently,

$$Pr(\theta^{(n+1)} = y | \theta^{(n)} = x, \theta^{(n-1)} = x_{n-1}, ..., \theta^{(0)} = x_0) \tag{7}$$
$$= Pr\left(\theta^{(n+1)} = y | \theta^{(n)} = x\right) \tag{8}$$

for all $x_0, x_1, ..., x_{n-1}, x, y \in S$. This form is obviously appropriate ony for discrete state spaces.

- If a sequence of numbers follows the above graphical model, it is a Markov Chain.
  - That is, $p(x_5 | x_4, x_3, x_2, x_1) = p(x_5 | x_4)$.
  - So the probability of a certain state being reached, depends only on the previous state of the chain.

■Definition and Transition Probabilities

In general, the probabilities in 5 depend on x, A and n. When they do not depend on n, the chain is said to be homogeneous. In this case, a transition function or kernel $P(x, A)$ can be defined as:
1. for all $x \in S$, $P(x, .)$ is a probablity distribution over S;
2. for all $A \in S$, the function $x \mapsto P(x, A)$ can be evaluated.
It is also useful when dealing with discrete state space to identify $P(x, \{y\}) = P(x, y)$. This function is called a transition probability and satisfies:

- $P(x, y) \geq 0, \forall x, y \in S$;
- $\sum_{y \in S} P(x, y) = 1, \forall x \in S$; as any probability distribution $P(x, .)$ should

■Example 1: Random Walk

Consider a particle moving independently left and right on the line with successive displacements from its current position governed by a probability function f over the integers and $\theta^{(n)}$ representing its position at instant n, $n \in N$. Initially, $\theta^{(0)}$ is distributed according to some distribution π^0. The positions can be related as $\theta^{(n)} = \theta^{(n-1)} + w_n = w_1 + w_2 + ... + w_n$ where the w_i are independent random variables with probability function f . So, $\{\theta^{(n)} : n \in N\}$ is a Markov chain in Z.

■Example 1: Random Walk

The position of the chain at instant $t = n$ is described probabilistically by the distribution of $w_1 + w_2 + ... + w_n$. If $f(1) = p,\ f(-1) = q$ and $f(0) = r$ with $p + q + r = 1$ then the transition probabilities are given by

$$P(x, y) = \begin{cases} p, \text{if } y = x + 1 \\ q, \text{if } y = x - 1 \\ r, \text{if } y = x \\ 0, \text{if } y \neq x, x - 1, x + 1 \end{cases} \tag{9}$$

