0.1 Markov Chains

0.1.1 Generalities

A *Markov Chain* consists of a countable (possibly finite) set S (called the *state space*) together with a countable family of random variables $X_\circ, X_1, X_2, \cdots$ with values in S such that

$$P[X_{l+1} = s \mid X_l = s_l, X_{l-1} = s_{l-1}, \cdots, X_\circ = s_\circ] = P[X_{l+1} = s \mid X_l = s_l].$$

We refer to this fundamental equation as the *Markov property*. The random variables $X_\circ, X_1, X_2, \cdots$ are dependent. Markov chains are among the few sequences of dependent random variables which are of a general character and have been successfully investigated with deep results about their behavior. Later we will discuss martingales which also provide examples of sequences of dependent random variables. Martingales have many applications to probability theory.

One often thinks of the subscript l of the random variable X_l as representing the time (discretely), and the random variables represent the evolution of a system whose behavior is only probabilistically known. Markov property expresses the assumption that the knowledge of the present (i.e., $X_l = s_l$) is relevant to predictions about the future of the system, however additional information about the past ($X_j = s_j$, $j \leq l-1$) is irrelevant. What we mean by the system is explained later in this subsection. These ideas will be clarified by many examples.

Since the state space is countable (or even finite) it customary (but not always the case) to use the integers $\mathbb{Z}$ or a subset such as $\mathbb{Z}_+$ (non-negative integers), the natural numbers $\mathbf{N} = \{1, 2, 3, \cdots\}$ or $\{0, 1, 2, \cdots, m\}$ as the state space. The specific Markov chain under consideration often determines the natural notation for the state space. In the general case where no specific Markov chain is singled out, we often use $\mathbf{N}$ or $\mathbb{Z}_+$ as the state space. We set

$$P_{ij}^{l,l+1} = P[X_{l+1} = j \mid X_l = i]$$

For fixed l the (possibly infinite) matrix $P_l = (P_{ij}^{l,l+1})$ is called the *matrix of transition probabilities* (at time l). In our discussion of Markov chains, the emphasis is on the case where the matrix P_l is independent of l which means that the law of the evolution of the system is time independent. For this reason one refers to such Markov chains as *time homogeneous* or having *stationary transition probabilities*. Unless stated to the contrary, all Markov chains considered in these notes are time homogeneous and therefore the subscript l is omitted and we simply represent the matrix of transition probabilities as $P = (P_{ij})$. P is called the *transition matrix*. The non-homogeneous case is generally called *time inhomogeneous* or *non-stationary in time*!

The matrix P is not arbitrary. It satisfies

$$P_{ij} \geq 0, \quad \sum_j P_{ij} = 1 \quad \text{for all } i. \tag{0.1.1.1}$$

A Markov chain determines the matrix P and a matrix P satisfying the conditions of (0.1.1.1) determines a Markov chain. A matrix satisfying conditions of (0.1.1.1) is called *Markov* or *stochastic*. Given an initial distribution $P[X_\circ = i] = p_i$, the matrix P allows us to compute the the distribution at any subsequent time. For example, $P[X_1 = j, X_\circ = i] = p_{ij}p_i$ and more generally

$$P[X_l = j_l, \cdots, X_1 = j_1, X_\circ = i] = P_{j_{l-1}j_l}P_{j_{l-2}j_{l-1}} \cdots P_{ij_1}p_i. \tag{0.1.1.2}$$

Thus the distribution at time $l = 1$ is given by the row vector $(p_1, p_2, \cdots)P$ and more generally at time l by the row vector

$$(p_1, p_2, \cdots) \underbrace{PP\cdots P}_{l \text{ times}} = (p_1, p_2, \cdots)P^l. \tag{0.1.1.3}$$

For instance, for $l = 2$, the probability of moving from state i to state j in two units of time is the sum of the probabilities of the events

$$i \to 1 \to j, \; i \to 2 \to j, \; i \to 3 \to j, \cdots, i \to n \to j,$$

since they are mutually exclusive. Therefore the required probability is $\sum_k P_{ik}P_{kj}$ which is accomplished by matrix multiplication as given by (0.1.1.3) Note that $(p_1, p_2, \cdots)$ is a row vector multiplying P on the left side. Equation (0.1.1.3) justifies the use of matrices is describing Markov chains since the transformation of the system after l units of time is described by l-fold multiplication of the matrix P with itself.

This basic fact is of fundamental importance in the development of Markov chains. It is convenient to make use of the notation $P^l = (P^{(l)}_{ij})$. Then for $r + s = l$ (r and s non-negative integers) we have

$$P^l = P^r P^s \quad \text{or} \quad P^{(l)}_{ij} = \sum_k P^{(r)}_{ik} P^{(s)}_{kj}. \tag{0.1.1.4}$$

Example 0.1.1.1 Let $\mathbb{Z}/n$ denote integers mod n, let $Y_1, Y_2, \cdots$ be a sequence of independent indentically distributed (from now on iid) random variables with values in $\mathbb{Z}/n$ and density function

$$P[Y = k] = p_k.$$

Set $Y_\circ = 0$ and $X_l = Y_\circ + Y_1 + \cdots + Y_l$ where addition takes place in $\mathbb{Z}/n$. Using

$$X_{l+1} = Y_{l+1} + X_l,$$

the validity of the Markov property and time stationarity are easily verified and it follows that $X_\circ, X_1, X_2 \cdots$ is a Markov chain with state space $\mathbb{Z}/n = \{0, 1, 2, \cdots, n-1\}$. The equation $X_{l+1} = Y_{l+1} + X_l$ also implies that transition matrix P is

$$P = \begin{pmatrix} p_\circ & p_1 & p_2 & \cdots & p_{n-2} & p_{n-1} \\ p_{n-1} & p_\circ & p_1 & \cdots & p_{n-3} & p_{n-2} \\ p_{n-2} & p_{n-1} & p_\circ & \cdots & p_{n-4} & p_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ p_2 & p_3 & p_4 & \cdots & p_\circ & p_1 \\ p_1 & p_2 & p_3 & \cdots & p_{n-1} & p_\circ \end{pmatrix}$$

We refer to this Markov chain as the *general random walk on* $\mathbb{Z}/n$. Rather than starting at 0 ($X_\circ = Y_\circ = 0$), we can start at some other point by setting $Y_\circ = m$ where $m \in \mathbb{Z}/n$. A possible way of visualizing the random walk is by assigning to $j \in \mathbb{Z}/n$ the point $e^{\frac{2\pi i j}{n}}$ on the unit circle in the complex plane. If for instance $p_k = 0$ for $k \neq 0, \pm 1$, then imagine particles at any and all locations $j \leftrightarrow e^{\frac{2\pi i j}{n}}$, which after passage of one unit of time, stay at the same place, or move one unit counterclockwise or clockwise with probabilities $p_\circ, p_1$ respectively and independently of each other. The fact that moving counterclockwise/clockwise or staying at the same location have the same probabilities for all locations j expresses the property of spatial homogeneity which is specific to random walks and not shared by general Markov chains. This property is expressed by the rows of the transition matrix being shifts of each other as observed in the expression for P. For general Markov chains there is no relation between the entries of the rows (or columns) except as specified by (0.1.1.1). Note that the transition matrix of the general random walk on $\mathbb{Z}/n$ has the additional property that the column sums are also one and not just the row sums as stated in (0.1.1.1). A stochastic matrix with the additional property that column sums are 1 is called *doubly stochastic.*

Example 0.1.1.2 We continue with the preceding example and make some modifications. Assume $Y_\circ = m$ where $1 \leq m \leq n-2$, and $p_j = 0$ unless $j = 1$ or $j = -1$ (which is the same thing as $n-1$ since addition is mod n.) Set $P(Y = 1) = p$ and $P[Y = -1] = q = 1 - p$. Modify the matrix P by leaving P_{ij} unchanged for $1 \leq i \leq n-2$ and defining

$$P_{\circ\circ} = 1,\ P_{\circ j} = 0,\ P_{n-1\ n-1} = 1, P_{n-1\ k} = 0,\ \ j \neq 0,\ k \neq n-1.$$

This is still a Markov chain. The states 0 and $n-1$ are called *absorbing* states since transition outside of them is impossible. Note that this Markov chain describes the familiar *Gambler's Ruin Problem.* ♠

Remark 0.1.1.1 In example 0.1.1.1 we can replace $\mathbb{Z}/n$ with $\mathbb{Z}$ or more generally $\mathbb{Z}^m$ so that addition takes place in $\mathbb{Z}^m$. In other words, we can start with iid sequence of random variables $Y_1, Y_2, \cdots$ with values in $\mathbb{Z}^m$ and define

$$X_\circ = 0, \quad X_{l+1} = Y_{l+1} + X_l.$$

By the same reasoning as before the sequence $X_\circ, X_1, X_2, \cdots$ is a Markov chain with state space $\mathbb{Z}^m$. It is called the *general random walk* on $\mathbb{Z}^m$. If $m = 1$ and the random variable Y (i.e. any of the Y_j's) takes only values ± 1 then it is called a *simple random walk* on $\mathbb{Z}$ and if in addition the values ± 1 are assumed with equal probability $\frac{1}{2}$ then it is called the *simple symmetric random* walk on $\mathbb{Z}$. The analogous definition for $\mathbb{Z}^m$ is obtained by assuming that Y only takes $2m$ values

$$(\pm 1, 0, \cdots, 0), \ (0, \pm 1, 0, \cdots, 0), \cdots, \ (0, \cdots, 0, \pm 1),$$

each with probability $\frac{1}{2m}$. One similarly defines the notions of simple and symmetric random walks on $\mathbb{Z}/n$. $\heartsuit$

In a basic course on probability it is generally emphasized that the underlying probability space should be clarified before engaging in the solution of a problem. Thus it is important to understand the underlying probability space in the discussion of Markov chains. This is most easily demonstrated by looking at the Markov chain $X_\circ, X_1, X_2, \cdots$, with finite state space $\{1, 2, \cdots, n\}$, specified by an $n \times n$ transition matrix $P = (P_{ij})$. Assume we have n biased dice with each die having n sides. There is one die corresponding each state. If the Markov chain is in state i then the i^{th} die is rolled. The die is biased and side j of die number i appears with probability P_{ij}. For definiteness assume $X_\circ = 1$. If we are interested in investigating questions about the Markov chain in $L \leq \infty$ units of time (i.e., the subscript $l \leq L$), then we are looking at all possible sequences $1k_1k_2k_3 \cdots k_L$ if $L < \infty$ (or infinite sequences $1k_1k_2k_3 \cdots$ if $L = \infty$). The sequence $1k_1k_2k_3 \cdots k_L$ is the event that die number 1 was rolled and side k_1 appeared; then die number k_1 was rolled and side k_2 appeared; then die number k_2 was rolled and side number k_3 appeared and so on. The probability assigned to this event is

$$P_{1k_1} P_{k_1k_2} P_{k_2k_3} \cdots P_{k_{L-1}k_L}.$$

One can graphically represent each event $1k_1k_2k_3 \cdots k_L$ as a function consisting of broken line segments joining the point $(0, 1)$ to $(1, k_1)$, $(1, k_1)$ to $(2, k_2)$, $(2, k_2)$ to $(3, k_3)$ and so on. Alternatively one can look at the event $1k_1k_2k_3 \cdots k_L$ as a step function taking value k_m on the interval $[m, m + 1)$. Either way the horizontal axis represents time and the vertical axis the state or site. Naturally one refers to a sequence $1k_1k_2k_3\cdots k_L$ or its graph as a *path*, and each path represents a *realization* of the Markov chain. Graphic representations are useful devices for understanding Markov chains. *The underlying probability space* Ω *is the set of all possible paths in whatever representation one likes.* Probabilities (or measures in more sophisticated language) are assigned to events $1k_1k_2k_3\cdots k_L$ or paths (assuming $L<\infty$) as described above. We often deal with conditional probabilities such as $P[\star|X_\circ=i]$. The appropriate probability space in this, for example, will all paths of the form $ik_1k_2k_3\cdots$.

Example 0.1.1.3 Suppose $L=\infty$ so that each path is an infinite sequence $1k_1k_2k_3\cdots$ in the context described above, and Ω is the set of all such paths. Assume $P_{ij}^{(l)}=\alpha>0$ for some given i,j and l. How is this statement represented in the space Ω? In this case we consider all paths $ik_1k_2k_3\cdots$ such that $k_l=j$ and no condition on the remaining k_m's. The statement $P_{ij}^{(l)}=\alpha>0$ means this set of paths in Ω has probability α. ♠

What makes a random walk special is that instead of having one die for every site, the same die (or an equivalent one) is used for all sites. Of course the rolls of the die for different sites are independent. This is the translation of the space homogeneity property of random walks to this model. This construction extends in the obvious manner to the case when the state space is infinite (i.e., rolling dice with infinitely many sides). It should be noted however, that when $L=\infty$ any given path $1k_1k_2k_3\cdots$ extending to ∞ will generally have probability 0, and sets of paths which are specified by finitely many values $k_{i_1}k_{i_2}\cdots k_{i_m}$ will have non-zero probability. It is important and enlightening to keep this description of the underlying probability space in mind. It will be further clarified and amplified in the course of future developments.

