Markov Chain

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

February 22, 2025

Contents

- Random variables
- Definition and Transition Probabilities
- Decomposition of the State Space
- Stationary distributions
- Limiting Theorems
- Reversible chains
- Continuous State Spaces

■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.
- $x, y, \ldots$ denote the observed quantities whereas unobserved quantities or parameters are denoted by Greek letters $\theta,\ \phi,\ \ldots$.

■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 gives any of the outcomes $\{2, 3, ..., 12\}$ (see state space of

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

$$\begin{gathered}\frac{1}{16}=({}_4\mathrm{C}_0)\,\frac{1}{16},\;\;\frac{4}{16}=({}_4\mathrm{C}_1)\,\frac{1}{16}\\ \frac{6}{16}=({}_4\mathrm{C}_2)\,\frac{1}{16},\;\;\frac{4}{16}=({}_4\mathrm{C}_3)\,\frac{1}{16},\;\;\frac{1}{16}=({}_4\mathrm{C}_4)\,\frac{1}{16}\end{gathered} \tag{3}$$

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.

- 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 $n > m \geq 0$.

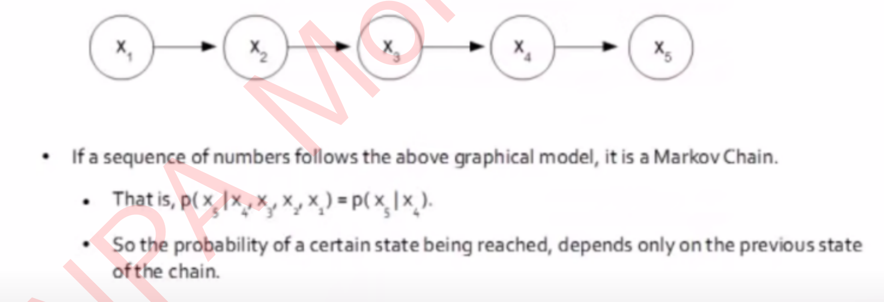
■Definition and Transition Probabilities