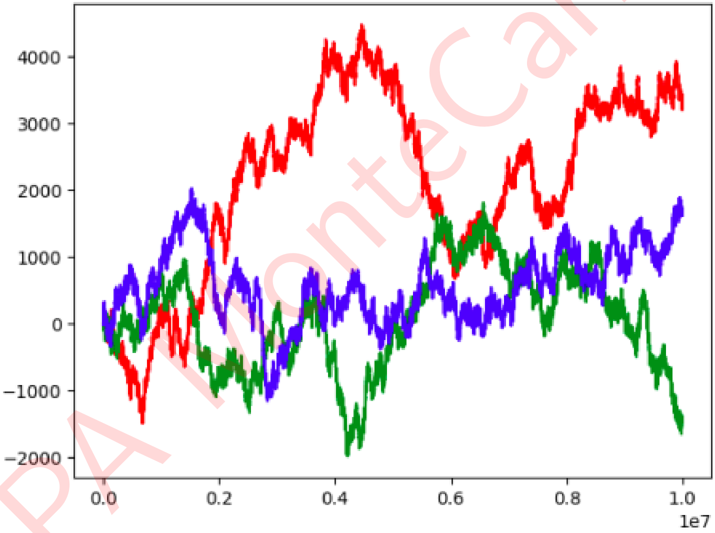
■Example 1: Random Walk

- Consider a random walk - forward direction +1, backward direction -1 and remaining in the same position as 0 as described in previous page.
- We can use python code to generate a random walk. After n steps the displacement will be zero but root mean squared deviation (RMSD) will not be. After a large numbers of such walks the average displacement will be RMSD.

```python
import numpy as np
import matplotlib.pyplot as plt
import random
def rw1D(n):
    x, y = 0, 0
    # Generate the time points [1, 2, 3, ... , n]
    timepoints = np.arange(n + 1)
    positions = [y]
    for i in range(1, n + 1):
        # Randomly select either UP or DOWN
        step = random.random()

        # Move the object up or down
        if step <=0.5:
            y += 1
        elif step >0.5:
            y -= 1
        # Keep track of the positions
        positions.append(y)
    return timepoints, positions
rw1 = rw1D(10000000)
rw2 = rw1D(10000000)
rw3 = rw1D(10000000)
plt.plot(rw1[0], rw1[1], 'r-', label="rw1")
plt.plot(rw2[0], rw2[1], 'g-', label="rw2")
plt.plot(rw3[0], rw3[1], 'b-', label="rw3")
plt.show()
```

■Example 1: Random Walk



■Definition and Transition Probabilities

CW/HW:You understand the meaning of $P(x, y)$. Also as an example discuss Random Walk problems in 1 and 2 dimensions. Students can write a python code for random walk in 1 & 2 dimensions.

■Markov chain Example 2- Ehrenfest model

Consider a total of r balls distributed in two urns with x balls in the first urn and r — x in the second urn. Take one of the r balls at random and put it in the other urn. Repeat the random selection process independently and indefinitely. This procedure was used by Ehrenfest to model the exchange of molecules between two containers. If $X^{(n)}$ represents the number of balls in the first urn after n exchanges then $\{X^{(n)} : n \in N\}$ is a Markov chain with state space $S = \{0, 1, 2, ...r\}$ and transition probabilities

$$P(x,y) = \begin{cases} x/d, \text{if } y = x+1 \\ 1 - x/d, \text{if } y = x-1 \\ 0, p, \text{if } |y-x| \neq 1 \end{cases} \tag{10}$$

■Simulation of Ehrenfest model

Gas molecules move about randomly in a box which is divided into two halves symmetrically by a partition. A hole is made in the partition. Suppose there are N molecules in the box. Think of the partitions as two urns containing balls labeled 1 through N. Molecular motion can be modeled by choosing a number between 1 and N at random and moving the corresponding ball from the urn it is presently in to the other. This is a historically important physical model introduced by Ehrenfest in the early days of statistical mechanics to study thermodynamic equilibrium.

The set of states of the Markov chain is $S = \{0, 1, 2, ..., N\}$ representing the number of molecules in one partition of the box.

■Simulation of Ehrenfest model (HW)

Do a computer simulation of the Markov chain for N $=100$. Start from state 0 (one of the partitions is empty) and follow the chain up to 1000 steps. Draw a graph of the number of molecules in the initially empty partition as a function of the number of steps. On the basis of your simulation, would you expect to observe during the course of the simulation a return to state 0?

It takes 2^{200} steps to make a box empty or to observe state $S=0$. If a molecule remains in a box for say $1/1000$ seconds then it takes around 10^{20} years!!! WOW?

■Transition probabilities

In the case o f discrete state spaces $S = x_1, x_2, ...$, a transition matrix P with (i, j)th element given by $P(x_i, x_j)$ can be defined. If S is finite with r elements, the transition matrix P is given by

$$\mathbf{P} = \begin{pmatrix} P(x_1, x_1) & \dots & P(x_1, x_r) \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ P(x_r, x_1) & \dots & P(x_r, x_r) \end{pmatrix} \tag{11}$$

- Transition matrices have all lines summing to one. Such matrices are called stochastic and have a few interesting properties.

- For instance, at least one eigenvalue of a stochastic matrix equals one and the product of stochastic matrices always produces a stochastic matrix. Of course, countable state spaces will lead to an infinite number of eigenvalues.