Example 0.1.1.4 Consider the simple symmetric random walk $X_\circ=0,X_1,X_2,\cdots$ where one may move one unit to the right or left with probability $\frac{1}{2}$. To understand the underlying probability space Ω, suppose a 0 or a 1 is generated with equal probability after each unit of time. If we get a 1, the path goes up one unit and if we a 0 then the path goes down one unit. Thus the space of all paths is the space of all sequences of 0's and 1's. Let $\omega=0a_1a_2\cdots$ denote a typical path. Expanding every real number $\alpha\in[0,1]$ in binary, i.e., in the form

$$\alpha=\frac{a_1}{2}+\frac{a_2}{2^2}+\frac{a_3}{2^3}+\cdots,$$

with $a_j=0$ or 1, we obtain a one to one correspondence between $[0,1]$ and the set of paths[1]. Under this correspondence the set of paths with $a_1=1$ is precisely the interval $[\frac{1}{2},1]$ and

[1]There is the minor problem that a rational number has more than one representation, e.g., $\frac{1}{2}=\frac{1}{4}+\frac{1}{8}+\cdots$ But such non-uniqueness occurs for only rational numbers which are countable and therefore have probability zero as will become clear shortly. Thus it does not affect our discussion.

the set of paths with $a_1 = 0$ is the interval $[0, \frac{1}{2}]$. Similarly, the set of paths with $a_2 = 0$ corresponds to $[0, \frac{1}{4}] \cup [\frac{1}{2}, \frac{3}{4}]$. More generally the subset of $[0, 1]$ corresponding to $a_k = 0$ or $a_k = 1$ is a union of 2^k disjoint intervals each of length $\frac{1}{2^{k+1}}$. Therefore the probability of the set of paths with $a_k = 0$ (or $a_k = 1$) is the just the sum of the lengths of these intervals. Thus in this case looking at the space of paths and corresponding probabilities as determined by the simple symmetric random walk is nothing more than taking lenths of unions of intervals in the most familiar way. ♠

With the above description of the undelying probability space Ω in mind, we can give a more precise meaning to the word *system* and its *evolution* as referenced earlier. Assume the state space is finite, $S = \{1, 2, \cdots, n\}$ for example, and imagine a large number Mn of dice with M identical dice for each state i. As before assume for definiteness that $X_\circ = 1$ and at time $l = 0$ all M dice corresponding to state 1 are rolled independently of each other. The outcomes are $k_1^1, k_1^2, \cdots, k_1^M$. At time $l = 1$, k_1^1 dice corresponding to state 1, k_1^2 dice corresponding state 2, k_1^3 dice corresponding state 3, etc. are rolled independently. The outcomes will be k_2^1 dice will show 1, k_2^2 will show number 2 etc. Repeating the process, we independently roll k_2^1 dice corresponding state 1, k_2^2 dice corresponding to state 2, k_2^3 dice corresponding to state 3 etc. The outcomes will be $k_3^1, k_3^2, \cdots, k_3^n$, and we repeat the process. In this fashion instead of obtaining a single path we obtain M paths independently of each other. At each time l, the numbers $k_l^1, k_l^2, k_l^3, \cdots, k_l^M$ define the system and the paths describe the evolution of the system. The assumption that $X_\circ = 1$ was made only for convenience and we could have assumed that at time $l = 0$, the system was in state $k_\circ^1, k_\circ^2, \cdots, k_\circ^M$ in which case at time $l = 0$ dice numbered $k_\circ^1, k_\circ^2, \cdots, k_\circ^M$ would have been rolled independently of each other. Since M is assumed to be an arbitrarily large number, from the set of paths that at time l are in state i, a portion approximately equal to P_{ij} transfer to state j in time $l + 1$ (Law of Large Numbers).

To give another example, assume we have M (a large number) of dice all showing number 1 at time $l = 0$. At the end of each unit of time, the number on each die will either remain unchanged, say with probability $p_\circ$, or will change by addition of ± 1 where addition is in $\mathbb{Z}/n$. We assume ± 1 are equally probable each having probability p_1 and $p_\circ + 2p_1 = 1$. As time goes on the composition of the numbers on the dice will change, i.e, the system will evolve in time. While any individual die will undergo many changes (with high probability), one may expect that the total composition of the numbers on the dice to settle down to something which can be understood, like for example, approximately the same number of 0's, 1's, 2's, $\cdots$, $n - 1$'s. In other words, while each individual die changes, the system as a whole will reach some form of equilibrium. An important goal of this course is provide an analytical framework which would allow us to effectively deal with phenomena of this nature.

EXERCISES

Exercise 0.1.1.1 *Consider the simple symmetric random walks on* $\mathbb{Z}/7$ *and* $\mathbb{Z}$ *with* $X_\circ = 0$. *Using a random number generator make graphs of ten paths describing realizations of the Markov chains from* $l = 0$ *to* $l = 100$.

Exercise 0.1.1.2 *Consider the simple symmetric random walk* $S_\circ = (0,0), S_1, S_2, \cdots$ *on* $\mathbb{Z}^2$ *where a path at* (i,j) *can move to either of four points* $(i \pm 1, j)$, $(i, j \pm 1)$ *with probability* $\frac{1}{4}$. *Assume we impose the requirement that the random walk cannot visit any site more than once. Is the resulting system a Markov chain? Prove your answer.*

Exercise 0.1.1.3 *Let* $S_\circ = 0, S_1, S_2, \cdots$ *denote the simple symmetric random walk on* $\mathbb{Z}$. *Show that the sequence of random variables* $Y_\circ, Y_1, Y_2, \cdots$ *where* $Y_j = |S_j|$ *is a Markov chain with state space* $\mathbb{Z}_+$ *and exhibit its transition matrix.*

Exercise 0.1.1.4 *Consider the simple symmetric random walk on* $\mathbb{Z}^2$ *(see exercise 0.1.1.2 for the definition). Let* $S_j = (X_j, Y_j)$ *denote the coordinates of* S_j *and define* $Z_l = X_l^2 + Y_l^2$. *Is* Z_l *a Markov chain? Prove your answer.* (Hint - You may use the fact that an integer may have more than one essentially distinct representation as a sum of squares, e.g., $25 = 5^2 + 0 = 4^2 + 3^2$.)

0.1.2 Classification of States

The first step in understanding the behavior of Markov chains is to classify the states. We say state j is *accessible* from state i if it possible to make the transition from i to j is finite units of time. This translates into $P_{ij}^{(l)} > 0$ for some $l \geq 0$. This property is denoted by $i \rightarrow j$. If j is accessible from i and i is accessible from j then we say i and j *communicate*. In case i and j communicate we write $i \leftrightarrow j$. Communication of states is an equivalence relation which means

1. $i \leftrightarrow i$. This is valid since $P^\circ = I$.

2. $i \leftrightarrow j$ implies $j \leftrightarrow i$. This follows from the definition of communicate.

3. If $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$. To prove this note that the hypothesis implies $P_{ij}^{(r)} > 0$ and $P_{jk}^{(s)} > 0$ for some integers $r, s \geq 0$. Then $P_{ik}^{(r+s)} \geq P_{ij}^{(r)} P_{jk}^{(s)} > 0$ proving k is accessible from i. Similarly i is accessible from k.

To classify the states we group them together according to the equivalence relation $\leftrightarrow$ (communication).

Example 0.1.2.1 Let the transition matrix of a Markov chain be of the form

$$P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$$

where P_1 and P_2 are $n \times n$ and $m \times m$ matrices. It is clear that none of the states $i \leq n$ is accessible from any of the states $n+1, n+2, \cdots, n+m$, and vice versa. If the matrix of a finite state Markov chain is of the form

$$P = \begin{pmatrix} P_1 & Q \\ 0 & P_2 \end{pmatrix},$$

then none of the states $i \leq n$ is accessible from any of the states $n+1, n+2, \cdots, n+m$, however, whether a state $j \geq n+1$ is accessible from a state $i \leq n$ depends on the matrices P_1, P_2 and Q. ♠

For a state i let $d(i)$ denote the greatest common divisor (gcd) of all integers $l \geq 1$ such that $P_{ii}^{(l)} > 0$. If $P_{ii}^{(l)} = 0$ for all $l \geq 1$, then we set $d(i) = 0$. If $d(i) = 1$ then we say state i

is *aperiodic*. If $d(i) \geq 2$ then we say state i is *periodic* with period $d(i)$. A simple example of a Markov chain where every state has period n is given by the $n \times n$ transition matrix

$$P = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

The process represented by this matrix is deterministic not probabilistic since it means that with the passage of each unit of time the transitions

$$1 \rightarrow 2,\ 2 \rightarrow 3,\ \cdots,\ n-1 \rightarrow n,\ n \rightarrow 1$$

take place with probability 1. Although this example is somewhat artificial, yet one should keep such chains in mind. A more realistic example of a periodic Markov chain (i.e., every state is periodic) is given by the following example:

Example 0.1.2.2 Consider a simple random walk on $\mathbb{Z}/n$ with $n = 2m$ an even integer, i.e., assume the random variable Y of the definition of general random walk on $\mathbb{Z}/n$ has density function

$$P[Y = 1] = p > 0,\ P[Y = n-1] = q = 1 - p > 0.$$

Looking at this random walk as taking place on the points $e^{\frac{2\pi ij}{n}}$ on the unit circle, we see that it describes the evolution of a system where after passage of each unit of time it moves counterclockwise one unit with probability p and clockwise with probability $q = 1-p$. Since both p and q are positive and n is even, every state is periodic with period 2. In fact, assuming $X_\circ = 0$, $X_{2l} \in \{0, 2, \cdots, 2m\}$ and $X_{2l-1} \in \{1, 3, \cdots, 2m-1\}$. If n were odd, then every state would be aperiodic. It is also clear that every state communicates with every other state. The same conclusions are valid for a simple random walk on $\mathbb{Z}^m$. ♠

The relationship between periodicity and communication is described by the following lemma:

Lemma 0.1.2.1 *If $i \leftrightarrow j$, then $d(i) = d(j)$.*

Proof - Let m, l and r be such that

$$P_{ij}^{(m)} > 0, \quad P_{ji}^{(l)} > 0, \quad P_{ii}^{(r)} > 0.$$

Then

$$P_{jj}^{(l+m)} > 0, \quad P_{jj}^{(l+r+m)} > 0.$$

Since $d(j)$ is the gcd of all k such that $P_{jj}^{(k)} > 0$, $d(j)$ divides $l+m$, $l+r+m$ and consequently $d(j)|(l+r+m-m-l) = r$. From $d(j)|r$ it follows that $d(j)|d(i)$. Because of the symmetry between i and j, $d(i)|d(j)$, and so $d(i) = d(j)$ as required. ♣

To further elaborate on the states of a Markov chain we introduce the notion of *first hitting* or *passage time* T_{ij} which is a function (or random variable) on the probability space Ω with values in $\mathbf{N}$. To each $\omega \in \Omega$, which as we explained earlier, is a path or sequence $\omega = ik_1k_2\cdots$, T_{ij}, assigns the smallest positive integer $l \geq 1$ such that $\omega(l) \stackrel{def}{=} k_l = j$. We also set

$$F_{ij}^l = P[T_{ij} = l] = P[X_l = j, X_{l-1} \neq j, \cdots, X_1 \neq j \mid X_\circ = i].$$

The quantity

$$F_{ij} = \sum_{l=1}^{\infty} F_{ij}^l$$

is the probability that at some point in time the Markov chain will visit or hit state j given that it started in state i. A state i is called *recurrent* if $F_{ii} = 1$; otherwise it is called *transient*. The relationship between recurrence and communication is given by the following lemma:

Lemma 0.1.2.2 *If $i \leftrightarrow j$, and i is recurrent, then so is j.*

Proof - Another proof of this lemma will be given shortly. Here we prove it using only the idea of paths. Let l be the smallest integer such that $P_{ij}^{(l)} > 0$. Therefore the set of paths Γ_{ij}^l which at time 0 are at i and at time l are at j has probability $P_{ij}^{(l)} > 0$. By the minimality of l and Markov property, the paths in Γ_{ij}^l do not return to i before hitting j. If j were not recurrent then a subset $\Gamma' \subset \Gamma_{ij}^l$ of positive probability will never return to j. But then this subset cannot return to i either since otherwise a fraction of positive probability of it will return to j. Therefore the paths in Γ_{ij}^l do not return to i which contradicts the recurrence of i. ♣

A subset $C \subset S$ is called *irreducible* if all states in C communicate. C is called *closed* if no state outside of C is accessible from any state in C. A simple and basic result about the classification of states of a Markov chain is

Proposition 0.1.2.1 *The state space of a Markov chain admits of the decomposition*

$$S = T \cup C_1 \cup C_2 \cup \cdots ,$$

where T is the set of transient states, and each C_i is an irreducible closed set consisting of recurrent states.