Equivalently,

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

$$= Pr\left(\theta^{(n+1)} = y|\theta^{(n)} = x\right) \quad (8)$$

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



■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

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

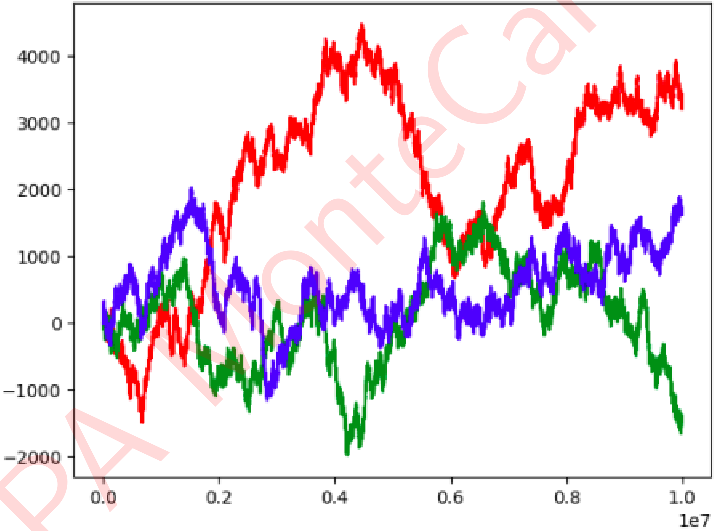
Example 1: Random Walk

- 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.

```
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 Ehrenfest model - book

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

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

■Simulation of Ehrenfest model (Explanation)

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 of discrete state spaces $S = x_1, x_2, \ldots$, 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

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

- 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

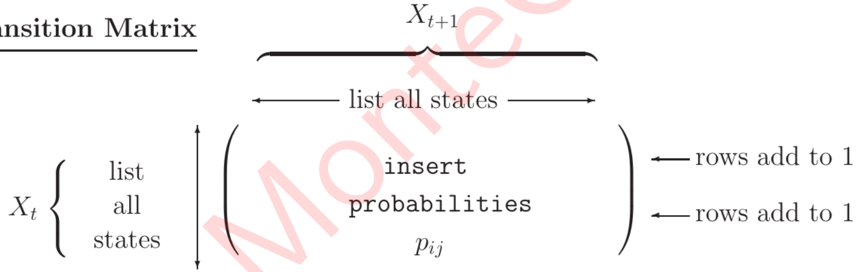
In the transition matrix P:

- the ROWS represent NOW, or FROM (Xt);
- the COLUMNS represent NEXT, or TO (Xt+1);
- entry (i, j) is the CONDITIONAL probability that NEXT = j,

given that
NOW = i: the probability of going FROM state i TO state j.

■Transition probabilities

Transition Matrix



■Transition probabilities

Notes:

1. The transition matrix P must list all possible states in the state space S.
2. P is a square matrix (N $\times$ N), because X_{t+1} and X_t both take values in the same state space S (of size N).

3. The **rows** of P should each *sum to 1:*

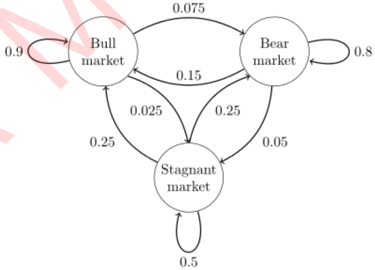
$$\sum_{j=1}^{N} p_{ij} = \sum_{j=1}^{N} \mathbb{P}(X_{t+1} = j \mid X_t = i) = \sum_{j=1}^{N} \mathbb{P}_{\{X_t = i\}}(X_{t+1} = j) = 1.$$

   This simply states that X_{t+1} *must* take one of the listed values.

4. The **columns** of P do **not** in general sum to 1.

■Transition probabilities

A state diagram for a simple example is shown in the figure using a directed graph to picture the state transitions. The states represent whether a hypothetical stock market is exhibiting a bull market, bear market, or stagnant market trend during a given week. According to the figure, a bull week is followed by another bull week 90% of the time, a bear week 7.5% of the time, and a stagnant week the other 2.5% of the time. Labeling the state space $\{1 = \text{bull}, 2 = \text{bear}, 3 = \text{stagnant}\}$ the transition matrix for this example i



■Transition probabilities

$$\mathbf{P} = \begin{array}{c} \\ Bull \\ Bear \\ Stagnant \end{array} \begin{pmatrix} Bull & Bear & stagnant \\ 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{pmatrix} \quad (12)$$

The distribution over states can be written as a stochastic row vector x with the relation $x^{(n+1)} = x^{(n)}P$. So if at time n the system is in state $x^{(n)}$, then three time periods later, at time $n+3$ the distribution is

$$x^{n+3} = x^{n+2}P = x^{n+1}P^2 = x^nP^3 \quad (13)$$

If at time n the system is in state 2 (bear), then at time n + 3 the distribution is

■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

$$
\begin{aligned}
&P^m(x,y) = Pr\left(\theta^{(m)} = y|\theta^{(0)} = x\right) \\
&= \sum_{x_1} ... \sum_{x_{m-1}} Pr\left(\theta^{(m)} = y, \theta^{(m-1)} = x_{m-1}, ...\theta^{(1)} = x_1|\theta^{(0)} = x\right) \\
&= \sum_{x_1} ... \sum_{x_{m-1}} Pr\left(\theta^{(m)} = y, \theta^{(m-1)} = x_{m-1})...Pr(\theta^{(1)} = x_1|\theta^{(0)} = x\right) \\
&= \sum_{x_1} ... \sum_{x_{m-1}} P(x,x_1)P(x_1,x_2)..P(x_{m-1},y) \qquad (14)