Transition probabilities

- Transition probabilities from state x to state y over m steps, denoted by $P^m(x,y)$, is given by the probability of a chain moving from state x to state y in exactly m steps. It can be obtained for $m \geq 2$ as

$$
\begin{aligned}
P^m(x,y) &= Pr\left(\theta^{(m)} = y | \theta^{(0)} = x\right) \\
&= \sum_{x_1} \ldots \sum_{x_{m-1}} Pr\left(\theta^{(m)} = y, \theta^{(m-1)} = x_{m-1}, \ldots \theta^{(1)} = x_1 | \theta^{(0)} = x\right) \\
&= \sum_{x_1} \ldots \sum_{x_{m-1}} Pr\left(\theta^{(m)} = y, \theta^{(m-1)} = x_{m-1}) \ldots Pr(\theta^{(1)} = x_1 | \theta^{(0)} = x\right) \\
&= \sum_{x_1} \ldots \sum_{x_{m-1}} P(x, x_1) P(x_1, x_2) .. P(x_{m-1}, y) \qquad (12)
\end{aligned}
$$

where the second equality is due to the Markovian property of the process.

■Transition probabilities

- The last equality means that the matrix containing elements $Pm(x,y)$ is also a stochastic matrix and is given by $P\ m$ obtained by the matrix product of the transition matrix $P\ m$ times. Also, for completeness, $PI(x,y) = P(x,y)$ and $P^0(x,y) = I(x=y)$.

The above derivation can be used to establish that

$$\begin{aligned} & P^{n+m}(x,y) \\ & = \sum_z Pr\left(\theta^{(n+m)} = y|\theta^{(n)} = z, \theta^{(0)} = x\right) Pr\left(\theta^{(n)} = z, \theta^{(0)} = x\right) \\ & = \sum_z P^n\left(x,z\right) P^m\left(z,y\right) \end{aligned} \tag{13}$$

■Transition probabilities

Equations 13 are usually called Chapman-Kolmogorov equations. All summations are with respect to the elements o f the state space S and results are valid for any stage o f the chain due to the assumed homogeneity. Higher transition matrices can be formed with these higher transition probabilities and it can be shown that they satisfy the relation $P^{n+m} = P^n P^m$ and, in particular, $P^{n+1} = P^n P$.

■Transition probabilities

The marginal distribution of the nth stage can be defined by the row vector $\pi^{(n)}$ with components $\pi^{(n)}(x_i)$, for all $x_i \in S$. For finite state spaces, this is a r-dimensional vector

$$\pi^{(n)} = \left(\pi^{(n)}(x_1), ...\pi^{(n)}(x_r)\right) \tag{14}$$

When $n = 0$, this is the initial distribution of the chain. Then

$$\begin{aligned} \pi^{(n)}(y) &= Pr(\theta^{(n)} = y) \\ &= \sum_{x \in S} Pr\left(\theta^{(n)} = y | \theta^{(0)} = x\right) Pr\left(\theta^{(0)} = x\right) \\ &= \sum_{x \in S} P^n(x, y)\pi^{(0)}(x) \end{aligned} \tag{15}$$

■Transition probabilities

The above equation can be written in matrix notation as $\pi^{(n)} = \pi^{(0)} P^n$. Also, since the same is valid for $n-1$, $\pi^{(n)} = \pi^{(0)} P^{n-1} P = \pi^{(n-1)P}$.
The probability of any event $A \in S$ for a Markov chain starting at x is denoted by $Pr_x(A)$. The hitting time of A is defined as $T_A = min\{n \geq 1 : \theta^{(n)} \in A\}$ if $\theta^{(n)} \in A$ for some $n > 0$. Otherwise, $T_A = \infty$. If $A = \{a\}$, the notation $T_{\{a\}} = T_a$ is used.