Proof - Let $C \subset S$ denote the set of recurrent states and T be the complement of C. In view of lemma 0.1.2.2 states in T and C do not communicate. Decompose C into equivalence classes $C_1, C_2, \cdots$ according to $\leftrightarrow$ so that each C_a is irreducible, i.e., all states within each C_a communicate with each other. It remains to show no state in C_b or T is accessible from any state in C_a for $a \neq b$. Assume $i \rightarrow j$ with $i \in C_a$ and $j \in C_b$ (or $j \in T$), then $P_{ij}^{(l)} > 0$ for some l, and let l be the smallest such integer. Since by assumption $j \not\rightarrow i$ then $P_{ji}^{(m)} = 0$ for all m, that is, there are no paths from state j back to state i, and it follows that

$$\sum_{k=1}^{\infty} F_{ii}^{k} \leq 1 - P_{ij}^{(l)} < 1,$$

contradicting recurrence of i. ♣

Next we turn our attention to Markov chains. Let $X_\circ, X_1, \cdots$ be a Markov chain and for convenience let $\mathbb{Z}_+$ be the state space. Recall that the random variable T_{ij} is the first hitting time of state j given that the Markov chain is in state i at time $l = 0$. The density function of T_{ij} is $F_{ij}^l = P[T_{ij} = l]$. Naturally we define the generating function for T_{ij} as

$$\mathsf{F}_{ij} = \sum_{l=1}^{\infty} F_{ij}^{l} \xi^{l}.$$

Note that the summation starts at $l = 1$ not 0. We also define the generating function

$$\mathsf{P}_{ij} = \sum_{l=0}^{\infty} P_{ij}^{(l)} \xi^{l}.$$

These infinite series converge for $|\xi| < 1$. Much of the theory of Markov chains that we develop is based on the exploitation of the relation between the generating functions $\mathsf{P}_\star$ and $\mathsf{F}_\star$ as given by the following theorem whose validity and proof depends strongly on the Markov property:

Theorem 0.1.2.1 *The following identities are valid:*

$$\mathsf{F}_{ii}\mathsf{P}_{ii} = \mathsf{P}_{ii} - 1, \quad \mathsf{P}_{ij} = \mathsf{F}_{ij}\mathsf{P}_{jj} \quad \text{for } i \neq j.$$

Proof - The coefficients of ξ^m in P_{ij} and in $\mathsf{F}_{ij}\mathsf{P}_{jj}$ are

$$P_{ij}^{(m)}, \quad \text{and} \quad \sum_{k=1}^{m} F_{ij}^{k} P_{jj}^{(m-k)}$$

respectively. The set of paths that start at i at time $l=0$ and are in state j at time $l=m$ is the disjoint union (as k varies) of the paths starting at i at time $l=0$, hitting state j for the first time at time $k \leq m$ and returning to state j after $m-k$ units of time. Therefore $P_{ij}^{(m)} = \sum_k F_{ij}^{k} P_{jj}^{(m-k)}$ proving the second identity. Noting that the lowest power of ξ in P_{ii} is zero, while the lowest power of ξ in F_{ii} is 1, one proves the first identity similarly. ♣

The following corollaries point to the significance of proposition 0.1.3.1:

Corollary 0.1.2.1 *A state i is recurrent if and only if $\sum_l P_{ii}^{(l)} = \infty$. Equivalently, a state k is transient if and only if $\sum_l P_{kk}^{(l)} < \infty$.*

Proof - From the first identity of proposition 0.1.3.1 we obtain

$$\mathsf{P}_{ii}(\xi) = \frac{1}{1 - \mathsf{F}_{ii}(\xi)},$$

from which the required result follows by taking the $\lim \xi \to 1^-$. ♣

Remark 0.1.2.1 In the proof of corollary 0.1.3.1, the evaluation of $\lim \xi \to 1^-$ requires justification since the series for $\mathsf{F}_{ii}(\xi)$ and $\mathsf{P}_{ii}(\xi)$ may be divergent for $\xi = 1$. According to a theorem of analysis (due to Abel) if a power series $\sum c_j \xi^j$ converges for $|\xi| < 1$ and $c_j \geq 0$, then

$$\lim_{\xi \to 1^-} \sum_{j=\circ}^{\infty} c_j \xi^j = \lim_{n \to \infty} \sum_{j=\circ}^{n} c_j = \sum_{j=\circ}^{\infty} c_j,$$

where we allow ∞ as a limit. This result removes any technical objection to the proof of corollary 0.1.3.1. Note the assumption $c_j \geq 0$ is essential. For example, substituting $x=1$ in $\frac{1}{1+x} = \sum (-1)^n x^n$, valid for $|x| < 1$, we obtain

$$\frac{1}{2} = 1 - 1 + 1 - 1 + 1 - 1 + \cdots,$$

which is absurd in the ordinary sense of convergence of series. ♡

Corollary 0.1.2.2 *If i is a recurrent state and $i \leftrightarrow j$, then j is recurrent.*

Proof - By assumption

$$P_{ij}^{(k)} > 0, \quad P_{ji}^{(m)} > 0$$

for some k and m. Therefore

$$\sum_l P_{jj}^{(l)} \geq \sum_r P_{jj}^{(k+r+m)} \geq P_{ji}^{(m)} P_{ij}^{(k)} \sum_r P_{ii}^{(r)} = \infty,$$

which proves the assertion by corollary 0.1.3.1. ♣

We use corollary 0.1.3.1 to show that, in a sense which will be made precise shortly, a transient state is visited only finitely many times with probability 1. It is important to understand clearly the sense in which this statement is true. Let $X_\circ, X_1, X_2, \cdots$ be a Markov chain with state space $\mathbb{Z}_+$, $X_\circ = 0$ and 0 a transient state. Let Ω be the underlying probability space and $\Omega_\circ$ be the subset consisting of all $\omega = 0k_1k_2\cdots$ such that $k_l = 0$ for infinitely many l's. Let $\Omega^{(m)} \subset \Omega$ be subset of $\omega = 0k_1k_2\cdots$ such that $k_m = 0$. The key observation is proving that the subset $\Omega_\circ$ has probability 0 is the identity of sets

$$\Omega_\circ = \bigcap_{l=1}^{\infty} \bigcup_{m=l}^{\infty} \Omega^{(m)}. \tag{0.1.2.1}$$

To understand this identity let $A_l = \cup_{m=l}^{\infty} \Omega^{(m)}$, then $A_l \supset A_{l+1} \supset \cdots$ and each A_l contains all paths which visit 0 infinitely often. Therefore their intersection contains all paths that visit 0 infinitely often. On the other hand, if a path ω visits 0 only finitely many times then for some N and all $l \geq N$, $\omega \notin A_l$ and consequently $\omega \notin \cap A_l$. This proves (0.1.3.5). Now since 0 is transient $\sum_l P_{\circ\circ}^{(l)} < \infty$ which implies

$$P[\cup_{m=l}^{\infty} \Omega^{(m)}] \leq \sum_{m=l}^{\infty} P_{\circ\circ}^{(m)} \longrightarrow 0 \tag{0.1.2.2}$$

as $l \to \infty$. It follows from (0.1.3.5) that

Corollary 0.1.2.3 *With the above notation and hypotheses,* $P[\Omega_\circ] = 0$.

In other words, corollary 0.1.3.3 shows that while the set of paths starting at a transient state 0 and visiting it infinitely often is not necessarily empty, yet it has probability zero.

Remark 0.1.2.2 In an infinite state Markov chain the set of paths visiting a given transient state at least m times may have positive probability for every m. It is shown later that if $p \neq \frac{1}{2}$ then for the simple random walk on $\mathbb{Z}$ every state is transient. It is a simple matter to see that if in addition $p \neq 0, 1$ then the probability of at least m visits to any given state is positive for every fixed $m < \infty$. ♡

EXERCISES

Exercise 0.1.2.1 *Consider a $n \times n$ chess board and a knight which from any position can move to all other legitimate positions (according to the rules of chess) with equal probabilities. Make a Markov chain out of the positions of the knight. What is the decomposition in proposition 0.1.2.1 in cases $n = 3$ and $n = 8$?*

Exercise 0.1.2.2 *Let i and j be distinct states and l be the smallest integer such that $P_{ij}^{(l)} > 0$ (which we assume exists). Show that*

$$\sum_{k=1}^{l} F_{ii}^{(k)} \leq 1 - P_{ij}^{(l)}.$$

Exercise 0.1.2.3 *Consider the Markov chain specified by the following matrix:*

$$\begin{pmatrix} \frac{9}{10} & \frac{1}{20} & 0 & \frac{1}{20} & 0 \\ 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ 0 & \frac{4}{5} & \frac{1}{5} & 0 & \\ 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}$$

Draw a directed graph with a vertex representing a state, and arrows representing possible transitions. Determine the decomposition in proposition 0.1.2.1 for this Markov chain

Exercise 0.1.2.4 *The transition matrix of a Markov chain is $\begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$, where $0 \leq p, q \leq 1$. Classify the states of two state Markov chains according to the values of p and q.*

Exercise 0.1.2.5 *Number the states of a finite state Markov chain according to the decomposition of proposition 0.1.2.1, that is, $1, 2, \cdots, n_1 \in T$, $n_1 + 1, \cdots, n_2 \in C_1$, etc. What general form can the transition matrix P have?*

Exercise 0.1.2.6 *Show that a finite state Markov chain has at least one recurrent state.*

Exercise 0.1.2.7 *For an integer $m \geq 2$ let $m = \overline{a_k a_{k-1} \cdots a_1 a_\circ}$ denote its expansion in base 10. Let $0 < p < 1$, $q = 1 - p$, and $\mathbb{Z}_{\geq 2} = \{2, 3, 4, \cdots\}$ be the set of integers ≥ 2. Consider the Markov chain with state space $\mathbb{Z}_{\geq 2}$ defined by the following rule:*

$$m \longrightarrow \begin{cases} \max(2, a_k^2 + a_{k-1}^2 + \cdots + a_1^2 + a_\circ^2) & \text{withprobability } p; \\ 2 & \text{withprobability } q. \end{cases}$$

Let $X_\circ$ be any distribution on $\mathbb{Z}_{\geq 2}$. Show that

$$C = \{2, 4, 16, 20, 37, 42, 58, 89, 145\}$$

is an irreducible closed set consisting of recurrent states, and every state $j \notin C$ is transient.

Exercise 0.1.2.8 *We use the notation and hypotheses of exercise 0.1.2.7 except for changing the rule defining the Markov chain as follows:*

$$m \longrightarrow \begin{cases} \max(2, a_k^2 + a_{k-1}^2 + \cdots + a_1^2 + a_\circ^2) & \text{withprobability } p; \\ \max(2, a_k + a_{k-1} + \cdots + a_1 + a_\circ) & \text{withprobability} q. \end{cases}$$

Determine the transient and recurrent states and implement the conclusion of proposition 0.1.2.1.

Exercise 0.1.2.9 *Consider the two state Markov chain $\{X_n\}$ with transition matrix*

$$\begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$$

where $0 < p, q < 1$. Let T_{ij} denote the first passage/hitting time of state j given that we are in state i and μ_{ij} be its expectation. Compute μ_{ij} by

1. *Using the density function for the random variable T_{ij};*
2. *Conditioning, i.e., using the relation $\mathsf{E}[\mathsf{E}[X|Y]] = \mathsf{E}[X]$.*

Exercise 0.1.2.10 *Consider the Markov chain with transition matrix*

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

Let T_{ij} denote the first passage/hitting time of state j given that we are in state i. Compute $P[T_{12} < \infty]$ and $P[T_{11} < \infty]$. What are the expectations of T_{12} and T_{11}?

Exercise 0.1.2.11 *Let P denote the transition matrix of a finite aperiodic irreducible Markov chain. Show that for some n all entries of P^n are positive.*

0.1.3 Generating Functions

Generating Functions are an important tool in probability and many other areas of mathematics. Some of their applications to various problems in stochastic processes will be discussed gradually in this course. The idea of generating functions is that when we have a number (often infinite) of related quantities, there may be a method of putting them together and get a nice function which can be used to draw conclusions that may not have possible, or would have been difficult, otherwise. To make this vague idea precise we introduce several examples which demonstrate the value of generating functions.