\end{aligned}
$$

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

$$
\begin{aligned}
&P^{n+m}(x,y) \\
&= \sum_z Pr\left(\theta^{(n+m)} = y|\theta^{(n)} = z, \theta^{(0)} = x\right) Pr\left(\theta^{(n)} = z, \theta^{(0)} = x\right) \\
&= \sum_z P^n\left(x,z\right) P^m\left(z,y\right) \qquad (15)
\end{aligned}
$$

Transition probabilities

Equations 15 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 (marginal distribution describes the probability distribution of a single variable from a set of related variables, ignoring the other variables) 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

$$\pi^{(n)} = \left(\pi^{(n)}(x_1), ...\pi^{(n)}(x_r)\right) \tag{16}$$

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

$$\begin{aligned} \pi^{(n)}(y) &= Pr(\theta^{(n)} = y) \\ &= \sum_{x \in S} Pr\left(\theta^{(n)} = y | \theta^{(0)} = x\right) Pr\left(\theta^{(0)} = x\right) \\ &= \sum_{x \in S} P^n(x, y) \pi^{(0)}(x) \end{aligned} \tag{17}$$

■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.

What is a Transition Probability Matrix?

- Describes the probabilities of transitioning from one state to another.
- Square matrix with non-negative entries.
- Each row sums to 1.

Example: Weather Forecasting

- States:
    - Sunny (S)
    - Rainy (R)
- Transition probabilities:
    - If today is sunny:
        - 70% chance of sunny tomorrow
        - 30% chance of rainy tomorrow
    - If today is rainy:
        - 40% chance of sunny tomorrow
        - 60% chance of rainy tomorrow

Transition Probability Matrix

The transition probability matrix is:

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

Where:

- $P_{11} = 0.7$: Sunny $\to$ Sunny
- $P_{12} = 0.3$: Sunny $\to$ Rainy
- $P_{21} = 0.4$: Rainy $\to$ Sunny
- $P_{22} = 0.6$: Rainy $\to$ Rainy

Probability After Two Days

If today is sunny, the probability distribution after two days is calculated as:

$$\text{Initial State} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

$$\text{After 2 days} = \begin{bmatrix} 1 & 0 \end{bmatrix} P^2$$

Where P^2 is the matrix product of P with itself.

Applications

- Weather prediction
- Customer behavior analysis
- Stock market analysis
- Biological systems (e.g., gene transitions)

Conclusion

- Transition probability matrices model dynamic systems.
- Useful for predicting state transitions over time.
- Widely used in various fields including science and economics.

■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 $\rho_{xy} = Pr_x\left(T_y < \infty\right)$ where T_y is the hitting time to y state;
(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:Recurrent and Transient state

Recurrent state: A state in a Markov chain where, if you start there, you are guaranteed to return to it at least once in the future (probability of returning $= 1$).

Transient state: A state in a Markov chain where, if you start there, there's a chance you might never return (probability of returning $<$ 1).

Analyzing these states helps understand the long-term behavior of Markov chains.

■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 recurrent 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. An irreducible Markov chain is one where all states communicate with each other. This means that, starting from any state in the chain, it's possible to eventually reach any other state with a non-zero probability of transition, no matter how many steps it may take.

■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.

