# 2

# *Stochastic Processes Background*

Perhaps the simplest version of a stochastic process is a Markov process. In this chapter we review some properties of Markov processes with discrete state space, often called Markov chains, and discrete or continuous time. We derive some basic properties of discrete time Markov chains, define continuous time Markov chains, and work out some relationships between the two kinds. Finally we describe some numerical approaches to matrix exponentiation and illustrate their importance in analyzing continuous time Markov chains.

## 2.1 Discrete-time Markov chains

### 2.1.1 Introduction to discrete Markov chains

**Definition.** A discrete time stochastic process $\{X_n\}_{n=0}^{\infty}$ is called a **Markov chain** if for all $n \geq 0$ and for all $i_0, i_1, ..., i_{n-1}, i, j$,

$$\Pr(X_{n+1} = j \,|\, X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = \Pr(X_{n+1} = j \,|\, X_n = i). \tag{2.1}$$

The Markov property encoded by the above formula says that given the full history of the process up to time step $n$, $X_1, \ldots, X_n$, we only need to know $X_n$ to define the distribution of $X_{n+1}$ — the random variable describing the state of the process at time $n+1$. Another way of stating the Markov property is to say that conditionally on $X_n$, $X_{n+1}$ is independent of $X_1, \ldots, X_{n-1}$ (see Exercise 1). Notice that sequences of independent random variables trivially satisfy the Markov property. On the other hand, Markov chains represent the simplest form of departure from the independence assumption.

We call $X_n$ a **homogeneous Markov chain** if $\Pr(X_{n+1} = j \,|\, X_n = i)$ is not a function of $n$ — $\Pr(X_{n+1} = j \,|\, X_n = i) = \Pr(X_{m+1} = j \,|\, X_m = i)$ for all $n$ and $m$ — and **inhomogeneous** otherwise.

Andrei Markov
1856 – 1922

The concept of chain dependence was first used by Quetelet (1846), and developed by Andrei Andreievich Markov to generalize the law of large numbers beyond independent random variables. Markov was a student of Chebyshev, and worked in the St. Petersburg University, where he succeeded Chebyshev as a teacher of probability. Markov was a member of the St. Petersburg academy, where he protested the honorary membership given to members of the royal family and other non-scientists. Markov's active role in political and social life earned him a nick name "militant academician."

For homogeneous Markov chains, we also define **1-step transition probabilities**

$$p_{ij} = \mathrm{P}(X_1 = j \,|\, X_0 = i), \quad \sum_j p_{ij} = 1, \; p_{ij} \geq 0 \text{ for all } i, j$$

and collect them into **transition probability matrix** $\mathbf{P} = \{p_{ij}\}$.

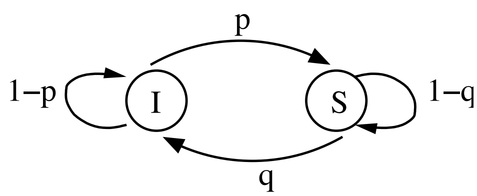


FIGURE 2.1: Transition graph of the SIS model.

**Example**: *SIS model* Suppose we observe an individual over a sequence of days $n = 1, 2, \ldots$ and classify this individual each day as

$$X_n = \begin{cases} I & \text{if infected,} \\ S & \text{susceptible.} \end{cases}$$

We would like to construct a stochastic model for the sequence $\{X_n\}_{n=1}^{\infty}$. One possibility is to assume that $X_n$s are independent and $\mathrm{P}(X_n = I) = 1 - \mathrm{P}(X_n = S) = p$. However, this model is not very realistic since we know from experience that the individual is more likely to stay infected if he or she is already infected. Since Markov chains are the simplest models that allow us to relax independence, we proceed by defining a transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}. \tag{2.2}$$

We can visualize transition probability matrices via transition graphs. In a transition graph, nodes represent elements of the state space. Directed edges designate one step transitions that occur with nonzero probabilities, which are depicted as edge labels. Figure 2.1 shows a transition graph corresponding to the transition probability matrix (2.2).

To illustrate how transition probabilities can be deduced from the procedural description of the Markov process, we consider three examples of Markov chains below.

**Example**: *Ehrenfest model of diffusion* Imagine a two dimensional rectangular box with a divider in the middle. The box contains $N$ balls (gas molecules) distributed somehow between the two halves. The divider has a small gap, through which balls can go through one at a time. We assume that at each time step we select a ball uniformly at random and force it to go through the gap to the opposite side of the divider. Let $X_n$ denote the total number of balls in the left half of the box.

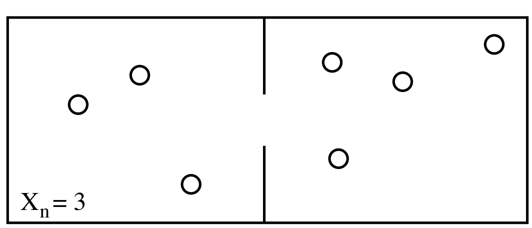


FIGURE 2.2: Illustration of the Ehrenfest diffusion model, where particles travel between two compartments by passing through a small opening in the wall dividing the compartments. $X_n$ keeps track of the number of particles in the left compratment.

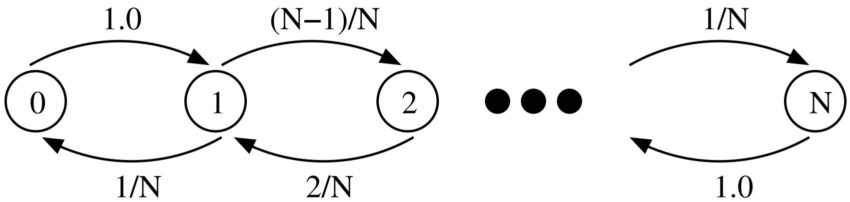


FIGURE 2.3: Transition graph of the Ehrenfest diffusion model.

Stochastic process $\{X_n\}$ is Markov by construction. Also by construction, $X_n$ can only go up or down by one unit. Moreover, since a ball is selected uniformly at random, we can use a simple counting argument to show that our Markov process is described by the following transition probabilities:

$$p_{ij} = \begin{cases} i/N, & \text{for } j = i-1, \\ 1 - i/N, & \text{for } j = i+1, \\ 0, \text{ otherwise.} \end{cases}$$

We show the transition graph of the Ehrenfest diffusion model in Figure 2.3.

**Example**: *Wright-Fisher model* Recall the Wright-Fisher model of genetic drift described in Section 1.1, where we describe the time evolution of $X_n$ — the number of $A$ alleles in a diploid population of $m$ individuals. According to our sampling with replacement construction, transition probabilities of the Wright-Fisher model are

$$p_{ij} = \Pr(X_{n+1} = j \,|\, X_n = i) = \binom{2m}{j} \left(\frac{i}{2m}\right)^j \left(1 - \frac{i}{2m}\right)^{2m-j}.$$

In other words

$$X_{n+1} \,|\, X_n \sim \text{Binom}\left(2m, \frac{X_n}{2m}\right).$$

We do not show the transition graph for the Wright-Fisher model, because it is fully connected, meaning that one can get with positive probability from any state to any other state in one step.