■Decomposition of the State Space

A few quantities of interest are important in the classification of states of a Markov chain with state space S and transition matrix P :
(i) The probability of the chain starting from state x hitting state y at any posterior step is $\rho_{xy} = Pr_x\left(T_y < \infty\right)$;
(ii) The number of visits of a chain to a state y is

$$N(y) = \#\{n > 0 : \theta^{(n)} = y\} = \sum_{n=1}^{\infty} I\left(\theta^{(n)} = y\right)$$

where I is indicator function

$$I(x \in A) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if} x \notin A \end{cases}$$

HW:Show that $E(T_y|\theta^0 = x) = \sum_{n=0}^{\infty} Pr_x(T_y > n)$ **and** $E(N(y)|\theta^0 = x) = \sum_{n=0}^{\infty} P^n(x,y)$.

■Decomposition of the State Space

A state $y \in S$ is said to be recurrent if the Markov chain, starting in y, returns to y with probability 1 $(\rho_{yy} = 1)$ and is said to be transient if it has positive probability of not returning to y $(\rho_{yy} < 1)$. An absorbing state $y \in S$ is recurrent because

$$Pr_y(T_y = 1) = Pr_y(\theta^{(1)} = y) = P(y, y) = 1$$

and therefore $(\rho_{yy} = 1)$.

If a Markov chain starts at a recurrent state y, the hitting (or return, in this case) time of y, T_y, is a finite random quantity whose mean μ_y can be evaluated. If this mean is finite, the state y is said to be positive recurrent and otherwise the state is said to be null recurrent. Positive recurrence is a very important property for establishing limiting results (Next lecture).

■Decomposition of the State Space

An important result describing analytically the difference between a recurrent and a transient state is that

- if $y \in S$ is a transient state then, for all $x \in S$,

$$Pr_x(N(y) < \infty) = 1 \text{ and } E\left[N(y)|\theta^{(0)} = x\right] = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty$$

- if $y \in S$ is a transient state then, for all $x \in S$,

$$Pr_x(N(y) = \infty) = 1 \text{ and } E\left[N(y)|\theta^{(0)} = y\right] = \infty$$

So, recurrent states are infinitely often (i.o.) visited with probability one.
The expected number of visits is finite if the state is transient.

■Decomposition of the State Space

It is interesting to investigate possible decompositions of S in subsets of recurrent and transient states. From this decomposition, probabilities of the chain hitting a given set of states can be evaluated. For states x and y in S, $x \neq y$, x is said to hit y, denoted $x \to y$, if $\rho_{xy} > 0$. A set $C \subseteq S$ is said to be closed if $\rho_{xy} = 0$ for $x \in C$ and $y \notin C$.

In obvious nomenclature, it is said to be irreducible if $x \to y$ for every pair $x, y \in C$. A chain is said to be irreducible if S is irreducible.

■Decomposition of the State Space

It is not difficult to show that the condition $\rho_{xy} > 0$ is equivalent to $P^n(x, y) > 0$ for some n ¿ 0. This can be used to show that if $x \in S$ is recurrent and $x \to y$ then y is also recurrent. In this case, $y \to x$ and one can write $x \leftrightarrow y$ when $x \to y$ and $y \to x$. In other words, recurrence defines an equivalence class with respect to the $\leftrightarrow$ operation. Also, $\rho_{xy} = \rho_{yx} = 1$. In fact, a stronger result is valid: null recurrence and positive recurrence also define equivalence classes. If $C \subseteq S$ is a closed, finite, irreducible set of states then all states of C are recurrent.

■Decomposition of the State Space -Example

Birth and Death Processes Consider a Markov chain that from the state x can only move in the next step to one o f the neighboring states x — 1 , representing a death, x or x + 1, representing a birth. The transition probabilities are given by

$$P(x,y) = \begin{cases} p_x, & \text{if } \; y = x+1 \\ q_x, & \text{if } \; y = x-1 \\ r_x, & \text{if } \; y = x \\ 0, & \text{if } \; |y-x| > 1 \end{cases}$$

where $p_x,\ q_x,$ and r_x are non-negative with $p_x + q_x, + r_x = 1$ and $q_0 = 0$. Note also that Ehrenfest model is special case of birth and death processes. Irreducible chains are obtained when $p_x > 0$ for $x \geq 0$ and $p_x > 0$ for $x > 0$.