Transition Matrix

Consider the transition matrix:

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.2 & 0.3 & 0.5 \\ 0 & 0 & 1 \end{bmatrix}$$

We analyze the recurrence and transience of each state.

Transition Matrix

Consider the transition matrix:

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.2 & 0.3 & 0.5 \\ 0 & 0 & 1 \end{bmatrix}$$

P Squared (P * P)

$$P^2 = P \cdot P = \begin{bmatrix} 0.35 & 0.4 & 0.25 \\ 0.16 & 0.25 & 0.59 \\ 0 & 0 & 1 \end{bmatrix}$$

This matrix represents the probabilities of transitioning between states in *two* steps.

P Cubed (P * P * P)

$$P^3 = P \cdot P \cdot P = \begin{bmatrix} 0.255 & 0.305 & 0.44 \\ 0.132 & 0.197 & 0.671 \\ 0 & 0 & 1 \end{bmatrix}$$

This matrix represents the probabilities of transitioning between states in *three* steps.

State 1 Analysis

Transition probabilities for State 1:

$$P_{11}^1 = 0.5, \quad P_{11}^2 = 0.35, \quad P_{11}^3 = 0.255, \ldots$$

General pattern:

$$P_{11}^n = 0.5 \times (0.7)^{n-1}$$

Summing the series:

$$\sum_{n=1}^{\infty} P_{11}^n = \frac{0.5}{1-0.7} = \frac{5}{3} \approx 1.67$$

Since the sum is finite, **State 1 is transient**.

State 2 Analysis

Transition probabilities for State 2:

$$P_{22}^1 = 0.3, \quad P_{22}^2 = 0.25, \quad P_{22}^3 = 0.197, \ldots$$

General pattern:

$$P_{22}^n = 0.3 \times (0.833)^{n-1}$$

Summing the series:

$$\sum_{n=1}^{\infty} P_{22}^n = \frac{0.3}{1 - 0.833} = \frac{0.3}{0.167} \approx 1.8$$

Since the sum is finite, **State 2 is transient**.

State 3 Analysis

Transition probabilities for State 3:

$$P_{33}^1 = 1, \quad P_{33}^2 = 1, \quad P_{33}^3 = 1, \ldots$$

General pattern:

$$P_{33}^n = 1 \text{ for all } n$$

Summing the series:

$$\sum_{n=1}^{\infty} P_{33}^n = \sum_{n=1}^{\infty} 1 = \infty$$

Since the sum is infinite, **State 3 is recurrent**.

Summary of Results

- **State 1:** Sum $= \frac{5}{3}$ (Finite) $\Rightarrow$ Transient
- **State 2:** Sum ≈ 1.8 (Finite) $\Rightarrow$ Transient
- **State 3:** Sum $= \infty$ (Infinite) $\Rightarrow$ Recurrent

Conclusion

State 1 and **State 2** are **Transient**. **State 3** is **Recurrent**. This is concluded from the expected number of visits being finite or infinite.

■Examples of Recurrent and Transient states

Example of Transient and Recurrent states

$P_{rx} \equiv P_{rxx} = \sum_{n=1}^{\infty} P_{rxx}^{(n)}$ = Prob. of recurrence to x.

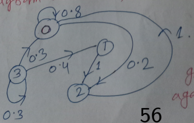
Then a state x is recurrent if $P_{rx} = 1$

and a . . . x . . . transient if $P_{rx} < 1$.

As discussed before, if $x \leftrightarrow y$ and x is recurrent $\Longrightarrow$ y recurrent.

Similarly, if $x \leftrightarrow y$ and x is transient $\Longrightarrow$ y transient.

Consider four states (say) Bhatbhateni, Bigmart, Saleways and daraz represented by states 0, 1, 2 and 3. The transition diagram looks like;



This means a person if shops in Bhatbhateni, the prob. that (s)he goes to bhatbhateni again is 80%.

■Examples of Recurrent and Transient states

(S) he goes to Saleways (probability) is 20% and so on.

$$\therefore P = \begin{array}{c|cccc} & BBSM & Bigmart & Saleways & Danaz \\ \hline BBSM & 0.8 & 0 & 0.2 & 0 \\ Bigmart & 0 & 0 & 1 & 0 \\ Saleways & 1 & 0 & 0 & 0 \\ Danaz & 0.3 & 0.4 & 0 & 0.3 \end{array}$$

To check whether Bhatbhateni is recurrent state.

For this find prob. $Pr_{xx} = Pr_x = \sum_{n=1}^{\infty} Pr_{xx}^{(n)}$

$\therefore$ here $Pr_{00}^{(1)} = 0.8$

$Pr_{00}^{(2)} = 0.2 \times 1 = 0.2$

$Pr_{00}^{(3)} = 0$