*Markov paths.*

Suppose we observe a finite realization of the discrete Markov chain and want to compute the probability of this random event:

$$\begin{aligned}
&\Pr(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \\
&= \Pr(X_n = i_n \,|\, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \Pr(X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \\
&= [\text{Markov property}] = p_{i_{n-1},i_n} \Pr(X_{n-1} = i_{n-1} \,|\, X_{n-2} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \\
&\times \Pr(X_{n-2} = i_{n-2}, \dots, X_1 = i_1, X_0 = i_0) = \cdots = \nu(i_0) p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{n-2},i_{n-1}} p_{i_{n-1},i_n},
\end{aligned}$$

where $\boldsymbol{\nu} = (\nu(1), \nu(2), \dots,)^T$ is the distribution of $X_0$, called the initial distribution of $\{X_n\}$.

*Note* 2.1. Every Markov chain is fully specified by its transition probability matrix $\mathbf{P}$ and initial distribution $\boldsymbol{\nu}$.

*Other consequences of the Markov property:*

1. The distribution of $X_n$ given a set of previous states depends only on the latest available state. We'll prove a restricted version of this statement: (see Exercise 1 (a) for a fuller statement)
$$\Pr(X_3 = i_3 \mid X_1 = i_1, X_0 = i_0) = \Pr(X_3 = i_3 \mid X_1 = i_1).$$
Notice that we need to prove this, because we are not conditioning on $X_2$.

   *Proof.*
$$\begin{aligned}
&\Pr(X_3 = i_3 \mid X_1 = i_1, X_0 = i_0) = [\text{law of total probability}] \\
&= \sum_{i_2} \Pr(X_3 = i_3, X_2 = i_2 \mid X_1 = i_1, X_0 = i_0) = [\text{def of cond prob}] \\
&= \sum_{i_2} \Pr(X_3 = i_3 \mid X_2 = i_2, X_1 = i_1, X_0 = i_0) \Pr(X_2 = i_2 \mid X_1 = i_1, X_0 = i_0) \\
&= [\text{Markov property}] = \sum_{i_2} \Pr(X_3 = i_3 \mid X_2 = i_2, X_1 = i_1) \Pr(X_2 = i_2 \mid X_1 = i_1) \\
&= [\text{def of cond prob}] = \sum_{i_2} \Pr(X_3 = i_3, X_2 = i_2 \mid X_1 = i_1) = \Pr(X_3 = i_3 \mid X_1 = i_1).
\end{aligned}$$
□

2. Past and future are independent given the present. Again, we'll prove a restricted version of this statement (and a fuller statement is in Exercise 1 (b))
$$\Pr(X_0 = i_0, X_3 = i_3, X_2 = i_2 \mid X_1 = i_1) = \Pr(X_0 = i_0 \mid X_1 = i_1) \Pr(X_3 = i_3, X_2 = i_2 \mid X_1 = i_1).$$

   *Proof.*
$$\begin{aligned}
&\Pr(X_0 = i_0, X_2 = i_2 \mid X_1 = i_1) = \frac{\Pr(X_0 = i_0, X_1 = i_1, X_2 = i_2, X_3 = i_3)}{\Pr(X_1 = i_1)} \\
&= \frac{\Pr(X_3 = i_3 \mid X_2 = i_2) \Pr(X_2 = i_2 \mid X_1 = i_1) \Pr(X_0 = i_0, X_1 = i_1)}{\Pr(X_1 = i_1)} \\
&= [\text{def of cond prob}] = \Pr(X_3 = i_3, X_2 = i_2 \mid X_1 = i_1) \Pr(X_0 = i_0 \mid X_1 = i_1).
\end{aligned}$$
□

**Definition.** We define ***n*-step transition probabilities** as $p_{ij}^{(n)} = \Pr(X_n = j \mid X_0 = i)$.

*Note* 2.2. For homogeneous Markov chains $p_{ij}^{(n)} = \Pr(X_{n+k} = j \mid X_k = i)$, i.e. $n$-step transition probabilities are invariant to a time shift.

*Proof.*
$$\begin{aligned}
&\Pr(X_{n+k} = j \mid X_k = i) = \sum_{i_{n+k-1},\ldots,i_{k+1}} \Pr(X_{n+k} = j, X_{n+k-1} = i_{n+k-1}, \ldots, X_{k+1} = i_{k+1} \mid X_k = i) \\
&= [\text{successive conditioning}] = \sum_{i_{n+k-1},\cdots,i_{k+1}} \Pr(X_{n+k} = j \mid X_{n+k-1} = i_{n+k-1}) \cdots \Pr(X_{k+1} = i_{k+1} \mid X_k = i) \\
&= [\text{def of homogeneity}] = \sum_{i_{n+k-1},\cdots,i_{k+1}} \Pr(X_n = j \mid X_{n-1} = i_{n+k-1}) \cdots \Pr(X_1 = i_{k+1} \mid X_0 = i) \\
&= \sum_{i_{n+k-1},\ldots,i_{k+1}} \Pr(X_n = j, X_{n-1} = i_{n+k-1}, \ldots, X_1 = i_{k+1} \mid X_0 = i) = \Pr(X_n = j \mid X_0 = i) = p_{ij}^{(n)}.
\end{aligned}$$
□

Sidney Chapman
1888 – 1970

British astro- and geoscientist Sidney Chapman coined the term "geomagnetism," studied thermal diffusion, the physics of the ozone layer and many other phenomena. Andrei Nikolaievich Kolmogorov was a Russian probabilist of the Moscow School, student of Luzin. He developed the axiomatization of probability theory, and made immense contributions to the theory of stochastic processes, turbulence, and complexity theory. The Chapman-Kolmogorov equation originated in work in economics by Bachelier and in physics by von Smoluchovski. The name is likely to have benn coined in the corridors of Stockholm University, where among others Will Feller and Harald Cramér worked, in the late 1930s.

Andrei Kolmogorov
1903 – 1987

*Chapman-Kolmogorov equation:*

$$p_{ij}^{(m+n)} = \sum_k p_{ik}^{(m)} p_{kj}^{(n)}.$$

*Proof.*

$$\begin{aligned} p_{ij}^{(m+n)} &= \Pr(X_{m+n} = j \mid X_0 = i) = [\text{law of total probability}] = \sum_k \Pr(X_{m+n} = j, X_m = k \mid X_0 = i) \\ &= \sum_k \Pr(X_{m+n} = j \mid X_m = k, X_0 = i) \Pr(X_m = k \mid X_0 = i) = [\text{Markov property consequence}] \\ &= \sum_k \Pr(X_{m+n} = j \mid X_m = k) \Pr(X_m = k \mid X_0 = i) = [\text{homogeneity}] \\ &= \sum_k \Pr(X_n = j \mid X_0 = k) \Pr(X_m = k \mid X_0 = i) = \sum_k p_{ik}^{(m)} p_{kj}^{(n)}. \end{aligned}$$

□

Setting $m$ and $n$ to 1 in turn, we arrive at Kolmogorov's

$$\text{forward equation} : p_{ij}^{(n+1)} = \sum_k p_{ik}^{(n)} p_{kj}, \text{ for } n = 1, 2, \ldots \text{ and}$$

$$\text{backward equation} : p_{ij}^{(n+1)} = \sum_k p_{ik} p_{kj}^{(n)}, \text{ for } n = 1, 2, \ldots.$$

*Note* 2.3. If we collect $n$-step transition probabilities into matrix $\mathbf{P}^{(n)} = \{p_{ij}^{(n)}\}$, then Kolmogorov's forward and backward equations can be rewritten in the matrix form as

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

where $\mathbf{P}^{(1)} = \mathbf{P}$. Therefore, $\mathbf{P}^{(n)} = \mathbf{P}^n$.