Let X be a random variable with values in $\mathbb{Z}_+$ and let f_X be its density function:

$$f_X(n) = P[X = n].$$

The most common way to make a generating function out of the quantities $f_X(n)$ is to define

$$\mathsf{F}_X(\xi) = \sum_{n=\circ}^{\infty} f_X(n)\xi^n = \mathsf{E}[\xi^X]. \tag{0.1.3.1}$$

This infinite series converges for $|\xi| < 1$ since $0 \leq f_X(n) \leq 1$ and $f_X(n) = 0$ for $n < 0$. The issue of convergence of the infinite series is not a serious concern for us. F_X is called the *probability generating function* of the random variable X. The fact that $\mathsf{F}_X(\xi) = \mathsf{E}[\xi^X]$ is significant. While the individual terms $f_X(n)$ may not be easy to evaluate, in some situations we can use our knowledge of probability, and specifically of the fundamental relation

$$\mathsf{E}[\mathsf{E}[Z \mid Y]] = \mathsf{E}[Z], \tag{0.1.3.2}$$

to evaluate $\mathsf{E}[Z]$ directly, and then draw conclusions about the random variable X. Examples 0.1.3.2 and 0.1.3.4 are simple demonstrations of this point.

Example 0.1.3.1 Just to make sure we understand the concept let us compute F_X for a couple of simple random variables. If X is binomial with parameter (n, p) then $f_X(k) = \binom{n}{k} p^k q^{n-k}$ where $q = 1 - p$, and

$$\mathsf{F}_X(\xi) = \sum_{k=\circ}^{n} \binom{n}{k} p^k q^{n-k} \xi^k = (q + p\xi)^n.$$

Similarly, if X is a Poisson random variable, then

$$f_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Consequently we obtain the expression

$$\mathsf{F}_X(\xi) = \sum e^{-\lambda}\frac{\lambda^k}{k!}\xi^k = e^{\lambda(\xi-1)},$$

for the generating function of a Poisson random variable. ♠

Let Y be another random variable with values in $\mathbb{Z}_+$ and let $\mathsf{F}_Y(\eta)$ be its probability generating function. The joint random variable (X, Y) takes values in $\mathbb{Z}_+ \times \mathbb{Z}_+$ and its density function is $f_{X,Y}(n, m) = P[X = n, Y = m]$. Note that we are not assuming X and Y are independent. The probability generating function for (X, Y) is defined as

$$\mathsf{F}_{X,Y}(\xi, \eta) = \sum_{n\geq \circ, m\geq \circ} f_{X,Y}(n, m)\xi^n\eta^m = \mathsf{E}[\xi^X\eta^Y].$$

An immediate consequence of the definition of independence of random variables is

Proposition 0.1.3.1 *The random variables X and Y are independent if and only if*

$$\mathsf{F}_{X,Y}(\xi, \eta) = \mathsf{F}_{X,Y}(\xi, 1)\mathsf{F}_{X,Y}(1, \eta).$$

An example to demonstrate the use of this proposition follows:

Example 0.1.3.2 A customer service manager receives X complaints every day and X is a Poisson random variable with parameter λ. Of these, he/she handles Y satisfactorily and the remaining Z unsatisfactorily. We assume that for a fixed value of X, Y is a binomial random variable with parameter (X, p). Let us compute the probability generating function for the joint random variable (Y, Z). We have

$$\begin{aligned}
\mathsf{F}_{Y,Z}(\eta, \zeta) &= \mathsf{E}[\eta^Y\zeta^Z] \\
&= \mathsf{E}[\eta^Y\zeta^{X-Y}] \\
&= \mathsf{E}[\mathsf{E}[\eta^Y\zeta^{X-Y}] \mid X] \\
&= \mathsf{E}[\zeta^X\mathsf{E}[(\tfrac{\eta}{\zeta})^Y] \mid X] \\
&= \mathsf{E}[\zeta^X(p\tfrac{\eta}{\zeta} + q)^X] \\
&= e^{\lambda(p\eta+q\zeta-1)} \\
&= e^{\lambda p(\eta-1)}e^{\lambda q(\zeta-1)} \\
&= \mathsf{F}_Y(\eta)\mathsf{F}_Z(\zeta).
\end{aligned}$$

From elementary probability we know that random variables Y and Z are Poisson, and thus the above calculation implies that the random variables Y and Z are *independent*! This is surprising since $Z = X - Y$. It should be pointed out that in this example one can also directly compute $P[Y = j, Z = k]$ to deduce the independence of Y and Z. ♠

Example 0.1.3.3 For future reference (see the discussion of Poisson processes) we calculate the generating function for the trinomial random variable. The binomial random variable was modeled as the number of H's in n tosses of a coin where H appeared with probability p. Now suppose we have a 3-sided die with side **i** appearing with probability p_i, $p_1 + p_2 + p_3 = 1$. Let X_i denotes the number of times side **i** has appeared in n rolls of the die. Then the probability density function for (X_1, X_2) is

$$P[X_1 = k_1, X_2 = k_2] = \binom{n}{k_1, k_2} p_1^{k_1} p_2^{k_2} p_3^{n-k_1-k_2}. \qquad (0.1.3.3)$$

The generating function for (X_1, X_2) is a function of two variables, namely,

$$\mathsf{F}_{X_1,X_2}(\xi, \eta) = \sum P[X_1 = k_1, X_2 = k_2]\xi^{k_1}\eta^{k_2},$$

where the summation is over all pairs of non-negative integers k_1, k_2 with $k_1 + k_2 \leq n$. Substituting from (0.1.3.3) we obtain

$$\mathsf{F}_{X_1,X_2}(\xi, \eta) = (p_1\xi + p_2\eta + p_3)^n, \qquad (0.1.3.4)$$

for the generating function of the trinomial random variable. ♠

Next we turn our attention to Markov chains. Let $X_\circ, X_1, \cdots$ be a Markov chain and for convenience let $\mathbb{Z}_+$ be the state space. Recall that the random variable T_{ij} is the first hitting time of state j given that the Markov chain is in state i at time $l = 0$. The density function of T_{ij} is $F_{ij}^l = P[T_{ij} = l]$. Naturally we define the generating function for T_{ij} as

$$\mathsf{F}_{ij} = \sum_{l=1}^{\infty} F_{ij}^l \xi^l.$$

Note that the summation starts at $l = 1$ not 0. We also define the generating function

$$\mathsf{P}_{ij} = \sum_{l=0}^{\infty} P_{ij}^{(l)} \xi^l.$$

These infinite series converge for $|\xi| < 1$. Much of the theory of Markov chains that we develop is based on the exploitation of the relation between the generating functions $\mathsf{P}_\star$ and $\mathsf{F}_\star$ as given by the following theorem whose validity and proof depends strongly on the Markov property:

Theorem 0.1.3.1 *The following identities are valid:*

$$\mathsf{F}_{ii}\mathsf{P}_{ii} = \mathsf{P}_{ii} - 1, \quad \mathsf{P}_{ij} = \mathsf{F}_{ij}\mathsf{P}_{jj} \quad \text{for } i \neq j.$$

Proof - The coefficients of ξ^m in P_{ij} and in $\mathsf{F}_{ij}\mathsf{P}_{jj}$ are

$$P_{ij}^{(m)}, \quad \text{and} \quad \sum_{k=1}^{m} F_{ij}^{k} P_{jj}^{(m-k)}$$

respectively. The set of paths that start at i at time $l = 0$ and are in state j at time $l = m$ is the disjoint union (as k varies) of the paths starting at i at time $l = 0$, hitting state j for the first time at time $k \leq m$ and returning to state j after $m - k$ units of time. Therefore $P_{ij}^{(m)} = \sum_k F_{ij}^k P_{jj}^{(m-k)}$ proving the second identity. Noting that the lowest power of ξ in P_{ii} is zero, while the lowest power of ξ in F_{ii} is 1, one proves the first identity similarly. ♣

The following corollaries point to the significance of proposition 0.1.3.1:

Corollary 0.1.3.1 *A state i is recurrent if and only if $\sum_l P_{ii}^{(l)} = \infty$. Equivalently, a state k is transient if and only if $\sum_l P_{kk}^{(l)} < \infty$.*

Proof - From the first identity of proposition 0.1.3.1 we obtain

$$\mathsf{P}_{ii}(\xi) = \frac{1}{1 - \mathsf{F}_{ii}(\xi)},$$

from which the required result follows by taking the $\lim \xi \to 1^-$. ♣

Remark 0.1.3.1 In the proof of corollary 0.1.3.1, the evaluation of $\lim \xi \to 1^-$ requires justification since the series for $\mathsf{F}_{ii}(\xi)$ and $\mathsf{P}_{ii}(\xi)$ may be divergent for $\xi = 1$. According to a theorem of analysis (due to Abel) if a power series $\sum c_j \xi^j$ converges for $|\xi| < 1$ and $c_j \geq 0$, then

$$\lim_{\xi \to 1^-} \sum_{j=\circ}^{\infty} c_j \xi^j = \lim_{n \to \infty} \sum_{j=\circ}^{n} c_j = \sum_{j=\circ}^{\infty} c_j,$$

where we allow ∞ as a limit. This result removes any technical objection to the proof of corollary 0.1.3.1. Note the assumption $c_j \geq 0$ is essential. For example, substituting $x = 1$ in $\frac{1}{1+x} = \sum (-1)^n x^n$, valid for $|x| < 1$, we obtain

$$\frac{1}{2} = 1 - 1 + 1 - 1 + 1 - 1 + \cdots,$$

which is absurd in the ordinary sense of convergence of series. ♡

Corollary 0.1.3.2 *If i is a recurrent state and $i \leftrightarrow j$, then j is recurrent.*

Proof - By assumption

$$P_{ij}^{(k)} > 0, \quad P_{ji}^{(m)} > 0$$

for some k and m. Therefore

$$\sum_l P_{jj}^{(l)} \geq \sum_r P_{jj}^{(k+r+m)} \geq P_{ji}^{(m)} P_{ij}^{(k)} \sum_r P_{ii}^{(r)} = \infty,$$

which proves the assertion by corollary 0.1.3.1. ♣

We use corollary 0.1.3.1 to show that, in a sense which will be made precise shortly, a transient state is visited only finitely many times with probability 1. It is important to understand clearly the sense in which this statement is true. Let $X_\circ, X_1, X_2, \cdots$ be a Markov chain with state space $\mathbb{Z}_+$, $X_\circ = 0$ and 0 a transient state. Let Ω be the underlying probability space and $\Omega_\circ$ be the subset consisting of all $\omega = 0k_1k_2\cdots$ such that $k_l = 0$ for infinitely many l's. Let $\Omega^{(m)} \subset \Omega$ be subset of $\omega = 0k_1k_2\cdots$ such that $k_m = 0$. The key observation is proving that the subset $\Omega_\circ$ has probability 0 is the identity of sets

$$\Omega_\circ = \bigcap_{l=1}^{\infty} \bigcup_{m=l}^{\infty} \Omega^{(m)}. \tag{0.1.3.5}$$

To understand this identity let $A_l = \cup_{m=l}^{\infty} \Omega^{(m)}$, then $A_l \supset A_{l+1} \supset \cdots$ and each A_l contains all paths which visit 0 infinitely often. Therefore their intersection contains all paths that visit 0 infinitely often. On the other hand, if a path ω visits 0 only finitely many times then for some N and all $l \geq N$, $\omega \notin A_l$ and consequently $\omega \notin \cap A_l$. This proves (0.1.3.5). Now since 0 is transient $\sum_l P_{\circ\circ}^{(l)} < \infty$ which implies

$$P[\cup_{m=l}^{\infty} \Omega^{(m)}] \leq \sum_{m=l}^{\infty} P_{\circ\circ}^{(m)} \longrightarrow 0 \tag{0.1.3.6}$$

as $l \to \infty$. It follows from (0.1.3.5) that

Corollary 0.1.3.3 *With the above notation and hypotheses, $P[\Omega_\circ] = 0$.*

In other words, corollary 0.1.3.3 shows that while the set of paths starting at a transient state 0 and visiting it infinitely often is not necessarily empty, yet it has probability zero.