$\therefore Pr_{00} = Pr_0 = Pr_{00}^{(1)} + Pr_{00}^{(2)} + Pr_{00}^{(3)}$

$= 1.0$ $\therefore$ Bhatbhateni is recurrent state.

Consider Bigmart: $Pr_{11}^{(1)} = 0$

$Pr_{11}^{(2)} = 0$; $Pr_{11}^{(3)} = 0$

$\sum_{n=1}^{3} Pr_{11}^{(n)} = 0 \Rightarrow Pr_{11} = 0 < 1$ means transient state.

■Decomposition of the State Space -Example 4.6 (Textbook)

As, $0 \longleftrightarrow 2$ and 0 is recurrent
2 i.e. Saleways is also recurrent.

Daraz: $Pr_{33}^{(1)} = 0.3$; $Pr_{33}^{(2)} = 0 = Pr_{33}^{(3)}$

$\therefore Pr_{33} = 0.3 + 0 = 0.3 < 1$

$\Rightarrow$ 3 also transient state.

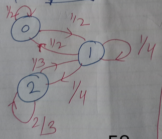
Now our aim is to decompose the state space into Recurrent and Transient states.

Consider example from Textbook (4.6)

Consider $S = \{0, 1, 2, \ldots, r\}$.

and $$P = \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & 1/2 & 1/2 & 0 \\ 1 & 1/2 & 1/4 & 1/4 \\ 2 & 0 & 1/3 & 2/3 \end{array}$$

Now represent the transition matrix by transition ~~proobab~~ state diagram.



Now represent the pair (x, y) by + & − such that
+ means $x \to y$
− means $x \not\to y$.

■Decomposition of the State Space -Example 4.6 (Textbook)

Then we get;

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix} \Rightarrow \begin{pmatrix} + & + & + \\ + & + & + \\ + & + & + \end{pmatrix}$$

This means all the states $S = \{0, 1, 2\}$ are recurrent. Hence $S = S_R$.

Now you prove that

$$(1, 0, 0)P^2 = ((1, 0, 0)P)P > 0.$$

Actually $0 \rightarrow 2$ after operating $(1, 0, 0)$ by P^2 and seeing that the prob. to state 2 is non zero.

$$\underset{\text{represents state } 0}{(1 \;\; 0 \;\; 0)} \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix} = (1/2 \;\; 1/2 \;\; 0)$$

Again; $(1/2 \;\; 1/2 \;\; 0) \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix} = \frac{1}{4}+\frac{1}{4} \quad \frac{1}{4}+\frac{1}{8} \quad \frac{1}{8}$

$= \left(\frac{1}{2}, \frac{3}{8}, \frac{1}{8}\right)$ means $0 \rightarrow 2$.

■Example 4.6(b)

(b)

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 & 0 \\ 0 & 1/5 & 2/5 & 1/5 & 0 & 1/5 \\ 0 & 0 & 0 & 1/6 & 1/3 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/4 & 0 & 3/4 \end{pmatrix} \tag{18}$$

CW: Draw the transition probability diagram as in previous case. Then write above matrix in terms of $+$ and $-$ as before.

$$\mathbf{P} = \begin{pmatrix} + & - & - & - & - & - \\ + & + & + & + & + & + \\ + & + & + & + & + & + \\ - & - & - & + & + & + \\ - & - & - & + & + & + \\ - & - & - & + & + & + \end{pmatrix} \tag{19}$$

■Example 4.6(b)

From above figures we get $S_R = \{0\} \cup \{3,4,5\}$ and $S_T = \{1,2\}$.

■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

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

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

$$\rho_{\{}0\}(x) = \rho_{x0} = \frac{\sum_{y=x}^{d-1} \gamma_y}{\sum_{y=0}^{d-1} \gamma_y}, \quad x = 1, \ldots d-1. \tag{20}$$

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

$$\sum_{x \in S} \pi(x) P(x, y) = \pi(y), \quad \forall y \in S \tag{21}$$

Equation 21 can be written in matrix form as

$$\pi P = \pi \tag{22}$$

The reason of the name is clear from the above equation. If the marginal distribution at any given step n is π 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

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \tag{23}$$

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

$$\pi(0)P(0, y) + \pi(1)P(1, y) = \pi(y), \quad y = 0, 1 \tag{24}$$

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

In [1]:
```
import numpy as np