*Marginal distribution of $X_n$.*

$$\Pr(X_n = j) = \sum_i \Pr(X_n = j, X_0 = i) = \sum_i \Pr(X_n = j \mid X_0 = i) \Pr(X_0 = i) = \sum_i \nu(i) p_{ij}^{(n)}.$$

Defining $\nu_n(i) = \Pr(X_n = i)$ and $\boldsymbol{\nu}_n^T = (\nu_n(1), \nu_n(2), \ldots)$, we rewrite the above formula in the matrix form as

$$\boldsymbol{\nu}_n^T = \boldsymbol{\nu}^T \mathbf{P}^n.$$

**Example**: *Reducing SIS model to iid Bernoulli* Any two-state homogeneous Markov chain on state space $\Omega = \{0,1\}$ is defined by

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \quad \text{and} \quad \boldsymbol{\nu} = \begin{pmatrix} \nu(0) \\ \nu(1) \end{pmatrix}.$$

Consider a restriction of this model with $\boldsymbol{\nu}^T = (1-p, p)$ and $q = 1-p$. After one step, the distribution of the Markov chain becomes

$$\boldsymbol{\nu}_1^T = \boldsymbol{\nu}^T\mathbf{P} = (1-p,p)\begin{pmatrix} 1-p & p \\ 1-p & p \end{pmatrix} = ((1-p)^2 + p(1-p), p(1-p)+p^2) = (1-p,p) = \boldsymbol{\nu}^T.$$

At the $n$th step we get

$$\boldsymbol{\nu}_n^T = \boldsymbol{\nu}^T\mathbf{P}^n = \boldsymbol{\nu}^T\mathbf{P}\mathbf{P}^{n-1} = \boldsymbol{\nu}^T\mathbf{P}^{n-1} = \cdots = \boldsymbol{\nu}^T.$$

So marginally, $\Pr(X_n = 1) = p$ and $\Pr(X_n = 0) = 1-p$ for all $n = 1,2,\ldots$.

$$\begin{aligned}&\Pr(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \nu(i_0)p_{i_0i_1}p_{i_1i_2}\cdots p_{i_{n-1}i_n} \\ &= p^{i_0}(1-p)^{1-i_0}p^{i_1}(1-p)^{1-i_1}\cdots p^{i_n}(1-p)^{1-i_n} = \Pr(X_0 = i_0)\Pr(X_1 = i_1)\cdots\Pr(X_n = i_n).\end{aligned}$$

So we proved that $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$.

*Note* 2.4. In general, if all rows of a transition probability matrix $\mathbf{P}$ are equal to the initial distribution, then the corresponding Markov chain is a sequence of iid random variables distributed according to the probability mass function defined by the first row of $\mathbf{P}$.

### 2.1.2 First-step analysis and Markov chains with absorbing states

First-step analysis is a general strategy for solving many Markov chain problems by conditioning on the first step of the Markov chain. We demonstrate this technique on a simple example.

**Example**: *Gambler's ruin* Two players bet one dollar in each round. Player 1 wins with probability $p$ and looses with probability $q = 1-p$. We assume that player 1 starts with $a$ dollars and player 2 starts with $b$ dollars. Let $X_n$ be fortune of player 1 after $n$ rounds. $X_n$ can take values from 0 to $a+b$. First note that $p_{00} = p_{a+b,a+b} = 1$ since once one player has all the money there is no change any more, i.e., these are absorbing states. Everywhere else (when both players have money)

$$p_{ij} = \Pr(X_1 = j \mid X_0 = i) = \begin{cases} p & \text{if } j = i+1, \\ q & \text{if } j = i-1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $T$ be a random number of rounds after which one of the players looses all their money. We are interested in probability that player 1 wins the game, which corresponds to the event $X_T = a+b$. Let

$$u(a) = \Pr(X_T = a+b \mid X_0 = a)$$

for $a > 0$. Then first-step analysis proceeds by

$$\begin{aligned}&u(a) = \Pr(X_T = a+b \mid X_0 = a) = \sum_{j=0}^{a+b}\Pr(X_T = a+b, X_1 = j \mid X_0 = a) \\ &= \sum_{j=0}^{a+b}\Pr(X_T = a+b \mid X_0 = a, X_1 = j)\Pr(X_1 = j \mid X_0 = a) \\ &= \Pr(X_T = a+b \mid X_1 = a-1)\Pr(X_1 = a-1 \mid X_0 = a) \\ &+ \Pr(X_T = a+b \mid X_1 = a+1)\Pr(X_1 = a+1 \mid X_0 = a) = u(a-1)q + u(a+1)p.\end{aligned}$$

So we constructed a difference equation

$$u(i) = u(i+1)p + u(i-1)q \text{ for } 0 < i < a+b$$

with boundary conditions

$$\begin{aligned} u(0) &= \Pr(X_T = a+b \,|\, X_0 = 0) = 0 \text{ and} \\ u(a+b) &= \Pr(X_T = a+b \,|\, X_0 = a+b) = 1. \end{aligned}$$

There are several ways to solve this equation. We will proceed with a simple substitution which will lead us to a telescoping sum argument. First, we notice that

$$\begin{aligned} & u(i) = u(i+1)p + u(i-1)q \iff pu(i) + qu(i) = u(i+1)p + u(i-1)q \\ & \iff p[u(i+1) - u(i)] = q[u(i) - u(i-1)] \end{aligned}$$

Letting $v(i) = u(i) - u(i-1)$ leads to

$$pv(i+1) = qv(i) \Rightarrow v(i+1) = \frac{q}{p}v(i) = \left(\frac{q}{p}\right)^2 v(i-1) = \cdots \left(\frac{q}{p}\right)^i v(1) \text{ or } v(i) = \left(\frac{q}{p}\right)^{i-1} v(1).$$

$$u(i) = u(i) - u(0) = \sum_{j=1}^{i} [u(j) - u(j-1)] = \sum_{j=1}^{i} v(j) = v(1) \sum_{j=1}^{i} \left(\frac{q}{p}\right)^{j-1} = \begin{cases} v(1)i & \text{if } p = q \\ v(1)\frac{1-\left(\frac{q}{p}\right)^i}{1-\frac{q}{p}} & \text{if } p \neq q. \end{cases}$$

To find $v(1)$ we use the boundary conditions:

$$1 = u(a+b) - u(0) = \sum_{i=1}^{a+b} [u(i) - u(i-1)] = \sum_{i=1}^{a+b} v(i) = v(1) \sum_{i=1}^{a+b} \left(\frac{q}{p}\right)^{i-1} = \begin{cases} v(1)(a+b) & \text{if } p = q \\ v(1)\frac{1-\left(\frac{q}{p}\right)^{a+b}}{1-\frac{q}{p}} & \text{if } p \neq q. \end{cases}$$