An important general observation about generating functions is that the moments of a random variable X with values in $\mathbb{Z}_+$ can be recovered from the knowledge of the generating function for X. In fact, we have

$$\mathsf{E}[X] = (\frac{d\mathsf{F}_X(\xi)}{d\xi})_{\xi=1^-}, \quad \text{if } P[X=\infty]=0. \tag{0.1.3.7}$$

Occasionally one naturally encounters random variables for which $P[X=\infty]>0$ while the series $\sum nP[X=n]<\infty$. In such cases $\mathsf{E}[X]=\infty$ for obvious reasons. If furthermore $\mathsf{E}[X]<\infty$, then

$$\mathsf{Var}[X] = \left[\frac{d^2\mathsf{F}_X(\xi)}{d\xi^2} + \frac{d\mathsf{F}_X(\xi)}{d\xi} - \left(\frac{d\mathsf{F}_X(\xi)}{d\xi}\right)^2\right]_{\xi=1^-}. \tag{0.1.3.8}$$

Another useful relation involving generating functions is

$$\sum_n P[X>n]\xi^n = \frac{1-\mathsf{E}[\xi^X]}{1-\xi}. \tag{0.1.3.9}$$

The identities are proven by simple and formal manipulations. For example to prove (0.1.3.9), we expand right hand side to obtain

$$\frac{1-\mathsf{E}[\xi^X]}{1-\xi} = \left(1-\sum_{n=0}^{\infty}P[X=n]\xi^n\right)\left(\sum_{n=0}^{\infty}\xi^n\right).$$

The coefficient of ξ^m is on right hand side is

$$1-\sum_{j=0}^{m}P[X=j]=P[X>m],$$

proving (0.1.3.9). The coefficient $P[X>n]$ of ξ^n on left hand side of (0.1.3.9) is often called *tail probabilities*. We will see examples of tail probabilities later.

Example 0.1.3.4 As an application of (0.1.3.7) we consider a coin tossing experiment where H's appear with p and T's with probability $q=1-p$. Let the random variable X denote the time of the first appearance of a sequence of m consecutive H's. We compute $\mathsf{E}[X]$ using (0.1.3.7) and by evaluating $\mathsf{F}_X(\xi)=\mathsf{E}[\xi^X]$, and the latter calculation is carried out by conditioning. Let H^rT^s be the event that first r tosses were H's followed by s T's. It is clear that for $1\le j\le m$

$$\mathsf{E}[\xi^X \mid H^{j-1}T] = \xi^j\mathsf{E}[\xi^X], \quad \mathsf{E}[\xi^X \mid H^m] = \xi^m$$

Therefore

$$\begin{aligned} \mathsf{E}[\xi^X] &= \mathsf{E}[\mathsf{E}[\xi^X \mid Y]] \\ &= \sum_{j=1}^m qp^{j-1}\xi^j \mathsf{E}[\xi^X] + p^m\xi^m. \end{aligned}$$

Solving this equation for $\mathsf{E}[\xi^X]$ we obtain

$$\mathsf{F}_X(\xi) = \mathsf{E}[\xi^X] = \frac{p^m\xi^m(1-p\xi)}{1-\xi+qp^m\xi^{m+1}}. \tag{0.1.3.10}$$

Using (0.1.3.7), we obtain after a simple calculation,

$$\mathsf{E}[X] = \frac{1}{p} + \frac{1}{p^2} + \cdots + \frac{1}{p^m}.$$

Similarly we obtain

$$\mathsf{Var}[X] = \frac{1}{(qp^m)^2} - \frac{2m+1}{qp^m} - \frac{p}{q^2}$$

for the variance of X. ♠

In principle it is possible to obtain the generating function for the time of the first appearance of any given pattern of H's and T's by repeated conditioning as explained in the preceding examples. However, it is more beneficial to introduce a more efficient machinary for this calculation. The idea is most clearly explained by following through an example. Another application of this idea is given in the subsection on Patterns in Coin Tossing.

Suppose we want to compute the time of the first appearance of the pattern A, for example, $A = HHTHH$. We treat H and T as non-commuting indeterminates. We let X be the formal sum of all finite sequences (i.e., monomials in H and T) which end with the first appearance of the pattern A. We will doing formal algebraic operations on these formal sums in two non-commuting variables H and T, and also introduce 0 as the zero element which when multiplied by any quantity gives 0, and is the additive identity. In the case of the pattern $HHTHH$ we have

$$\begin{aligned} X \;=\; & HHTHH + HHHTHH + \\ & THHTHH + HHHHTHH + \\ & HTHHTHH + THHHTHH + \\ & TTHHTHH + \;\ldots \end{aligned}$$

Similarly let Y be the formal sum of all sequences (including the empty sequence which is represented by 1) which do not contain the given pattern A. For instance for $HHTHH$ we get

$$\begin{aligned} Y &= 1 + H + T + HH + HT + TH + TT + \\ & \quad \ldots + HHTHT + HHTTH + \ldots \end{aligned}$$

There is an obvious relation between X and Y independently of the chosen pattern, namely,

$$1 + Y(H + T) = X + Y. \tag{0.1.3.11}$$

The verification of this identity is almost trivial and is accomplished by noting that a monomial summand of $X + Y$ of length l either contains the given pattern for the first time at its end or does not contain it, and then looking at the first $n - 1$ elements of the monomial. There is also another linear relation between X and Y which depends on the nature of the the desired pattern. Denote a given pattern by A and let A^j (resp. A_j) denote the first j elements of the pattern starting from right (respectively left). Thus for $HHTHH$ we get

$$\begin{aligned} &A^1 = H, \quad A^2 = HH, \quad A^3 = THH, \quad A^4 = HTHH; \\ &A_1 = H, \quad A_2 = HH, \quad A_3 = HHT, \quad A_4 = HHTH. \end{aligned}$$

Let Δ_j be 0 unless $A_j = A^j$ in which case it is 1. We obtain

$$YA = S(1 + A^1\Delta_{n-1} + A^2\Delta_{n-2} + \ldots + A^{n-1}\Delta_1). \tag{0.1.3.12}$$

For example in this case we get

$$YHHTHH = S(1 + A^3HH + A^4H).$$

Some experimentation will convince the reader that this identity is really the content of conditioning argument involved in obtaining the generating function for the time of first occurrence of a given pattern. At any rate its validity is easy to see. Equations (0.1.3.11) and (0.1.3.12) give us two linear equations which we can solve easily to obtain expressions for X and Y. Our primary interest in the expression for X. Therfore substituting for Y in (**??**) from (**??**) we obtain

$$A(1 - X) = X\left[A + \big(1 + \sum_{j=1}^{n-1} A^j\Delta_j\big)\big(1 - H - T\big)\right] \tag{0.1.3.13}$$

which gives an expression for X. Now assume H appears with probability p and T with probability $q = 1 - p$. Since X is the formal sum of all finite sequences ending in the first appearance of the desired pattern, by substituting $p\xi$ for H and $q\xi$ for T in the expression for X we obtained the desired probability generating function F (for the time τ of the first appearance of the pattern A). Denoting the result of this substitution in $A^j, A, \ldots$ by $A^j(\xi), A(\xi), \ldots$ we obtain

$$\mathsf{F}(\xi) = \frac{A(\xi)}{A(\xi) + \big(1 + \sum_{j=1}^{n-1} A^j(\xi)\Delta_{n-j}\big)\big(1-\xi\big)}. \tag{0.1.3.14}$$

For example in this case $A = HHTHH$ from the equations

$$1 + Y(T+H) = X + Y, \quad YHHTHH = X(1 + HHT + HHTH),$$

we obtain the expression

$$\mathsf{F}(\xi) = \frac{p^4 q\xi^5}{p^4 q\xi^5 + (1 + p^2 q\xi^3 + p^3 q\xi^4)(1-\xi)},$$

for the generating function of the time of the first appearance of $HHTHH$. From (0.1.3.13) one easily obtains the expectation and variance of τ. In fact we obtain

$$\mathsf{E}[\tau] = \frac{1+\sum_{j=1}^{n-1} A^j(1)\Delta_{n-j}}{A(1)}, \quad \mathsf{Var}[\tau] = \mathsf{E}[\tau]^2 - \frac{1+\sum_{j=1}^{n-1}(2j-1)A^j\Delta_{n-j}}{A(1)}. \tag{0.1.3.15}$$

In principle it is possible to obtain the generating function for the time of the first appearance of any given pattern of H's and T's by repeated conditioning as explained in the preceding examples. However, it is more beneficial to introduce a more efficient machinary for this calculation. The idea is most clearly explained by following through an example. Another application of this idea is given in the subsection on Patterns in Coin Tossing.

Suppose we want to compute the time of the first appearance of the pattern A, for example, $A = HHTHH$. We treat H and T as non-commuting indeterminates. We let X be the formal sum of all finite sequences (i.e., monomials in H and T) which end with the first appearance of the pattern A. We will doing formal algebraic operations on these formal sums in two non-commuting variables H and T, and also introduce 0 as the zero element which when multiplied by any quantity gives 0, and is the additive identity. In the case of the pattern $HHTHH$ we have

$$\begin{aligned} X \ = \ & HHTHH \ + \ HHHTHH \ + \\ & THHTHH \ + \ HHHHTHH + \\ & HTHHTHH + THHHTHH + \\ & TTHHTHH + \ \ \ldots \end{aligned}$$

Similarly let Y be the formal sum of all sequences (including the empty sequence which is represented by 1) which do not contain the given pattern A. For instance for $HHTHH$ we get

$$\begin{aligned} Y &= 1 + H + T + HH + HT + TH + TT + \\ &\quad \ldots + HHTHT + HHTTH + \ldots \end{aligned}$$

There is an obvious relation between X and Y independently of the chosen pattern, namely,

$$1 + Y(H+T) = X + Y. \tag{0.1.3.16}$$

The verification of this identity is almost trivial and is accomplished by noting that a monomial summand of $X+Y$ of length l either contains the given pattern for the first time at its end or does not contain it, and then looking at the first $n-1$ elements of the monomial. There is also another linear relation between X and Y which depends on the nature of the the desired pattern. Denote a given pattern by A and let A^j (resp. A_j) denote the first j elements of the pattern starting from right (respectively left). Thus for $HHTHH$ we get

$$\begin{aligned} &A^1 = H, \quad A^2 = HH, \quad A^3 = THH, \quad A^4 = HTHH; \\ &A_1 = H, \quad A_2 = HH, \quad A_3 = HHT, \quad A_4 = HHTH. \end{aligned}$$

Let Δ_j be 0 unless $A_j = A^j$ in which case it is 1. We obtain

$$YA = S(1 + A^1\Delta_{n-1} + A^2\Delta_{n-2} + \ldots + A^{n-1}\Delta_1). \tag{0.1.3.17}$$

For example in this case we get

$$YHHTHH = S(1 + A^3HH + A^4H).$$

Some experimentation will convince the reader that this identity is really the content of conditioning argument involved in obtaining the generating function for the time of first occurrence of a given pattern. At any rate its validity is easy to see. Equations (0.1.3.16) and (0.1.3.17) give us two linear equations which we can solve easily to obtain expressions for X and Y. Our primary interest in the expression for X. Therefore substituting for Y in (**??**) from (**??**) we obtain

$$A(1-X) = X\Big[A + \big(1 + \sum_{j=1}^{n-1} A^j\Delta_j\big)\big(1 - H - T\big)\Big] \tag{0.1.3.18}$$

which gives an expression for X. Now assume H appears with probability p and T with probability $q = 1-p$. Since X is the formal sum of all finite sequences ending in the first appearance of the desired pattern, by substituting $p\xi$ for H and $q\xi$ for T in the expression for X we obtained the desired probability generating function F (for the time τ of the first appearance of the pattern A). Denoting the result of this substitution in $A^j, A, \ldots$ by $A^j(\xi), A(\xi), \ldots$ we obtain

$$\mathsf{F}(\xi) = \frac{A(\xi)}{A(\xi) + \left(1 + \sum_{j=1}^{n-1} A^j(\xi)\Delta_{n-j}\right)\left(1-\xi\right)}. \tag{0.1.3.19}$$

For example in this case $A = HHTHH$ from the equations

$$1 + Y(T+H) = X + Y, \quad YHHTHH = X(1 + HHT + HHTH),$$

we obtain the expression

$$\mathsf{F}(\xi) = \frac{p^4 q\xi^5}{p^4 q\xi^5 + (1 + p^2 q\xi^3 + p^3 q\xi^4)(1-\xi)},$$

for the generating function of the time of the first appearance of $HHTHH$. From (0.1.3.18) one easily obtains the expectation and variance of τ. In fact we obtain

$$\mathsf{E}[\tau] = \frac{1+\sum_{j=1}^{n-1} A^j(1)\Delta_{n-j}}{A(1)}, \quad \mathsf{Var}[\tau] = \mathsf{E}[\tau]^2 - \frac{1+\sum_{j=1}^{n-1}(2j-1)A^j\Delta_{n-j}}{A(1)}. \tag{0.1.3.20}$$

There are elaborate mathematical techniques for obtaining information about a sequence of quantities of which a generating function is known. Here we just demonstrate how by a simple argument we can often deduce good approximation to a sequence of quantities q_n provided the generating function $\mathsf{Q}(\xi) = \sum_n q_n \xi^n$ is a rational function

$$\mathsf{Q}(\xi) = \frac{U(\xi)}{V(\xi)},$$

with $\deg U < \deg V$. For simplicity we further assume that the polynomial V has distinct roots $\alpha_1, \cdots, \alpha_m$ so that $\mathsf{Q}(\xi)$ has a partial fraction expansion

$$\mathsf{Q}(\xi) = \sum_{j=1}^{m} \frac{b_j}{\xi - \alpha_j}, \quad \text{with} \;\; b_j = \frac{-U(\alpha_j)}{V'(\xi_j)}.$$

Expanding $\frac{1}{\alpha_j - \xi}$ in a geometric series

$$\frac{1}{\alpha_j - \xi} = \frac{1}{\alpha_j}\frac{1}{1-\frac{\xi}{\alpha_j}} = \frac{1}{\alpha_j}[1 + \frac{\xi}{\alpha_j} + \frac{\xi^2}{\alpha_j^2} + \cdots]$$

we obtain the following expression for q_n:

$$q_n = \frac{b_1}{\alpha_1^{n+1}} + \frac{b_2}{\alpha_2^{n+1}} + \cdots + \frac{b_m}{\alpha_m^{n+1}} \tag{0.1.3.21}$$

To see how (0.1.3.21) can be used to give good approximations to the actual values of q_n's, assume $|\alpha_1| < |\alpha_j|$ for $j \neq 1$. Then we use the approximation $q_n \sim \frac{b_1}{\alpha_1^{n+1}}$.