def get_stationary(n):
    row = n
    pi = np.full((1, row), 1 / row)
    P = np.array([[1/4,1/2,1/4],
                  [1/3,0,2/3],
                  [1/2,0,1/2]])
    while True:
        new_pi = np.dot(pi, P)
        if np.allclose(pi, new_pi):
            return pi
            break
        pi = new_pi
print(get_stationary(3))
```

```
[[0.37500019 0.18750166 0.43749814]]
```

■Equilibrium Distributions

Let us consider a stochastic process at discrete times labeled consecutively $t_1, t_2, t_3, \ldots$, for a system with a finite set of possible states $S_1, S_2, S_3, \ldots.$, and we denote by X_t the state the system is in at time t. We consider the conditional probability that $X_{t_n} = S_{i_n}$,

$$P(X_{t_n} = S_{i_n} | X_{t_{n-1}} = S_{i_{n-1}}, X_{t_{n-2}} = S_{i_{n-2}}, \ldots., X_{t_1} = S_{i_1}) \qquad (25)$$

given that at the preceding time the system state $X_{t_{n-1}}$ was in state $S_{i_{n-1}}$, etc. Since this process is Markov process the conditional probability is in fact independent of all states but the immediate predecessor i.e. $P = P(X_{t_n} = S_{i_n} | X_{t_{n-1}} = S_{i_{n-1}}$.

Above conditional probability can be interpreted as the transition probability to move from state i to state j,

$$P_{ij} = P(S_i \rightarrow S_j) = P(X_{t_n} = S_j | X_{t_{n-1}} = S_i) \qquad (26)$$

■Equilibrium Distributions and Master Equation

We further require that $P_{ij} \geq 0, \quad \sum_j P_{ij} = 1$ as usual for the transition probabilities. We may then construct the total probability $P(X_{t_n} = S_j)$ that at time t_n the system is in state S_j as

$$\begin{aligned} P(X_{t_n} = S_j) &= P(X_{t_n} = S_j | X_{t_{n-1}} = S_i) P(X_{t_{n-1}} = S_i) \\ &= P_{ij} P(X_{t_{n-1}} = S_i) \end{aligned} \tag{27}$$

The master equation considers the change of this probability with time t (treating time as continuous rather than discrete variable and writing then $P(X_{t_n} = S_j) = P(S_j, t)$)

■Equilibrium Distributions and Master Equation

$$\frac{dP(S_j,t)}{dt} = -\sum_i P_{ij}P(S_j,t) + \sum_i P_{ij}P(S_i,t) \tag{28}$$

This equation also represent continuity of probability as it says total probability $\sum_j P(S_j,t) = 1$ at all times. All probability of a state i that is 'lost' by transition to state j is gained in the probability of that state, and vice versa.

Basic property of Markov process: knowledge of state at time t completely determines the future time evolution. The main significance of equation 28 is that the importance sampling Monte Carlo process can be interpreted as a Markov process, with a particular choice of transition probabilities.

■Equilibrium Distributions and Master Equation

Hence for equilibrium probability P_{eq} we get

$$P_{ji}P_{eq}(S_j) = P_{ij}P_{eq}(S_i) \tag{29}$$

called *detailed balance equation*. This means the master equation yields

$$\frac{dP_{eq}(S_j, t)}{dt} \equiv 0 \tag{30}$$

This equation ensures that gain and loss terms in equation 28 cancel exactly.

■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

$$\begin{aligned} P((i,j),(k,l)) &= Pr(\theta^{(n)} = (k,l)|\theta^{(n-1)} = (i,j)) \\ &= Pr(\theta_2^{(n)} = l|\theta_1^{(n)} = i)Pr(\theta_1^{(n)} = k|\theta_1^{(n)} = j) \\ &= \frac{\pi_{kl}}{\pi_{k+}}\frac{\pi_{kj}}{\pi_{+j}} \end{aligned} \quad (31)$$

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

- **Aperiodic 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

$$\bar{t}_n \to (a.s.) E_\pi[t(\theta)] \text{ as } n \to \infty \qquad (32)$$

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 π 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

$$\begin{aligned} P_n^*(x, y) &= Pr(\theta^{(n)} = y|\theta^{(n+1)} = x) \\ &= \frac{Pr(\theta^{(n+1)} = x|\theta^{(n)} = y)Pr(\theta^n = y)}{Pr(\theta^{(n+1)} = x} \\ &= \frac{\pi^{(n)}(y)P(y, x)}{\pi^{(n+1)}(x)} \end{aligned} \tag{33}$$

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