Therefore,

$$v(1) = \begin{cases} \frac{1}{a+b} & \text{if } p = q, \\ \frac{1-\frac{q}{p}}{1-\left(\frac{q}{p}\right)^{a+b}} & \text{if } p \neq q. \end{cases} \Rightarrow u(i) = \begin{cases} \frac{i}{a+b} & \text{if } p = q, \\ \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^{a+b}} & \text{if } p \neq q. \end{cases}$$

#### 2.1.2.1 Markov chains with absorbing states

In the Gambler's ruin example above, we were interested in the probability of a Markov chain getting absorbed into one of its two absorbing states. Let's consider a generalization of this problem, where we have a Markov chain $\{X_n\}$ on a finite state space $\Omega = \{1, \ldots, n\}$. Some subset of the state space forms a set of all absorbing states. We will call the rest of the states transient. In this setting there are two interesting questions one may ask:

1. What is the probability of the Markov chain getting into the absorbing set via one particular absorbing state?
2. How long does the Markov chain stay in the transient states?

We first formalize these questions and then derive general Let

$$\begin{aligned} B &= \{1, \ldots, m\} \text{ — transient states} \\ A &= \{m+1, \ldots, n\} \text{ — absorbing states,} \end{aligned}$$

meaning that $p_{ij}=0$ for all $i \in A$ and $j \in B$ and for every $i \in B$ there exists $j \in A$ such that $i \to j$, but $i \not\leftarrow j$. Arranging the order of the states so that $\Omega = \{B, A\}$, we can write the transition probability matrix in a block matrix form:

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{S} \end{pmatrix},$$

where $\mathbf{Q}$ is an $m \times m$ matrix, $\mathbf{R}$ is an $m \times (n-m)$ matrix, and $\mathbf{S}$ is an $(n-m) \times (n-m)$ matrix.

*Hitting probabilities*

Let $h_{ij}$ be probability of starting at the transient state $i \in B$ and entering $A$ through $j \in A$. First-step analysis yields

$$h_{ij} = p_{ij} + \sum_{k=1}^{m} p_{ik} h_{kj}$$

or in matrix form:

$$\mathbf{H} = \mathbf{R} + \mathbf{QH} \Rightarrow \mathbf{H} = (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{R},$$

where $\mathbf{H} = \{h_{ij}\}$ is $m \times (n-m)$ matrix and $\mathbf{I}$ is an $m \times m$ identify matrix. This result relies on the fact that $(\mathbf{I} - \mathbf{Q})^{-1}$ exists.

**Proposition 2.1.** *Let* $\mathbf{Q}$ *be a square matrix with* $\lim_{k\to\infty} \mathbf{Q}^k = \mathbf{0}$. *Then* $(\mathbf{I}-\mathbf{Q})^{-1}$ *exists and* $(\mathbf{I}-\mathbf{Q})^{-1} = \sum_{k=0}^{\infty} \mathbf{Q}^k$.

*Proof.* We start with

$$(\mathbf{I}-\mathbf{Q})(\mathbf{I}+\mathbf{Q}+\cdots+\mathbf{Q}^{k-1}) = \mathbf{I}-\mathbf{Q}^k. \tag{2.3}$$

Taking determinant of both sides, we get

$$\det(\mathbf{I}-\mathbf{Q}) \times \det(\mathbf{I}+\mathbf{Q}+\cdots+\mathbf{Q}^{k-1}) = \det(\mathbf{I}-\mathbf{Q}^k). \tag{2.4}$$

Since determinant is a continuous function,

$$\lim_{k\to\infty} \det(\mathbf{I}-\mathbf{Q}^k) = \det[\lim_{k\to\infty}(\mathbf{I}-\mathbf{Q}^k)] = \det(\mathbf{I}-\mathbf{0}) = 1.$$

Therefore, there exists $k$ such that $\det(\mathbf{I}-\mathbf{Q}^k) > 0$, which together with equation (2.4) implies that $\det(\mathbf{I}-\mathbf{Q}) \neq 0 \Rightarrow (\mathbf{I}-\mathbf{Q})^{-1}$ exists.

Moreover, sending $k$ to $\infty$ in equation (2.3), we obtain

$$\sum_{k=0}^{\infty} \mathbf{Q}^k = (\mathbf{I}-\mathbf{Q})^{-1}.$$

□

To apply the above proposition, we need to show that for $\mathbf{Q}$ that we defined for our absorbing Markov chain, $\mathbf{Q}^k \to \mathbf{0}$. Notice that the of the transition probability matrix leads to

$$\mathbf{P}^k = \begin{pmatrix} \mathbf{Q}^k & ? \\ \mathbf{0} & \mathbf{S}^k \end{pmatrix},$$

which shows that $p_{ij}^{(k)} = q_{ij}^{(k)}$, where $\mathbf{Q}^k = \{q_{ij}^{(k)}\}$ and $i, j \in B$. Recall that $\lim_{k\to\infty} p_{ij}^{(k)} = 0$ when $j$ is transient $\Rightarrow \lim_{k\to\infty} q_{ij}^{(k)} = 0 \Rightarrow \lim_{k\to\infty} \mathbf{Q}^k = \mathbf{0}$. To summarize, hitting probabilities can be legitimately computed as

$$\mathbf{H} = (\mathbf{I}-\mathbf{Q})^{-1}\mathbf{R},$$

**Definition.** $\sum_{k=0}^{\infty} \mathbf{Q}^k$ is called a **fundamental matrix of the absorbing Markov chain**.

*Hitting times*

Let $t_{ij}$ be the expected number of visits to transient state $j$ starting from transient state $i$:

$$t_{ij} = E_i\left(\sum_{n=0}^{\infty} 1_{\{X_n=j\}}\right) \text{ for } i,j \in B.$$

First-step analysis yields

$$t_{ij} = 1_{\{i=j\}} + \sum_{k=1}^{m} p_{ik}t_{kj} = 1_{\{i=j\}} + \sum_{k=1}^{m} q_{ik}t_{kj}$$

or in matrix form:

$$\mathbf{T} = \mathbf{I} + \mathbf{Q}\mathbf{T} \Rightarrow \mathbf{T} = (\mathbf{I}-\mathbf{Q})^{-1},$$

where $\mathbf{T} = \{t_{ij}\}$.

Let $t_i = \sum_{j=1}^{m} t_{ij}$ be the mean number of steps it takes $X_n$ to get absorbed starting from state $i$ and $\mathbf{t}^T = (t_1,\ldots,t_m)$. Then

$$\mathbf{t} = \mathbf{T}\mathbf{1} = (\mathbf{I}-\mathbf{Q})^{-1}\mathbf{1},$$

where $\mathbf{1}^T = (1,\ldots,1)$.

Another interesting object to study is $f_{ij}$ - probability of starting at transient state $i \in B$ and ever visiting transient state $j$ before absorption. Applying the law of total expectation, we get

$$t_{ij} = f_{ij}t_{jj} + (1-f_{ij}) \times 0,$$

which leads to $f_{ij} = t_{ij}/t_{jj}$.