Example 0.1.3.5 To illustrate the above idea of using partial fractions consider example 0.1.3.4 above. We can write the generating function (0.1.3.10) for the time of first appearance of pattern of m consecutive H's in the form

$$\mathsf{F}_X(\xi) = \frac{p^m \xi^m}{1 - q\xi(1 + p\xi + \cdots + p^{m-1}\xi^{m-1})}.$$

Denoting the denominator by $Q(\xi)$, we note that $Q(1) > 0$, $\lim_{\xi\to\infty} Q(\xi) = -\infty$ and Q is a decreasing function of $\xi \in \mathbf{R}_+$. Therefore Q has a unique positive root $\alpha > 1$. If $\gamma \in \mathbf{C}$ with $|\gamma| \leq \alpha$, then

$$|q\gamma(1 + p\gamma + \cdots + p^{m-1}\gamma^{m-1})| \leq |q\alpha(1 + p\alpha + \cdots + p^{m-1}\alpha^{m-1})| = 1,$$

with = only if all the terms have the same argument and $|\gamma| = \alpha$. It follows that α is the root of $Q(\xi) = 0$ with smallest absolute value. Applying the procedure described above we obtain the approximation

$$F_l \sim \frac{(\alpha - 1)(1 - p\alpha)}{(m + 1 - m\alpha)q}\alpha^{-l-1},$$

where F_l is the probability that first of pattern $H \cdots H$ is at time l so that $\mathsf{F}_X(\xi) = \sum F_l \xi^l$. This is a good approximation. For instance for $m = 2$ and $p = \frac{1}{2}$ we have $F_5 = .09375$ and the above approximation gives $F_5 \sim .09579$, and the approximation improves as l increases. ♠

For a sequence of real numbers $\{f_j\}_{j\geq\circ}$ satisfying a linear recursion relation, for example,

$$\alpha f_{j+1} + \beta f_j + \gamma f_{j-1} = 0, \tag{0.1.3.22}$$

it is straighforward to explicitly compute the generating function $\mathsf{F}(\xi)$. In fact, it follows from (0.1.3.22) that

$$\alpha\mathsf{F}(\xi) + \beta\xi\mathsf{F}(\xi) + \gamma\xi^2\mathsf{F}(\xi) = \alpha f_\circ + (\alpha f_1 + \beta f_\circ)\xi.$$

Solving this equation for F we obtain

$$\mathsf{F}(\xi) = \frac{\alpha f_\circ + (\alpha f_1 + \beta f_\circ)\xi}{\alpha + \beta\xi + \gamma\xi^2}. \tag{0.1.3.23}$$

Here we assumed that the coefficients α, β and γ are independent of j. It is clear that the method of computing $\mathsf{F}(\xi)$ is applicable to more complex recursion relations as long as the coefficients are independent of j. If these coefficients have simple dependence on j, e.g., depend linearly on j, then we can obtain a differential equation for F. To demonstrate this the point we consider the following simple example with probabilistic implications:

Example 0.1.3.6 Assume we have the recursion relation (the probabilistic interpretation of which is given shortly)

$$(j+1)f_{j+1} - jf_j - f_{j-1} = 0, \quad j = 2, 3, \cdots \tag{0.1.3.24}$$

Let $\mathsf{F}(\xi) = \sum_{j=1}^{\infty} f_j \xi^j$. To compute F note

$$\begin{array}{rcccccccc} \mathsf{F}' & = & f_1 & + & 2f_2\xi & + & 3f_3\xi^2 & + \cdots \\ \xi\mathsf{F}' & = & & & f_1\xi & + & 2f_2\xi^2 & + \cdots \\ \xi\mathsf{F} & = & & & & & f_1\xi^2 & + \cdots \end{array}$$

It follows that

$$(1-\xi)\frac{d\mathsf{F}}{d\xi} - \xi\mathsf{F} = f_1 + (f_1 + 2f_2)\xi. \tag{0.1.3.25}$$

As an application to probability we consider the *matching problem* where n balls numbered $1, 2, \cdots, n$ are randomly put in boxes numbered $1, 2, \cdots, n$; one in each box. Let f_n be the probability that the numbers on balls and boxes containing them have no matches. To obtain a recursion relation for f_j's let A_j be the event of no matches, and B_j be the event that the first ball is put in a box with a non-matching number. Then

$$f_{j+1} = P[A_{j+1} \mid B_{j+1}]\frac{j}{j+1}. \tag{0.1.3.26}$$

On the other hand,

$$P[A_{j+1} \mid B_{j+1}] = \frac{1}{j}f_{j-1} + \frac{j-1}{j}P[A_j \mid B_j]. \tag{0.1.3.27}$$

Equations (0.1.3.26) and (0.1.3.27) imply validity of (0.1.3.24) and (0.1.3.25) with

$$f_1 = 0; \quad f_2 = \frac{1}{2}. \tag{0.1.3.28}$$

Therefore to compute the generating function $\mathsf{F}(\xi)$ we have to solve the differential equation

$$(1-\xi)\frac{d\mathsf{F}}{d\xi} = \xi\mathsf{F} + \xi,$$

with $\mathsf{F}(0) = 0$. Making the substitution $H(\xi) = (1-\xi)\mathsf{F}(\xi)$, the differential equation becomes $H' + H = \xi$ which is easily solved to yield

$$\mathsf{F}(\xi) = \frac{e^{-\xi}}{1-\xi} - 1.$$

Expanding as a power series, we obtain after a simple calculation

$$f_k = \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^k}{k!}. \qquad (0.1.3.29)$$

Thus for k large, the probability of no matches is approximately $\frac{1}{e}$. Of course one can derive (0.1.3.29) by a more elementary (but substantially the same) argument. ♠

Example 0.1.3.7 Consider the simple random walk on the integer which moves one unit to the right with probability p and one unit to the left with probability $q = 1-p$ and is initially at 0. Let p_l denote the probability that the walk is at 0 at time l and $P_{\circ\circ}(\xi) = \sum p_l\xi^l$ denote the corresponding generating function. It is clear that $p_{2l+1} = 0$ and

$$p_{2l} = \binom{2l}{l} p^l q^l.$$

Therefore

$$\mathsf{P}_{\circ\circ}(\xi) = \frac{1}{\sqrt{1-4pq\xi^2}}.$$

Let F_l denote the probability that first return to 0 occurs at time l. It follows that theorem 0.1.3.1 that

$$\mathsf{F}_{\circ\circ}(\xi) \stackrel{\text{def.}}{=} \sum F_l\xi^l = 1 - \sqrt{1-4pq\xi^2}.$$

Consequently the probability of eventual return to the origin is $1-|p-q|$. Let the random variable $T_{\circ\circ}$ be the time of the first return to the origin. Let $p = q = \frac{1}{2}$. Differentiating $F_{\circ\circ}(\xi)$ with respect to ξ and setting $\xi = 1$ we obtain

$$\mathsf{E}[T_{\circ\circ}] = \infty.$$

In other words, although with probability 1 every path will return to the origin, the expectation of the time return is infinite. For $p \neq q$ there is probability $|p-q| > 0$ of never returning to the origin and therefore the expected time of return to the origin is again infinite. ♠

A consequence of the the computation of the generating function $\mathsf{F}_{\circ\circ}(\xi)$ is the classification of the states of the simple random walk on $\mathbb{Z}$:

Corollary 0.1.3.4 *For $p \neq q$ the simple random walk on $\mathbb{Z}$ is transient. For $p = q = \frac{1}{2}$, every state is recurrent.*

Proof - the first statement follows from the fact the with probability $|q - p| > 0$ a path will never return to the origin. Setting $p = q = \frac{1}{2}$ and $\xi = 1$ in $\mathsf{F}_{\circ\circ}(\xi)$ we obtain $\mathsf{F}_{\circ\circ}(1) = 1$ proving recurrence of 0 and therefore all states. ♣

Example 0.1.3.8 Consider the simple random walk $S_1, S_2, \cdots$ on $\mathbb{Z}$ where $X_\circ = 0$, $X_j = \pm 1$ with probabilities p and $q = 1 - p$, and $S_l = X_\circ + X_1 + \cdots + X_l$. Let T_n be the random variable denoting the time of first visit to state $n \in \mathbb{Z}$ given that $X_\circ = 0$. In this example we investigate the generating function for T_n, namely,

$$\mathsf{F}_{\circ n}(\xi) = \sum_{l=1}^{\infty} P[T_n = l]\xi^l$$

be its probability generating function. It is clear that

$$P[T_n = l] = \sum_{j=1}^{l-1} P[T_{n-1} = l - j]P[T_1 = j].$$

From this identity it follows that

$$\mathsf{F}_{\circ n}(\xi) = [\mathsf{F}_{\circ 1}(\xi)]^n. \tag{0.1.3.30}$$

which reduces the computation of $\mathsf{F}_{\circ n}$ to that of $\mathsf{F}_{\circ 1}$. It is immediate that

$$P[T_1 = l] = \begin{cases} qP[T_2 = l - 1], & \text{if } l > 1; \\ P[T_{\circ 1} = 1] = p, & \text{if } l = 1. \end{cases}$$

This together with (0.1.3.30) imply

$$\mathsf{F}_{\circ 1}(\xi) = p\xi + q\xi[\mathsf{F}_{\circ 1}(\xi)]^2.$$

Solving the quadratic equation we obtain

$$\mathsf{F}_{\circ 1}(\xi) = \frac{1 - \sqrt{1 - 4pq\xi^2}}{2q\xi}. \tag{0.1.3.31}$$

Substituting $\xi = 1$ we see that the probability that the simple random walk ever visits $1 \in \mathbb{Z}$ is $\min(1, \frac{p}{q})$. ♠

Example 0.1.3.9 We shown that the simple symmetric random walk on $\mathbb{Z}$ is recurrent and exercise 0.1.3.11 show that the same conclusion is valid for for the simple symmetric random walk on $\mathbb{Z}^2$. In this example we consider the simple symmetric random walk on $\mathbb{Z}^3$. To carry out the analysis we make use of an elementary fact regarding multinomial coefficients. Let $\binom{N}{n_1\ n_2 \cdots\ n_k}$ denote the multinomial coefficient

$$\binom{N}{n_1\ n_2 \cdots\ n_k} = \frac{N!}{n_1!n_2! \cdots n_k!},$$

where $N = n_1 + n_2 + \cdots + n_k$ and all integers n_j are non-negative. Just as in the case of binomial coefficients the maximum of $\binom{N}{n_1\ n_2 \cdots\ n_k}$ occurs when the the quantities $n_1, \cdots, n_k$ are (approximately) equal. We omit the proof of this elementary fact and make use of it for $k = 3$. To determine recurrence/transience of the random walk on $\mathbb{Z}^3$ we proceed as before by looking at $\sum P^{(l)}_{\circ\circ}$. We have $P^{(2l+1)}_{\circ\circ} = 0$ and

$$P^{(2l)}_{\circ\circ} = \sum_{i+j+k=l} \binom{2l}{i\ i\ j\ j\ k\ k} \frac{1}{6^{2l}}.$$

Multiplying the above expression by $\frac{(l!)^2}{(l!)^2}$ and simplifying we obtain

$$P^{(2l)}_{\circ\circ} = \sum_{i,j=0}^{l} \binom{2l}{l} \frac{l!^2}{[i!j!(l-i-j)!]^2} \frac{1}{6^{2l}}.$$