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad \text{for all} \quad x, y \in S \tag{34}$$

■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 34 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 34 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 34 to give 21. Construction of Markov chains with a given stationary distribution π reduces to finding transition probabilities $P(x, y)$ satisfying 34.

■Reversible Chains- Example : Metropolis Algorithm

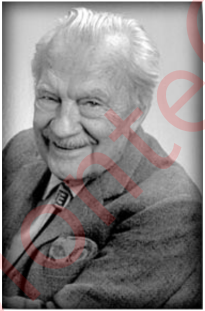


Figure: Metropolis was a Greek-American physicist. In Los Alamos he led the group in the Theoretical Division that designed and built the MANIAC I computer in 1952 that was modeled on the IAS machine, and the MANIAC II in 1957. At Los Alamos in the late 1940s and early 1950s a group of researchers led by Metropolis, including John von Neumann and Stanislaw Ulam, developed the Monte Carlo method.

■Reversible Chains- Example : Metropolis Algorithm

- Consider a given distribution $p_x,\ x \in\ S$ with $\sum_x p_x = 1$ where the state space S can be a subset of the line or even a d-dimensional subset of R^d. The problem posed and solved by Metropolis and coworkers in 1953 was how to construct a Markov chain with stationary distribution π such that $\pi(x) = p_x,\ x \in S$.

  Let Q be any irreducible transition matrix on S satisfying the symmetry condition $Q(x, y) = Q(y, x)$, for $x, y \in S$. Define a

  Markov chain $(\theta^(n))_n \geq 0$ as having transition from x to y proposed according to the probabilities Q(x, y). This proposed value for $(\theta^{(n+1)})$ is accepted with probability $min\{1, \frac{p_y}{p_x}\}$ and rejected otherwise, leaving the chain in state x.

■Reversible Chains- Example : Metropolis Algorithm

- The transition probabilities $P(x, y)$ of the above chain $(\theta^{(n)})_n \geq 0$ are

$$\begin{aligned} P(x,y) &= Pr(\theta^{(n+1)} = y, TA|\theta^{(n)} = x) \\ &= Pr(\theta^{(n+1)} = y, |\theta^{(n)} = x)Pr(TA) \\ &= Q(x,y)min\{1, \frac{p_y}{p_x}\} \end{aligned} \tag{35}$$

for $y \neq x$ and TA denotes the event [transition accepted.]

■Reversible Chains- Example : Metropolis Algorithm

- If $y = x$, then

$$P(x,x) = Pr(\theta^{(n+1)} = x, TA|\theta^{(n)} = x) + Pr(\theta^{(n+1)} \neq x, \bar{T}A|\theta^{(n)} = x) \tag{36}$$

$$= Pr(\theta^{(n+1)} = x, |\theta^{(n)} = x)Pr(TA) + \sum_{y \neq x} Pr(\theta^{(n+1)} = y, \bar{T}A|\theta^{(n)} = x)$$

$$= Q(x,x) + \sum_{y \neq x} Q(x,y)[1 - min\{1, p_y/p_x\}]$$

In the above case $y = x$, we need to consider state $y = x$ with prob(TA) as well as $y \neq x$ with prob $1 - prob(TA) \equiv prob(\bar{T}A)$.

■Reversible Chains- Example : Metropolis Algorithm

The first step to obtaining the stationary distribution of this chain is to prove that the probabilities p_x satisfy the reversibility condition. For $x = y$, equation

$$\pi(x)P(x, y) = \pi(y)P(y, x) \tag{37}$$

that is

$$p_x P(x, y) = p_y P(y, x) \tag{38}$$

for all $x, y \in S$ is trivially satisfied.

■Reversible Chains- Example : Metropolis Algorithm

For $y \neq x$, there will be two cases. (i) $p_y > p_x$ and case (ii) $p_x > p_y$.
Case (i) $p_y > p_x$:

$$p_x P(x,y) = p_x Q(x,y) = Q(y,x) min\{1, p_x/p_y\} p_y = p_y P(y,x) \tag{39}$$

In writing above equation, we used $p_x = min(1, \frac{p_x}{p_y}) * p_x$ as we are considering $p_y > p_x$ from $min(1, \frac{p_x}{p_y})$ one gets $\frac{p_x}{p_y}$ and hence $p_x = min(1, \frac{p_x}{p_y}) * p_x$.
Case (ii) can be followed analogously. Therefore the chain is reversible and the probabilities $p_x, x \in S$ provide the stationary distribution of the chain. If Q is aperiodic, so will be P and the stationary distribution is also the limiting distribution.

■ Example : Metropolis Algorithm

One can use Metropolis algorithm to generate "target distribution" say gaussian distribution from "initial (random) distribution".

The algorithm is simple.

- Initialization (Input):