■Decomposition of the State Space -Example

It is possible to determine if a state y is recurrent or transient even for an infinite state space by studying the convergence of the series $\sum_{y=0}^{\infty} \gamma_y$ where

$$\gamma_y = \begin{cases} 1 \text{ if } \ y = 0 \\ \frac{q_1 \dots q_y}{p_1 \dots p_y} \ \text{ if } \ y > 0 \end{cases}$$

If the sum diverges, the chain is recurrent. Otherwise, the chain is transient.

If S is finite and 0 is an absorbing state, the absortion probability is

$$\rho_{\{}0\}(x) = \rho_{x0} = \frac{\sum_{y=x}^{d-1} \gamma_y}{\sum_{y=0}^{d-1} \gamma_y}, \quad x = 1, \dots d-1. \tag{16}$$

This example just discussed is optional.

■Stationary (Equilibrium) Distributions

A fundamental problem for Markov chains in the context of simulation is the study of the asymptotic behavior of the chain as the number of steps or iterations $n \to \infty$. A key concept is that of a stationary distribution π. A distribution π is said to be a stationary distribution of a chain with transition probabilities $P(x, y)$ if

$$\sum_{x \in S} \pi(x) P(x, y) = \pi(y), \quad \forall y \in S \tag{17}$$

Equation 17 can be written in matrix form as

$$\pi P = \pi \tag{18}$$

The reason o f the name is clear from the above equation. If the marginal distribution at any given step n is tt then the distribution at the next step is $\pi P = \pi$.

■Stationary (Equilibrium) Distributions

Once the chain reaches a stage where π is the distribution of the chain, the chain retains this distribution for all subsequent stages. This distribution is also known as the invariant or equilibrium distribution for similar interpretations.

One can show that if the stationary distribution π exists and $\lim_{n\to\infty} P^n(x,y) = \pi(y)$ then, independently of the initial distribution of the chain, $\pi^{(n)}$ will approach π as $n \to \infty$. In this sense, the distribution is also referred to as the limiting distribution.

■Equilibrium Distributions- Example

Consider $\{\theta^{(n):n\geq 0}\}$, a Markovian chain in $S = \{0, 1\}$ with initial distribution $\pi^{(0)}$ given by $\pi^{(0)} = \left(\pi^{(0)}(0), \pi^{(0)}(1)\right)$ and transition matrix P given by

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \tag{19}$$

The stationary distribution π is the solution of the system $\pi P = \pi$ that gives the equations

$$\pi(0)P(0, y) + \pi(1)P(1, y) = \pi(y), \quad y = 0, 1 \tag{20}$$

The solution is $\pi = (q, p)/(p + q)$, a distribution that can be shown to be invariant for the stages of the chain.

CW/HW: Prove above solution. Also write a python code to get numerical values of π for given q=0.5, p=0.5. Also try different values of p and q. Do you still get stationary values of π for $(p + q) = 2$?

■Equilibrium Distributions

Also discuss the python code for 3×3 transition matrix.// Further derivation from Binder's book.

■Equilibrium Distributions: Example - Gibbs Sampler

This example provides a very simple special case of the Gibbs sampler. The complete form of the Gibbs sampler will be our next topics, here let us consider a simple and special case of Gibbs Sampler. In this special case, the state space is $S = \{0, 1\}$ and define a probability distribution π over S as

	θ_2	
θ_1	*0*	*1*
0	π_{00}	π_{01}
1	π_{10}	π_{11}

The probability vector π contains the above probabilities in any fixed order, say $(\pi_{00},\ \pi_{01},\ \pi_{10},\ \pi_{11})$.