To estimate this expression, we make use of the obvious fact

$$1 = (\frac{1}{3} + \frac{1}{3} + \frac{1}{3})^l = \sum_{i,j=0}^{l} \frac{l!}{i!j!(l-i-j)!} \frac{1}{3^l}.$$

This allows us to write

$$P^{(2l)}_{\circ\circ} \leq \binom{2l}{l} \frac{1}{2^{2l}} \frac{1}{3^l} M_l,$$

where

$$M_l = \max_{0 \leq i+j \leq l} \frac{l!}{i!j!(l-i-j)!}.$$

Using the fact that the maximum M_l is achieved for approximately $i = j = \frac{l}{3}$, we obtain

$$P^{(2l)}_{\circ\circ} \leq \frac{l!}{[(l/3)!]^3 2^{2l} 3^l} \binom{2l}{l}.$$

Now recall Stirling's formula

$$n! = \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} e^{\rho(n)}, \quad where \frac{1}{12(n+\frac{1}{2})} < \rho(n) < \frac{1}{12n}. \tag{0.1.3.32}$$

Applying Stirling's formula we obtain the bound

$$\sum_l P_{oo}^{(l)} = \sum_l P_{oo}^{(2l)} \leq \gamma \sum_l \frac{1}{l^{3/2}} < \infty,$$

for some constant γ. Thus 0 and therefore all states in the simple symmetric random walk on $\mathbb{Z}^3$ are transient. By a similar argument, the simple symmetric random walk is transient in dimensions ≥ 3. ♡

EXERCISES

Exercise 0.1.3.1 *Let $P = (P_{ij})$ be a (possibly infinite) Markov matrix, and $P^l = (P_{ij}^{(l)})$. Show that if j is a transient state then for all i we have*

$$\sum_l P_{ij}^{(l)} < \infty.$$

Exercise 0.1.3.2 *Show that if states i and j of a Markov chain communicate and they are recurrent, then $F_{ij} = 1$, i.e., with probability 1, every path starting at i will visit j.*

Exercise 0.1.3.3 *Consider the Markov chain on the vertices of a square with vertices $A = (0,0)$, $B = (1,0)$, $C = (0,1)$ and $D = (1,1)$, where one moves along an horizontal edge with probability p and along a vertical edge with probability $q = 1-p$, and is initially at A. Let F_l denote the probability that first return to state A occurs at time l, and $p_l = P_{AA}^{(l)}$ denote the probability that that the Markov chain is in state A at time l. Show that the generating functions functions $\mathsf{F}(\xi) = \sum F_l\xi^l$ and $P(\xi) = \sum p_l\xi^l$ are*

$$P(\xi) = \frac{1}{2}\left(\frac{1}{1-(1-2p)^2\xi^2} + \frac{1}{1-\xi^2}\right), \quad F(\xi) = \frac{P(\xi)-1}{P(\xi)}.$$

Exercise 0.1.3.4 *Consider the coin tossing experiment where H's appear with probability p and T's with probability $q = 1-p$. Let S_n denote the number of T's before the appearance of the n^{th} H. Show that the probability generating function for S_n is*

$$\mathsf{E}[\xi^{S_n}] = \left(\frac{p}{1-q\xi}\right)^n.$$

Exercise 0.1.3.5 *Consider the coin tossing experiment where H's appear with probability p and T's with probability $q = 1-p$. Compute the probaility generating function for the time of first appearance of the following patterns:*

1. *THH;*
2. *$THHT$;*
3. *$THTH$.*

Exercise 0.1.3.6 *Show that the generating function for the pattern $HTTHT$ is We can easily solve this for $\mathsf{E}[\xi^T]$:*

$$\mathsf{F}_T(\xi) = \mathsf{E}[\xi^T] = \frac{p^2q^3\xi^3}{1 + p^2q^3\xi^5 + pq^2\xi^3 - \xi - pq^2\xi^4}.$$

Exercise 0.1.3.7 *Let a_n denote the number of ways an $(n+1)$-sided convex polygon with vertices $P_\circ, P_1, \cdots, P_n$ can be decomposed into triangles by drawing non-intersecting line segments joining the vertices.*

1. *Show that*

$$a_n = a_1 a_{n-1} + a_2 a_{n-2} + \cdots + a_{n-1} a_1, \quad \text{with} \quad a_1 = 1.$$

2. *Let $\mathsf{A}(\xi) = \sum_{n=1}^{\infty} a_n \xi^n$ be the corresponding generating function. Show that $\mathsf{A}(\xi)$ satisfies the quadratic relation*

$$\mathsf{A}(\xi) - \xi = [\mathsf{A}(\xi)]^2.$$

3. *Deduce that*

$$\mathsf{A}(\xi) = \frac{1 - \sqrt{1 - 4\xi}}{2}, \quad \text{and} \quad a_n = \frac{1}{n}\binom{2(n-1)}{n-1}.$$

Exercise 0.1.3.8 *Let q_n denote the probability that in n tosses of a fair coin we do not get the sequence HHH.*

1. *Use conditioning to obtain the recursion relation*

$$q_n = \frac{1}{2}q_{n-1} + \frac{1}{4}q_{n-2} + \frac{1}{8}q_{n-3}.$$

2. *Deduce that the generating function $Q(\xi) = \sum q_j \xi^j$ is*

$$Q(\xi) = \frac{2\xi^2 + 4\xi + 8}{-\xi^3 - 2\xi^2 - 4\xi + 8}.$$

3. *Show that the root of the denominator of $Q(\xi)$ with smallest absolute value is $\alpha_1 = 1.0873778$.*

4. *Deduce that the approximations $q_n \sim \frac{1.23684}{(1.0873778)^{n+1}}$ yield, for instance,*

$$q_3 \sim .8847, \ q_4 \sim .8136, \ q_{12} \sim .41626$$

(The actual values $q_3 = .875$, $q_4 = 8125$ and $q_{12} = .41626$.)

Exercise 0.1.3.9 *In a coin tossing experiment heads appear with probability p. Let A_n be the event that there are an even number of heads in n trials, and a_n be the probability of A_n. State and prove a linear relation between a_n and a_{n-1}, and deduce that*

$$\sum a_n\xi^n = \frac{1}{2}\Big(\frac{1}{1-\xi} + \frac{1}{1-(1-2p)\xi}\Big).$$

Exercise 0.1.3.10 *In a coin tossing experiment heads appear with probability p and $q = 1-p$. Let X denote the time of first appearance of the pattern HTH. Show that the probability generating function for X is*

$$\mathsf{F}_X(t) = \frac{p^2q\xi^3}{1-t+pq\xi^2-pq^2\xi^3}.$$

Exercise 0.1.3.11 *Consider the random walk on $\mathbb{Z}^2$ where a point moves from (i,j) to any of the points $(i\pm 1, j), (i, j\pm 1)$ with probability $\frac{1}{4}$. Show that the random walk is recurrent.* (Use the idea of example 0.1.3.9.)

0.1.4 Stationary Distribution

It was noted earlier that one of the goals of the theory of Markov chains is to establish that under certain hypotheses, the distribution of states tends to a limiting distribution. If indeed this is the case then there is a row vector $\pi = (\pi_1, \pi_2, \cdots)$ with $\pi_j \geq 0$ and $\sum \pi_j = 1$, such that $\pi^{(\circ)} P^n \rightarrow \pi$ as $n \rightarrow \infty$. Here $\pi^{(\circ)}$ denotes the initial distribution. If such π exists, then it has the property $\pi P = \pi$. For this reason we define the *stationary* or *equilibrium distribution* of a Markov chain with transition matrix P (possibly infinite matrix) as a row vector $\pi = (\pi_1, \pi_2, \cdots)$ such that

$$\pi P = \pi, \quad \text{with} \quad \pi_j \geq 0, \text{ and } \sum_{j=1}^{\infty} \pi_j = 1. \tag{0.1.4.1}$$

The existence of such a vector π does not imply that the distribution of states of the Markov chain necessarily tends to π as shown by the following example:

Example 0.1.4.1 Consider the Markov chain given by the 3×3 transition matrix $P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$. Then for $\pi^{(\circ)} = (1, 0, 0)$ the Markov chain moves between the states $1, 2, 3$ periodically. On the other hand, for $\pi^{(\circ)} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ $\pi^{(\circ)} P = \pi^{(\circ)}$. So for periodic Markov chains, stationary distribution has no implication about a limiting distribution. This examples easily generalizes to $n \times n$ matrices. Another case to keep in mind when the matrix P admits of a decomposition $P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$. Each P_j is necessarily a stochastic matrix, and if $\pi^{(j)}$ is a stationary distribution for P_j, then $(t\pi^{(1)}, (1-t)\pi^{(2)})$ is one for P, for $0 \leq t \leq 1$. Thus the long term behavior of this chain depends on the initial distribution. ♠

Our goal is to identify a set of hypotheses which imply the existence and uniqueness of the stationary distribution π and such that the long term behavior of the Markov chain is accurately represented by π. To do so we first discuss the issue of the existence of solution to (0.1.4.1) for a finite state Markov chain. Let $\mathbf{1}$ denote the column vector of all 1's, then $P\mathbf{1} = \mathbf{1}$ and 1 is an eigenvalue of P. This implies the existence of a row vector $v = (v_1, \cdots, v_n)$ such that $vP = v$, however, *a priori* there is no guarantee that the eigenvector v can be chosen such that all its components $v_j \geq 0$. Therefore we approach the problem differently. The existence of π satisfying (0.1.4.1) follows from a very general theorem with a simple statement and diverse applications and generalizations. We state the theorem without proof since its proof has no relevance to stochastic processes.

Theorem 0.1.4.1 (Brouwer Fixed Point Theorem) - *Let $K \subset \mathbf{R}^n$ be a convex compact*[2] *set, and $F : K \rightarrow K$ be a continuous map. Then there is $x \in K$ such that $F(x) = x$.*

Note that only continuity of F is required for the validity of the theorem although we apply it for F linear. To prove existence of π we let

$$K = \{(x_1, \cdots, x_n) \in \mathbf{R}^n \mid \sum x_j = 1,\ x_j \geq 0\}.$$

Then K is a compact convex set and let F be the mapping $v \rightarrow vP$. The fact that P is a stochastic matrix implies that P maps K to itself. In fact, for $v \in K$ let $w = (w_1, \cdots, w_n) = vP$, then $w_j \geq 0$ and

$$\begin{aligned} \textstyle\sum_i w_i &= \textstyle\sum_{i,j} v_j P_{ji} \\ &= \textstyle\sum_j v_j \sum_i P_{ij} \\ &= \textstyle\sum_j v_j \\ &= 1, \end{aligned}$$

proving $w \in K$. Therefore Brouwer's Fixed Point Theorem is applicable to ensure existence of π for a finite state Markov chain.

In order to give a probabilistic meaning to the entries π_j of the stationary distribution π, we recall some notation. For states $i \neq j$ let T_{ij} be the random variable of first hitting time of j starting at i. Denote its expectation by μ_{ij}. If $i = k$ then denote the expectation of first return time to i by μ_i and define $\mu_{ii} = 0$.

Proposition 0.1.4.1 *Assume a solution to (0.1.4.1) exists for the Markov chain defined by the (possibly infinite) matrix P, and furthermore*

$$\mu_{ij} < \infty, \quad \mu_j < \infty \quad \text{for all } i, j.$$

Then $\pi_i \mu_i = 1$ for all i.

Proof - For $i \neq j$ we have

$$\begin{aligned} \mu_{ij} &= \mathsf{E}[\mathsf{E}[T_{ij} \mid X_1]] \\ &= 1 + \textstyle\sum_k P_{ik} \mu_{kj}, \end{aligned}$$

and

$$\mu_j = 1 + \sum_k P_{jk} \mu_{kj}.$$

[2]A closed and bounded subset of $\mathbf{R}^n$ is called *compact*. $K \imath \mathbf{R}^n$ is convex if for $x, y \in K$ the line segment $tx + (1-t)y$, $0 \leq t \leq 1$, lies in K. The assumption of convexity can be relaxed but compactness is essential.