**Example**: *Monitoring deaths in SIS model* Suppose an individual in a population has four possible states $S$ - susceptible, $I$ - infectious, $D_I$ - dead from disease complications, $D_S$ - dead from other causes. The individual moves between these state space $\Omega = \{S, I, D_S, D_I\}$ according to a Markov chain with the following transition probability matrix:

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & 0 \\ p_{21} & p_{22} & 0 & p_{24} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

So in our absorbing Markov chain notation

$$\mathbf{Q} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}, \qquad \mathbf{R} = \begin{pmatrix} p_{13} & 0 \\ 0 & p_{24} \end{pmatrix}, \qquad \mathbf{S} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The fundamental matrix is

$$(\mathbf{I}-\mathbf{Q})^{-1} = \frac{1}{(1-p_{11})(1-p_{22}) - p_{12}p_{21}} \begin{pmatrix} 1-p_{22} & p_{12} \\ p_{21} & 1-p_{11} \end{pmatrix}.$$

So the hitting probabilities can be computed as

$$\mathbf{H} = (\mathbf{I}-\mathbf{Q})^{-1} \begin{pmatrix} p_{13} & 0 \\ 0 & p_{24} \end{pmatrix}$$

and hitting times as $\mathbf{T} = (\mathbf{I}-\mathbf{Q})^{-1}$. The transient visit probabilities are collected in the matrix

$$\mathbf{F} = \mathbf{T}\left[\text{diag}(\mathbf{T})\right]^{-1} = \begin{pmatrix} 1 & \frac{t_{12}}{t_{22}} \\ \frac{t_{21}}{t_{11}} & 1 \end{pmatrix}.$$

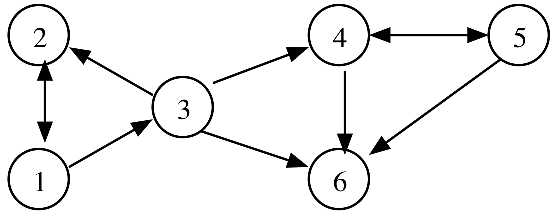


FIGURE 2.4: Transition graph of a Markov chain, with communication classes $\{1,2,3\}$, $\{4,5\}$, and $\{6\}$.

### 2.1.3 State communication, recurrence, and limiting behavior of Markov chains

Now that we have seen how to study Markov chains with absorbing states, we would like to understand how Markov chains behave when there are no absorbing states. In particular, it is often of interest to understand if any pattern of behavior emerges when the Markov chain is running for "a long time." It turns out that concepts of state communication (irreducibility) and recurrence are important ingredients to study long term behavior of Markov chains. We start by defining accessibility and communication.

**Definition.** For a pair of states $i$ and $j$, we say that $j$ is **accessible** from $i$ ($i \rightarrow j$) if there exists $m \geq 0$ such that $p_{ij}^m > 0$. We say that $i$ **communicates** with $j$ ($i \leftrightarrow j$) if $j$ is accessible from $i$ and $i$ is accessible from $j$.

Since communication is a binary relationship between Markov chain states, it is useful to recall that some binary relationships allow us to combine objects into "classes."

**Definition. Equivalence relation** "$\sim$" is a binary relation between elements of a set satisfying

1. reflexivity: $i \odot i$ for all $i$
2. symmetry: $i \odot j \Rightarrow j \odot i$
3. transitivity: $i \odot j$, $j \odot k \Rightarrow i \odot k$.

**Definition.** For a set $\mathscr{S}$ and $a \in \mathscr{S}$, $\{s \in \mathscr{S} : s \sim a\}$ is called an equivalence class.

Equivalence relations allow us to split Markov chain state spaces into equivalence classes.

**Proposition 2.2.** *Markov chain state communication is an equivalence relation.*

*Proof.* Reflexivity and symmetry follow immediately from the definition of communication. Transitivity needs a proof. Let $i \leftrightarrow j$ and $j \leftrightarrow k$. We want to show that $i \leftrightarrow k$. $i \rightarrow j \Rightarrow$ there exists a path $i, i_1, \ldots, i_{M-1}, j$ such that $p_{ii_1} p_{i_1 i_2} \cdots p_{i_{M-1} j} > 0$. Similarly, $j \rightarrow k \Rightarrow$ there exists a path $j, j_1, \ldots, j_{M-1}, k$ such that $p_{jj_1} p_{j_1 j_2} \cdots p_{j_{M-1} k} > 0$. Therefore, $p_{ii_1} p_{i_1 i_2} \cdots p_{i_{M-1} j} p_{jj_1} p_{j_1 j_2} \cdots p_{j_{M-1} k} > 0$, which implies that

$$p_{ik}^{M+N} = \sum_{l_1, \ldots, l_{M+N-1}} p_{il_1} p_{l_1 l_2} \cdots p_{l_{M+N-1} k} > 0.$$

Hence $i \rightarrow k$. Similarly, we can show that $k \rightarrow i$ and conclude that $i \leftrightarrow k$. $\square$

**Definition.** $\{X_n\}$ is called **irreducible** if it has only one communication class, i.e. $\forall i, j\ i \leftrightarrow j$.

In other words, if $\{X_n\}$ is an irreducible Markov chain, then all its states communicate with each other.

Now that we know what it means for states to communicate, we would like to study "returning" behavior of Markov chains. Intuitively, such behavior is important, because we can hope to find a pattern of a Markov chain limiting behavior only if this Markov chain keeps visiting each its state infinitely many times. Two important random variables that will be helpful to us are return times to a particular state and number of visits to a particular state.

**Definition.** The return time to state $i$ is defined as $T_i = \inf\{n \geq 1 : X_n = i\}$

**Definition.** Let $N_i = \sum_{n=1}^{\infty} 1_{\{X_n = i\}}$ be the number of visits of $\{X_n\}$ to state $i$, not counting the initial step.

*Note* 2.5. We'll start using the following notation:

$$\Pr(\cdot \,|\, X_0 = i) = P_i(\cdot),$$
$$\mathrm{E}(\cdot \,|\, X_0 = i) = E_i(\cdot).$$

We list the following relationships between $T_i$ and $N_i$. Some of these relationships follow directly from the definition of these random variables, but some require proofs that we leave as exercises to the reader.

1. $\{N_i = 0\} = \{T_i = \infty\} \qquad \{N_i > 0\} = \{T_i < \infty\}$
2. $P_j(N_i = r) = \begin{cases} f_{ji} f_{ii}^{r-1}(1 - f_{ii}) & \text{if } r \geq 1, \\ 1 - f_{ji} & \text{if } r = 0, \end{cases}$
   where $f_{ji} = P_j(T_i < \infty)$ is the probability of reaching $i$ in finite time starting from state $j$.
3. Setting $i = j$ in the previous formula we obtain
   $$P_i(N_i = r) = f_{ii}^{r}(1 - f_{ii}),$$
   $$P_i(N_i > r) = f_{ii}^{r+1}.$$
4. $P_i(T_i < \infty) = 1 \Leftrightarrow P_i(N_i = \infty) = 1$ and $P_i(T_i < \infty) < 1 \Leftrightarrow P_i(N_i = \infty) = 0 \Leftrightarrow E_i(N_i) < \infty$.
5. $P_i(N_i = \infty) = 0$ or $P_i(N_i = \infty) = 1$, nothing in between.
6. $\{T_i < \infty\} \neq \{N_i = \infty\}$, but $P_i(T_i < \infty) = 1 \iff P_i(N_i = \infty) = 1$. In other words, if a Markov chain returns to state $i$ in finite time, then the chain visits this state infinitely often.