- $randomsamples$: A list containing initial random values.
- mu : The mean of the target Gaussian distribution.
- $sigma$: The standard deviation of the target Gaussian distribution.
- $numsteps$: The number of iterations to run the Metropolis algorithm.
- Output: states: A list containing the final samples that approximate a Gaussian distribution.

Example : Metropolis Algorithm

- Create a copy of the $randomsamples$ list to avoid modifying the original data (store it in states).
- Iterate for $numsteps$: - Randomly select an index (index) from the states list.
- - Propose a new state by adding a random Gaussian noise ($deltax$) to the current state at the selected index (states[index]).
- - Calculate the acceptance probability based on the target Gaussian distribution:
- - Calculate the probability of the new state ($targetprobnew$) using its distance from the target mean (mu) and standard deviation (sigma).

■ Example : Metropolis Algorithm

- - Calculate the probability of the old state ($targetprobold$) using the same approach. - Divide $targetprobnew$ by $targetprobold$ to get the acceptance probability.
- - Generate a random number between 0 and 1.
- - If the random number is less than the acceptance probability, accept the proposed state and update states[index] with the new value.
- Return the final list states containing the samples that resemble a Gaussian distribution.
- Plot initial and final distributions

Discuss the results from code (20240315MetropolisExample.ipynb) - initial distribution "uniform" and final distribution "gaussian". Also check supplementary lectures (20240315Metropolis.tex)

■Continuous state spaces

In continuous state spaces, sequences of random quantities that form a Markov chain in R but still retain a discrete parameter space T. There are a few changes required with respect to the discrete case but the main results of the previous sections are still valid. In particular, convergence to the limiting distribution, the ergodic theorem and the central limit theorem need basically technical changes in the conditions of the chain to hold.

■Continuous state spaces- transition kernels

Markov chains can be defined analogously as discrete case. If the conditional probabilities do not depend on the step n, the chain is homogeneous.
Then the transition kernel $P(x, A)$ is again used to define the chain. The analogy with the discrete case breaks when trying to consider $P(x, \{y\})$, which is always null in the continuous case and not useful in this context. Therefore, transition matrices cannot be constructed and transition kernels must be used instead.

■Continuous state spaces- transition kernels

However, given that $P(x, \cdot)$ defines a probability distribution, the notation $P(x, y)$ can be used as

$$P(x, y) = Pr(\theta^{(n+1)} \leq y | \theta^n = x) = Pr(\theta^{(1)} \leq y | \theta^0 = x) \tag{40}$$

for $x, y \in S$ when P is absolutely continuous with respect to y. Also associated with this conditional distribution, one can obtain the conditional density

$$p(x, y) = \frac{\partial P(x, y)}{\partial y} \tag{41}$$

for $x, y \in S$. This density can be used to define the transition kernel of the chain instead of $P(x, A)$. The state space S does not need to be the entire line. It can be any interval or collection of intervals for results below to hold.

■Continuous state spaces- transition kernels

The conditional transition probability over m steps is given by

$$P^m(x,y) = Pr(\theta^{(m+n)} \leq y|\theta^{(n)} = x) \;\; \text{for} \;\; x,y \in S \tag{42}$$

and the transition kernel over to steps is given by

$$P^m(x,y) = \frac{\partial P^m(x,y)}{\partial y} \;\; \text{for} \;\; x,y \in S \tag{43}$$

■Continuous state spaces- Stationary distribution

The stationary or invariant distribution $p(x, y)$ must satisfy

$$\pi(y) = \int_{-\infty}^{\infty} \pi(x)p(x, y)dx \tag{44}$$

which is continuous version of stationary distribution.