The chain now consists of a bidimensional vector $\theta^{(n)} = (\theta_1^{(n)}, \theta_2^{(n)})$.

Although this introduces some novelties in the presentation they can

■Equilibrium Distributions: Example - Gibbs Sampler

be removed by considering a scalar chain $\psi^{(n)}$ that assumes values that are in correspondence with the $\theta^{(n)}$ chain, e.g. $\psi^{(n)} = 10\ \theta_1^{(n)}\ +\ \theta_2^{(n)}$. This is always possible for discrete state spaces. Therefore we do not make any distinction between scalar and vector chains.

Consider the following transition probabilities:

- For the first component θ_1, the transition probabilities are given by the conditional distribution π_1 of $\theta_1|\theta_2 = j$,

$$\pi_1(0|j) = \frac{\pi_{0j}}{\pi_{+j}} \text{ and } \pi_1(1|j) = \frac{\pi_{1j}}{\pi_{+j}}$$

  where $\pi_{+j} = \pi_{0j} + \pi_{1j}, j = 0, 1$.
- For the second component θ_2, the transition probabilities are given by the conditional distribution

■Equilibrium Distributions: Example - Gibbs Sampler

π_2 of $\theta_2|\theta_1 = i$,

$$\pi_2(0|i) = \frac{\pi_{i0}}{\pi_{i+}} \text{ and } \pi_2(1|i) = \frac{\pi_{i1}}{\pi_{i+}}$$

where $\pi_{i+} = \pi_{i0} + \pi_{i1}, i = 0, 1.$

The overall transition probability o f the chain is

$$\begin{aligned} P((i,j),(k,l)) &= Pr(\theta^{(n)} = (k,l)|\theta^{(n-1)} = (i,j)) \\ &= Pr(\theta_2^{(n)} = l|\theta_1^{(n)} = i)Pr(\theta_1^{(n)} = k|\theta_1^{(n)} = j) \\ &= \frac{\pi_{kl}}{\pi_{k+}}\frac{\pi_{kj}}{\pi_{+j}} \end{aligned} \tag{21}$$

for $(i,j),(k,l) \in S$. Thus a 4×4 matrix P can be formed.

■Equilibrium Distributions

The existence and uniqueness of stationary distributions can be studied through weaker results. Let $N_n(y)$ be the number of visits to state y in n steps and define $G_n(x,y) = E_x[N_n(y)]$, the average number of visits of the chain to state y and $m_y = E_y(T_y)$, the average return time to state y. Then, $G_n(x,y) = \sum_{k=1}^{n} P^k(x,y)$ and $\lim_{n\to\infty} G_n(x,y)/n$ provides the limiting occupation of state y in a chain observed for an infinitely long number of steps.

■Limiting Theorems

- There are situations where stationary distributions are available but limiting distributions are not (See above examples).
- In order to establish limiting results, one characterization of states still absent and that must be introduced. This is the notion of periodicity.

- The period of a state x, denoted by d_x is the largest common divisor of the set
$$\{n \geq 1 : P^n(x,x) > 0\}$$

It is obvious that $P(x,x) > 0$ implies that $d_x = 1$ and that if $x \leftrightarrow y$ then $d_x = d_y$. Therefore, the states of an irreducible chain have the same period.

■Limiting Theorems

- **Aperiodci state:** A state x is aperiodic if $d_x = 1$
- **Ergodic state:** An aperiodic and positive recurrent state is said to be ergodic state.

- A chain is periodic with period d if all its states are periodic with period $d > 1$ and aperiodic if all its states are aperiodic. Finally, a chain is ergodic if all its states are ergodic.

- In an ergodic scenario, the average outcome of the group is the same as the average outcome of the individual over time. An example of an ergodic systems would be the outcomes of a coin toss (heads/tails). If 100 people flip a coin once or 1 person flips a coin 100 times, you get the same outcome.