**Definition.** State $i$ is called **recurrent** if $P_i(T_i < \infty) = f_{ii} = 1$ and **transient** otherwise.

**Definition.** Recurrent state $i$ is called **positive recurrent** if $E_i(T_i) < \infty$ and **null recurrent** otherwise.

**Proposition 2.3.** *State i is recurrent* $\iff \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.

*Proof.* First, notice that

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \sum_{n=1}^{\infty} P_i(X_n = i) = \sum_{n=1}^{\infty} E_i(1_{\{X_n = i\}}) = E_i(N_i).$$

Applying what we know about the connection between $T_i$ and $N_i$, we get

$$i \text{ is recurrent} \Leftrightarrow P_i(T_i < \infty) = 1 \Leftrightarrow P_i(N_i = \infty) = 1 \Leftrightarrow E_i(N_i) = \infty \Leftrightarrow \sum_{i=1}^{\infty} p_{ii}^{(n)} = \infty.$$

$\square$

**Proposition 2.4.** *Recurrence is a communication class property, i.e. if $i \leftrightarrow j$ and $i$ is recurrent, then $j$ is recurrent.*

*Proof.* Homework exercise. □

**Example**: *Gambler's ruin* Recall that in the gambler's ruin problem, 0 and $a+b$ states are absorbing. Therefore, $\sum_{n=1}^{\infty} p_{00}^{(n)} = \sum_{n=1}^{\infty} p_{a+b,a+b}^{(n)} = \sum_{n=1}^{\infty} 1 = \infty$. Hence, 0 and $a+b$ are recurrent states. Consider state 1:

$$P_1(T_1 < \infty) = 1 - P_1(T_1 = \infty) \leq 1 - q < 1 \text{ if } q \in (0,1).$$

Therefore, by definition 1 is a transient state. Since states $\{1,\ldots,c-1\}$ form a communication class, all states in this class are also transient.

**Example**: *1-D random walk* Let $X_n$ be a random walk on the set of all integers $\mathbb{Z}$ such that

$$p_{ij} = \begin{cases} p & \text{if } j = i+1, \\ q & \text{if } j = i-1, \end{cases}$$

where $p+q=1$. Let's study recurrence of state 0. We know that $p_{00}^{(2n+1)} = 0 \; \forall \; n \geq 0$. We also know that $X_{2n} \,|\, X_0 = 0 = \xi_1 + \cdots \xi_{2n}$, where $\xi_1,\ldots,\xi_n$ are iid with $\Pr(\xi_i = 1) = 1 - \Pr(\xi = -1) = p$. So

$$p_{00}^{(2n)} = \Pr(X_{2n} = 0 \,|\, X_0 = 0) = \binom{2n}{n} p^n q^n.$$

Recall that Stirlings formula says that $n! = O(n^{n+\frac{1}{2}} e^{-n})$.

Therefore,

$$p_{00}^{(2n)} = \frac{2n!}{n!n!} p^n q^n = O\left(\frac{(2n)^{2n+\frac{1}{2}} e^{-2n}}{n^{2n+1} e^{-2n}} (pq)^n\right) = O\left(\frac{2^{2n+\frac{1}{2}} n^{2n+\frac{1}{2}}}{n^{2n+1} 2^{\frac{1}{2}}} (pq)^n\right) = O\left(\frac{(pq)^n 2^{2n}}{\sqrt{n}}\right) = O\left(\frac{(4pq)^n}{\sqrt{n}}\right).$$

Consequently,

$$\sum_{n=1}^{\infty} p_{00}^{(n)} = \sum_{n=1}^{\infty} p_{00}^{(2n)} = \infty \Leftrightarrow 4pq \geq 1 \Leftrightarrow p = q = \frac{1}{2}.$$

So only a *symmetric* random walk is recurrent on $\mathbb{Z}$. Interestingly, a symmetric random walk on $\mathbb{Z}^2$ is also recurrent, but it is transient on $\mathbb{Z}^n$ for $n \geq 3$.

**Definition.** A vector $\mathbf{x} \neq \mathbf{0}$ is called an **invariant measure** of the stochastic matrix $\mathbf{P}$ if

$$x_i \in [0,\infty) \; \forall i \quad \text{and} \quad \mathbf{x}^T \mathbf{P} = \mathbf{x}^T \left(\forall \; i \; x_i = \sum_j x_j p_{ji}\right)$$

**Definition.** A probability vector $\boldsymbol{\pi}$ on the Markov chain state space is called a **stationary distribution** if $\boldsymbol{\pi}^T \mathbf{P} = \boldsymbol{\pi}^T$ or equivalently if $\pi(i) = \sum_j \pi(j) p_{ji}$ for all $i$. This system of linear equations is called **global balance**.

*Note* 2.6. If $\mathbf{x}$ is an invariant measure and $c = \sum_i x_i < \infty$, then $c^{-1}\mathbf{x}$ is a stationary distribution. However, $c = \infty$ for some invariant measures, so one can not normalize them.

**Example**: *Two state Markov chain*

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \quad 0 < p < 1, \quad 0 < q < 1.$$

The global balance equation is

$$\big(\pi(1), \pi(2)\big) \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} = \big(\pi(1), \pi(2)\big) \text{ or}$$

$$\begin{cases} (1-p)\pi(1) + q\pi(2) & = \pi(1) \\ p\pi(1) + (1-q)\pi(2) & = \pi(2) \end{cases} \Rightarrow p\pi(1) = q\pi(2) \Rightarrow \pi(1) = \frac{q}{p}\pi(2)$$

Using the fact that $\pi(1) + \pi(2) = 1$, we obtain

$$\frac{q}{p}\pi(2) + \pi(2) = 1 \Rightarrow \pi(2) = \frac{p}{p+q}$$

Therefore

$$\boldsymbol{\pi}^T = \left(\frac{q}{p+q}, \frac{p}{p+q}\right),$$

and this solution of the global balance equations is unique.

**Example**: *Gambler's ruin* Let the total fortune of both players be $a+b=4$. Then

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

The vectors $\boldsymbol{\pi}_\alpha = (\alpha, 0, 0, 0, 1-\alpha)$ satisfy global balance: $\boldsymbol{\pi}_\alpha^T \mathbf{P} = \boldsymbol{\pi}_\alpha^T$ for any $\alpha \in [0,1]$. So the gambler's ruin chain has an uncountable number of stationary distributions.

We would like to answer the following questions:

1. Does every Markov chain have an invariant measure?
2. Does every Markov chain have a stationary distribution?
3. Under what conditions does a Markov chain have a unique stationary distribution?

It turns out that all these questions have to do with irreducibility and recurrence of Markov chain states.

**Theorem 2.1.** *(***Stationary distribution criterion***) An irreducible and homogeneous Markov chain is positive recurrent if and only if there exists a stationary distribution. Moreover, if stationary distribution $\boldsymbol{\pi}^T = (\pi_1, \pi_2, \dots)$ exists, it is unique and $\pi_i > 0$ for all $i \in \Omega$.*

*Note* 2.7. An irreducible Markov chain can have an invariant measure and be transient or null recurrent. For example, we showed that a 1-D random walk on all integers $\mathbb{Z}$ with $p \neq q$ is transient and recurrent if $p = q = 0.5$. However, this Markov chain has an invariant measure $\mathbf{y}^T = (1, 1, \dots)$ for any $p$ and $q$. Since this measure is not normalizable, 1-D random walk can not be positive recurrent.

**Theorem 2.2.** *(***Mean return time***) If $\{X_n\}$ is homogeneous, irreducible, and positive recurrent, then*

$$\pi_i = \frac{1}{E_i(T_i)},$$

*where $\boldsymbol{\pi}^T = (\pi_1, \pi_2, \dots)$ is the stationary distribution of $\{X_n\}$ and $T_i = \inf\{n \geq 1 : X_n = i\}$ is the return time to state i.*

**Theorem 2.3.** *An irreducible homogeneous Markov chain on a finite state space is positive recurrent.*

*Proof.* Homework exercise. □

In summary, the main results are

1. Irreducibility + *positive* recurrence $\Leftrightarrow$ irreducibility + $\exists$ a stationary distribution $\boldsymbol{\pi}$ and it is unique. Moreover, when $\boldsymbol{\pi}$ exists, $\pi_i > 0$ and $\pi_i = 1/E_i(T_i)$.
2. Irreducibility + finite state space $\Rightarrow$ positive recurrence.

**Theorem 2.4.** *(***Ergodic theorem***) Let $\{X_n\}$ be an irreducible, homogeneous, and positive recurrent Markov chain on state space $\Omega$ with stationary distribution $\boldsymbol{\pi}$. Let $f : \Omega \to \mathbb{R}$ such that $\sum_{i\in\Omega} |f(i)|\pi_i < \infty$. Then for any initial distribution*

$$\frac{1}{N}\sum_{k=1}^{N} f(X_k) \overset{a.s.}{\to} \sum_{i\in\Omega} f(i)\pi_i \text{ when } N \to \infty.$$

George D. Birkhoff
1884-1944

George Birkhoff was the leading US mathematician in the first third of the twentieth century. In 1931 he proved the ergodic theorem, tying together time and (phase) space averages in statistical physics. The first theorem directly applicable to Markov chains of the kind we call the ergodic theorem in this book was, we think, published by the Stanford probabilist Kai-Lai Chung in 1950.

Kai-Lai Chung
1917 – 2009

**Example**: Suppose we are interested in how many times on average an irreducible and positive recurrent Markov chain $\{X_n\}$ visits a certain subset $A \subset E$. Setting $f = 1_{\{X_n \in A\}}$ and applying the ergodic theorem we have

$$\frac{1}{N}\sum_{k=1}^{N} 1_{\{X_n\in A\}} \overset{\text{a.s.}}{\to} \sum_{i\in\Omega} 1_{\{i\in A\}}\pi_i = \sum_{i\in A}\pi_i \text{ as } N \to \infty.$$

**Example**: *Limiting behavior of Ehrenfest diffusion* Recall our Ehrenfest diffusion Markov chain $X_n$ that tracks the number of gas molecules in the left half of the box. Let's apply the ergodic theorem to this model. We start with $N = 100$ gas molecules. The chain is irreducible and positive recurrent (do you understand why?), so we can apply the ergodic theorem. We start with $X_1 = 90$ and run the chain for $10^4$ time steps and record the number of molecules at each step. First, we plot the frequency of visiting each state in the right panel of Figure 2.5. This histogram has a bell-like shape, suggesting that if scaled appropriately the stationary distribution of the Ehrenfest diffusion can be approximated

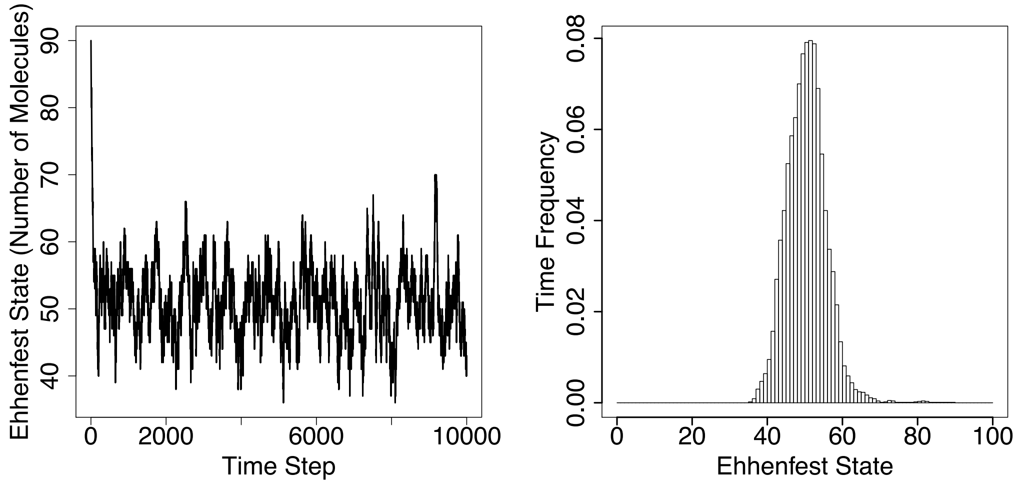


FIGURE 2.5: Results of Ehrenfest diffusion simulation. The left panel shows the states of the system (number of gas molecules) across time steps ranging from 1 to $10^4$. The right panel shows the histogram of the number of molecules corresponding to time frequencies of visiting the Ehrenfest model states.

by a Gaussian distribution. We can use the ergodic theorem to approximate various properties of the stationary distribution. For example, if we define a random variable $Y$ that follows the stationary distribution of the Ehrenfest diffusion Markov chain, then

$$\Pr(Y=43) = \mathrm{E}\left(1_{\{Y=43\}}\right) \approx \frac{1}{10^4}\sum_{i=1}^{10^4} 1_{\{X_i=43\}} = 0.03.$$

Notice that we have used $f(x) = 1_{\{x=43\}}$ to apply the ergodic theorem here. If we use $f(x) = x$, we can use the ergodic theorem to approximate the first moment of the stationary distribution:

$$\mathrm{E}(Y) \approx \frac{1}{10^4}\sum_{i=1}^{10^4} X_i = 51.$$

Finally, if we take $f(x) = x^2$, we can as easily approximate the second moment:

$$\mathrm{E}\left(Y^2\right) \approx \frac{1}{10^4}\sum_{i=1}^{10^4} X_i^2 = 2630.$$

Notice that by computing the first and second moment, we can find an approximation for the variance, even thought the ergodic theorem does give us a recipe to compute the stationary variance. More specifically, $\mathrm{Var}(Y) = \mathrm{E}\left(Y^2\right) - [\mathrm{E}(Y)]^2 \approx 2630 - 51^2 = 28$.