■Limiting Theorems

- Once ergodicity of the chain is established, important limiting theorems can be stated. The first and most important one is the ergodic theorem. The ergodic average of a real-valued function $t(\theta)$ is the average

$$\bar{t}_n = \left(\frac{1}{n}\right) \sum_{i=1}^{n} t(\theta^{(i)}).$$

- If the chain is esgodic and $E_\pi[t(\theta)] < \infty$ for the unique limiting distribution π then

$$\bar{t}_n \to (a.s.) E_\pi[t(\theta)] \text{ as } n \to \infty \tag{22}$$

where $a.s.$ refers to almost sure.

■Limiting Theorems

- This result is a Markov chain equivalent of the law of large numbers. It states that averages of chain values also provide strongly consistent estimates of parameters of the limiting distribution tt despite their dependence.

- If $t(\theta) = I(\theta = x)$ then the ergodic averages are simply counting the relative frequency of values of xs in realizations of the chain. By the ergodic theorem, this relative frequency converges almost surely to $\pi(x) = \frac{1}{m_x}$, the average frequency of visits to state x.

  HW/CW: You collect sales of shoes from a store say Bhatbhateni for three months. It may increase but if you analyze difference it remains almost stationary.

■Reversible Chains

- Let $\left(\theta^{(n)}\right)_n \geq 0$ be an homogeneous Markov chain with transition probabilities $P(x, y)$ and stationary distribution π. Assume that one wishes to study the sequence of states $\theta^{(n)}, \theta^{(n-1)}, \ldots$ in reverse order. One can show that this sequence satisfies

$$Pr(\theta^{(n)} = y|\theta^{(n+1)} = x, \theta^{(n+2)} = x_2, \ldots) = Pr(\theta^{(n)} = y|\theta^{(n+1)} = x)$$

which defines a Markov chain. The Transition probabilities are

$$\begin{aligned} P_n^*(x, y) &= Pr(\theta^{(n)} = y|\theta^{(n+1)} = x) \\ &= \frac{Pr(\theta^{(n+1)} = x|\theta^{(n)} = y)Pr(\theta^n = y)}{Pr(\theta^{(n+1)} = x} \\ &= \frac{\pi^{(n)}(y)P(y, x)}{\pi^{(n+1)}(x)} \end{aligned} \quad (23)$$

and in general the chain is not homogeneous.

■Reversible Chains

- If $n \to \infty$ or alternatively, $\pi^{(0)} = \pi$, then $P_n^*(x,y) = P*(x,y) = \pi(y)P(y,x)/\pi(x)$ and the chain becomes homogeneous. If$P*(x,y) = P(x,y)$ for all x and $y \in S$, the time reversed Markov chain has the same transition probabilities as the original Markov chain. Markov chains with such a property are said to be reversible and the reversibility condition is usually written as

$$\pi(x)P(x,y) = \pi(y)P(y,x) \quad \text{for all} \quad x,y \in S \tag{24}$$

■Reversible Chains

- It can be interpreted as saying that the rate at which the system moves from x to y when in equilibrium, $\pi(x)P(x, y)$, is the same as the rate at which it moves from y to x, $\pi(y)P(y, x)$. For that reason, equation 24 is sometimes referred to as the detailed balance equation; balance because it equates the rates of moves through states and detailed because it does it for every possible pair of states.

  HW/CW: You can prove that the irreducible birth and death chains are reversible.
  Note that the detailed balance is required for getting and maintaining the stationary (equilibrium) Distribution.

■Reversible Chains

- Reversible chains are useful because if there is a distribution π satisfying equation 24 for an irreducible chain, then the chain is positive recurrent, reversible with stationary distribution π. This is easily obtained by summing over y both sides of 24 to give 17. Construction of Markov chains with a given stationary distribution π reduces to finding transition probabilities $P(x, y)$ satisfying 24.

■Reversible Chains- Example : Metropolis Algorithm

-