#### 2.1.3.1 Reversibility

**Definition.** A probability vector $\boldsymbol{\pi}$ on $\Omega$ is said to satisfy **detailed balance** with respect to stochastic matrix **P** if

$$\pi_i p_{ij} = \pi_j p_{ji} \text{ for all } i, j. \tag{2.5}$$

**Proposition 2.5.** *(Detailed balance $\Rightarrow$ global balance) Let* $\mathbf{P}$ *be a transition probability matrix of* $\{X_n\}$ *on* $\Omega$ *and let* $\boldsymbol{\pi}$ *be a probability distribution on* $\Omega$. *If* $\boldsymbol{\pi}$ *satisfies detailed balance, then* $\boldsymbol{\pi}$ *is a stationary distribution of* $\{X_n\}$.

*Proof.* $\pi_i p_{ij} = \pi_j p_{ji} \Rightarrow \sum_{j\in\Omega} \pi_i p_{ij} = \pi_i$. for all $i \in \Omega$. □

The main advantage of detailed balance is that it is easier to check it than the global balance for a candidate probability vector. We illustrate this using the example below.

**Example**: *Stationary distribution of Ehrenfest diffusion* Recall that the transition probabilities of the Ehrenfest model are

$$p_{ij} = \begin{cases} \frac{i}{N} & \text{if } j = i-1, \\ 1-\frac{i}{N} & \text{if } j = i+1. \end{cases}$$

$\{X_n\}$ is an irreducible Markov chain on the finite state space, so the chain is positive recurrent. Therefore, there exists a stationary distribution and it is unique. One way to find this stationary distribution is to solve the global balance equations $\boldsymbol{\pi}^T\mathbf{P} = \boldsymbol{\pi}^T$. Alternatively, we can try to "guess" the stationary distribution. After applying ergodic theorem to this model, we know that the stationary distribution is bell shaped, has mean of around $N/2$ and variance approximately half the size of the mean. First, a binomial distribution is a natural candidate since its probability mass function can be bell shaped for some parameters. Next, we know that the support of this binomial distribution should be $\{0,1,\ldots,N\}$, so the size parameter of the binomial should be number of particles $N$. Since we approximated the stationary mean to be around $N/2$, the success probability of the binomial distribution should be 0.5. So we conjecture that at equilibrium $X_n \sim \text{bin}(\frac{1}{2}, N)$. Let's try to verify this candidate stationary distribution via the detailed balance. Notice that we do not know whether the Ehrenfest chain is reversible, but we will go ahead with the detailed balance check anyway. Entries of our candidate probability vector are

$$\pi_i = \binom{N}{i}\left(\frac{1}{2}\right)^i\left(1-\frac{1}{2}\right)^{N-i} = \binom{N}{i}\frac{1}{2^N} \tag{2.6}$$

Since $X_n$ can only increase or decrease by one at each time step, we need to check detailed balance only for $i$ and $j = i+1$.

$$\pi_i p_{i,i+1} = \frac{1}{2^N}\binom{N}{i}\frac{N-i}{N} = \frac{1}{2^N}\frac{N!}{i!(N-i)!}\frac{N-i}{N} = \frac{1}{2^N}\frac{N!}{(i+1)!(N-i-1)!}\frac{i+1}{N}$$
$$= \binom{N}{i+1}\frac{1}{2^N}\frac{i+1}{N} = \pi_{i+1}p_{i+1,i},$$

confirming our guess.

**Example**: *Random walk on a graph* Let $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ be an undirected graph with a finite number of vertices $\mathscr{V}$ and edges $\mathscr{E} = \{(i,j) : i,j \in \mathscr{V}\}$. Let $d(i)$ be a degree of node $i$, defined as the number of edges that connect to $i$. Define a Markov chain $\{X_n\}$ on $\mathscr{V}$ with transition probabilities

$$p_{ij} = \begin{cases} \frac{1}{d(i)} & \text{if } (i,j) \in \mathscr{E}, \\ 0 & \text{if } (i,j) \notin \mathscr{E}. \end{cases}$$

If we assume that the graph is connected, then $\{X_n\}$ is irreducible and positive recurrent, because

$\mathscr{V}$ is finite. Let's try to solve detailed balance equations even though we don't know if this is going to work. For $(i,j) \in \mathscr{E}$

$$\pi_i p_{ij} = \pi_j p_{ji} \Rightarrow \pi_i \frac{1}{d(i)} = \pi_j \frac{1}{d(j)} \Rightarrow \pi_i = cd(i).$$

$\pi_i = cd(i) \Rightarrow c\sum_i d(i) = 1 \Rightarrow c = 1/2m$, where $m = |\mathscr{E}|$ is the number of edges in $\mathscr{E}$. Therefore a probability vector with entries

$$\pi_i = \frac{d(i)}{2m}$$

is a stationary distribution of $\{X_n\}$.

**Example**: *Birth-death chain* Consider a homogeneous Markov chain on the nonnegative integers $\{0,1,2,\dots\}$ with transition probabilities

$$p_{ij} = \begin{cases} r_i & \text{if } j = i, \\ p_i & \text{if } j = i+1, \\ q_i & \text{if } j = i-1, \\ 0 & \text{otherwise.} \end{cases}$$

Let's check if detailed balance holds.

$$\begin{aligned} \pi_o p_o = \pi_1 q_1 &\Rightarrow \pi_1 = \pi_0 \frac{p_0}{q_1} \\ \pi_1 p_1 = \pi_2 q_2 &\Rightarrow \pi_2 = \pi_1 \frac{p_1}{q_2} = \pi_0 \frac{p_0 p_1}{q_1 q_2} \\ &\vdots \\ \pi_i &= \pi_0 \prod_{j=1}^{i} \frac{p_{j-1}}{q_j}. \end{aligned}$$

Since we require $\sum_i \pi_i = 1$, the inequality $\pi_0 \sum_{i=0}^{\infty} \prod_{j=0}^{i} \frac{p_{j-1}}{q_j} < \infty$ must hold.

For example, if $p_i = p$ and $q_i = q$ for all $i \geq 0$, then $\pi_i = \pi_0 (p/q)^i$ and stationary distribution exists if and only if $\sum_{i=0}^{\infty} (p/q)^i < \infty \Leftrightarrow p < q$. In the case of $p < q$,

$$\sum_{i=0}^{\infty} \pi_i = \pi_0 \sum_{i=0}^{\infty} \left(\frac{p}{q}\right)^i = \pi_0 \times \frac{1}{1-p/q} = 1 \Rightarrow \pi_0 = 1 - \frac{p}{q} = \frac{q-p}{q} \Rightarrow \pi_i = \pi_0 \left(\frac{p}{q}\right)^i = \frac{(q-p)p^i}{q^{i+1}} = \left(1 - \frac{p}{q}\right)\left(\frac{p}{q}\right)^i.$$

So we showed that when $p_i = p$ and $q_i = q$, the stationary distribution of the birth-death chain is a geometric distribution with success probability $p/q$.