The two equations can be written simply as

$$\mu_{ij} + \delta_{ij}\mu_j = 1 + \sum_k P_{ik}\mu_{kj}, \quad \text{where } \delta_{ij} = \begin{cases} 1 & \text{if } i = j\ ; \\ 0 & \text{otherwise.} \end{cases} \tag{0.1.4.2}$$

Multiplying (0.1.4.2) by π_i and summing over i (j is fixed) we obtain

$$\begin{aligned} \textstyle\sum_i \pi_i\mu_{ij} + \sum_i \pi_i\delta_{ij}\mu_j &= 1 + \textstyle\sum_i\sum_k \pi_i P_{ik}\mu_{kj} \\ &= 1 + \textstyle\sum_k \pi_k\mu_{kj}. \end{aligned}$$

Cancelling $\sum_i \pi_i\mu_{ij}$ from both sides we get the desired result. ♣

The proposition in particular implies that if the quantities μ_{ij} and μ_j are finite, then a stationary distribution, if exists, is necessarily unique. Clearly if $P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$ then some of the quantities μ_{ij} will be infinite. Since for finite Markov chains, the existence of a solution to (0.1.4.1) has already been established, the main question is the determination of finiteness of μ_{ik} and μ_k and when the stationary distribution reflects the long term behavior of the Markov chain. It is convenient to introduce two definitions. Let $X_\circ, X_1, X_2, \cdots$ be a Markov chain with state space S and transition matrix P. Let $S = T \cup C_1 \cup \cdots$ be the decomposition of the state space into transient and recurrent classes as described in proposition 0.1.2.1. The Markov chain or its transition matrix is called *positive* (resp. *almost positive*)[3] if all the entries P_{ij} (resp. all the entries P_{ij} with i, j recurrent states) are positive. A Markov chain or its transition matrix P is called *regular* (resp. *almost regular*) if P^l is positive (resp. almost positive) for some l. Clearly if P^l is positive (resp. almost positive) then P^m is also positive (resp. almost positive) for all $m \geq l$. For a finite state almost regular Markov chain (so that there is at least one recurrent state by exercise 0.1.2.6) the set of recurrent classes form one equivalence class. The set of transient state may or may not be empty.

To understand the long term behavior of the Markov chain, we show that under certain hypotheses the entries of the matrix P^l have limiting values

$$\lim_{l\to\infty} P_{ij}^{(l)} = p_j. \tag{0.1.4.3}$$

Notice that the value p_j is independent of i so the matrix P^l tends to a matrix P^∞ with the same entry p_j along j^{th} column. This implies that if the initial distribution is any vector $\pi^\circ = (\pi_1^\circ, \pi_2^\circ, \cdots, \pi_N^\circ)$ then

$$\pi^\circ P^\infty = (p_1, \cdots, p_N).$$

Therefore the long term behavior of the Markov chain is accurately reflected in the vector $(p_1, \cdots, p_N)$ and $p_j = \pi_j$. More precisely we prove the following theorem:

[3]The terminology almost positive or almost regular is not standard.

Theorem 0.1.4.2 *Let P be the transition matrix of a finite state regular Markov chain. Then*

$$\lim_{l\to\infty} P_{ij}^{(l)} = \pi_j.$$

The theorem remains valid for almost regular chains with a minor modification. To fix notation, we let the $\{1, 2, \cdots, N\}$ denote the state space of a finite Markov chain $X_\circ, X_1, X_2, \cdots$. Assume states $\{1, \cdots, n\}$ are transient and $\{n+1, \cdots, N\}$ are recurrent. The transition matrix of the Markov chain is necessarily of the form

$$P = \begin{pmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{pmatrix},$$

and P_{22} is an $(N-n)\times(N-n)$ stochastic matrix and P_{22} is regular. Let $(\pi_{n+1}, \cdots, \pi_N)$ be the stationary distribution of of P_{22}. Then

$$\lim_{l\to\infty} P_{ij}^{(l)} = \begin{cases} 0 & \text{if } j \leq n; \\ \pi_j & \text{if } j \geq n+1. \end{cases}$$

The proof given below works without change for this slightly more general case.

The proof of the theorem requires proposition 0.1.4.2 below and the idea of coupling.

Proposition 0.1.4.2 *Let $Z_\circ, Z_1, Z_2, \cdots$ be an almost regular Markov chain with finite state space $\{1, 2, \cdots, M\}$ and transition matrix P where we are assuming states $m+1, \cdots, M$ are recurrent. Then there are constants $c < \infty$ and $\lambda < 1$ such that*

$$P[T_{iM} > l] < c\lambda^l.$$

In particular all moments of T_{iM} exist and $P[T_{iM} > l] \to 0$ as $l \to \infty$.

We postpone the proof of the proposition and introduce a version of *coupling* in order to prove theorem 0.1.4.2. Let us describe the idea of coupling. Assume two Markov chains $X_1, X_2, \cdots$ and $Y_1, Y_2, \cdots$ have the same law (i.e., the same state space S and transition matrix) but start with different initial distributions. We combine the Markov chain together in the form (X_j, Y_j) with state space $S \times S$. The description of the underlying probability space for the coupled Markov chain (X_j, Y_j) requires some care. Assume $S = \{1, 2, \cdots, N\}$ (for simplicity although finiteness is not necessary). In our description of the underlying probability space of a Markov chain, e.g. $X_1, X_2, \cdots$ or $Y_1, Y_2, \cdots$ we used N-sided dice. If a path for the coupled chain is in state (a, b), $a \neq b$, at time l then we roll dice corresponding to states a and b. The path moves to state (c, d) at time $l + 1$ if the outcomes of the rolls of the dice are c and d respectively. If the path is in state (a, a) at time l, then only one die (corresponding to state a) is rolled and the path moves to state (b, b) where b is the outcome of the roll of the die corresponding to state a. Thus if the coupled chain (X_j, Y_j) enters the set $D = \{(a, a) \mid a \in S\}$ at time l then in all subsequent times it will be in D. One often refers to D as the *diagonal*. The reason the idea of coupling is useful is that if we know the development of a Markov chain for one initial distribution (for example, for Y_j), and if we know that the two chains merge, then we can deduce the long term term behavior of X_j. We will now use this idea to prove theorem 0.1.4.2. Notice that if P is regular, then the coupled chain (X_j, Y_j) is almost regular. In fact, we have

Lemma 0.1.4.1 *Assume P is regular. Then the coupled chain (X_j, Y_j) is almost regular; the transient states are $T = \{(a, b) \mid a \neq b, a, b \in S\}$ and the recurrent states are $D = \{(a, a) \mid a \in S\}$.*

Proof - Assume $X_\circ = i$ and let $j \neq i$ and a be states. Let $\Omega^{(l)}_{i,j;a}$ be the set of paths of the coupled chain which at time 0 are in (i, j) and at time l in (a, a). Then

$$P[\Omega^{(l)}_{i,j;a}] \geq \pi_j P^{(l)}_{ia} P^{(l)}_{ja} > 0,$$

for l sufficiently large. Therefore the diagonal is accessible from any state. It is clear that non-diagonal states are not accessible from the diagonal. The required result follows from regularity of the original Markov chain. ♣

Proof of Theorem 0.1.4.2 - Consider the coupled chain (X_j, Y_j) where we assume that the initial distribution $X_\circ = i$ and $Y_\circ = (\pi_1, \cdots, \pi_N)$. Let T denote the first hitting time of D. In view of lemma 0.1.4.1, with probability 1 paths of the coupled chain enter D. We have

$$\begin{aligned} |P^{(l)}_{ij} - \pi_j| &= |P[X_l = j] - P[Y_l = j]| \\ &\leq |P[X_l = j,\ T \leq l] - P[Y_l = j,\ T \leq l]| + \\ &\quad\ |P[X_l = j,\ T > l] - P[Y_l = j,\ T > l]|. \end{aligned}$$

It follows from propsition 0.1.4.2 and lemma 0.1.4.1 that $P[T > l] \to 0$ as $l \to \infty$. Therefore each term $P[X_l = j,\ T > l]$ and $P[Y_l = j,\ T > l]$ goes to zero as $l \to \infty$. If a path ω enters D at time l' then it remains in D for all $l \geq l'$. Since with probability 1 every path eventually enters D, the set of paths for which $X_l(\omega) \neq Y_l(\omega)$ will have arbitrarily small probability by taking l sufficiently large. Therefore

$$\lim_{l \to \infty} |P[X_l = j,\ T \leq l] - P[Y_l = j,\ T \leq l]| = 0$$

It follows that $\lim_{l \to \infty} P^{(l)}_{ij} = \pi_j$. ♣

We have shown that the stationary distribution exists for regular finite state Markov chains and the entries of the stationary distribution are the reciprocals of the expected return times to the corresponding states. We can in fact get more information from the stationary distribution. For example, for states a and i let $R_i(a)$ be the number of visits to state i before first return to a given that initially the Markov chain was in state a. $R_i(a)$ is a random variable and we let

$$\rho_i(a) = \mathsf{E}[R_i(a)].$$

We want to calculate $\rho_i(a)$. Observe

Lemma 0.1.4.2 *We have*

$$\rho_i(a) = \sum_{l=1}^{\infty} P[X_l = i, T_a \geq l \mid X_\circ = a]$$

where T_a is the first return time to state a.

Proof - Let $\Omega^{(l)}$ denote the set of paths which are in state i at time l, and first return to a occurs at time $l' > l$. Define the random variable I_l by

$$I_l(\omega) = \begin{cases} 1 & \text{if } \omega \in \Omega^{(l)}; \\ 0 & \text{otherwise.} \end{cases}$$

Then $R_i(a) = \sum_{l=1}^{\infty} I_l$. Conequently,

$$\rho_i(a) = \sum_{l=1}^{\infty} \mathsf{E}[I_l] = \sum_{l=1}^{\infty} P[X_l = i, T_a \geq l \mid X_\circ = a]$$

as required. ♣

It is clear that

$$P[X_1 = i, T_a \geq 1 \mid X_\circ = a] = P_{ai}.$$

For $l \geq 2$ we use conditional probability

$$\begin{aligned} P[X_l = i, T_a \geq l \mid X_\circ = a] &= \sum_{j \neq a} P[X_l = i, T_a \geq l, X_{l-1} = j \mid X_\circ = a] \\ &= \sum_{j \neq a} P[X_l = i \mid T_a \geq l, X_{l-1} = j, X_\circ = a]. \\ &\qquad P[T_a \geq l-1, X_{l-1} = j \mid X_\circ = a] \\ &= \sum_{j \neq a} P_{ji} P[T_a \geq l-1, X_{l-1} = j \mid X_\circ = a]. \end{aligned}$$

Substituting in lemma 0.1.4.2 and noting $\rho_a(a) = 1$ we obtain

$$\begin{aligned}\rho_i(a) &= P_{ai} + \sum_{j\neq a} P_{ji} \sum_{l\geq 2} P[X_{l-1} = j, T_a \geq l-1 \mid X_\circ = a] \\ &= P_{ai} + \sum_{j\neq a} \rho_j(a) P_{ji} \\ &= \sum \rho_j(a) P_{ji},\end{aligned}$$

where the last summation is over all j including $j = a$. This means the vector $\rho = (\rho_1(a), \rho_2(a), \cdots)$ satisfies

$$\rho P = \rho, \quad \rho_i(a) \geq 0.$$

We now prove

Corollary 0.1.4.1 *Assume the Markov chain has a unique stationary distribution and the expected hitting times μ_i and μ_{ij} are finite. Then*

$$\rho_i(a) = \frac{\mu_a}{\mu_i}.$$

Proof - $\rho P = \rho$ and the hypotheses imply that ρ is a multiple of the stationary distribution. Since $\rho_a(a) = 1$ the required result follows. ♣

We need some preliminary considerations for the proof of proposition 0.1.4.2. First assume P is a regular $N \times N$ matrix. Define $(N-1) \times (N-1)$ matrices $Q^{(l)} = (Q^{(l)}_{ij})$, where $1 \leq i, j \leq N-1$ by

$$Q^{(l)}_{ij} = P[X_l = j,\ T_{iN} > l \mid X_\circ = i].$$

Since the indices $i, j \leq N-1$ we have $Q^{(1)}_{ij} = P_{ij}$ and $Q^{(l)} = (Q^{(1)})^l$, or equivalently,

$$Q^{(l)}_{ij} = \sum_{j_1\neq N} \sum_{j_2\neq N} \cdots \sum_{j_{l-1}\neq N} P_{ij_1} P_{j_1 j_2} \cdots P_{j_{l-1}N}. \tag{0.1.4.4}$$

We need two simple technical lemmas.

Lemma 0.1.4.3 *If P is positive then there is $\rho < 1$ such that*

$$\sum_{j=1}^{N-1} Q^{(l)}_{ij} < \rho^l,$$

and consequently $\sum_{l=1}^{\infty} \sum_{j=1}^{N-1} Q^{(l)}_{ij}$ converges.

Proof - Since P is positive

$$\sum_{j=1}^{N-1} P_{ij} \leq \rho < 1$$

for some ρ and all i. It follows from (0.1.4.4) that

$$\sum_{j=1}^{N-1} Q_{ij}^{(l)} \leq \rho^l.$$

The required result follows from the convergence of a geometric series. ♣

Lemma 0.1.4.4 *For a regular matrix P, $\sum_{j=1}^{N-1} Q_{ij}^{(l)}$ is a non-increasing function of l.*

Proof - Since $\sum_{j=1}^{N-1} Q_{ij}^{(1)} \leq 1$, we have

$$\sum_{j=1}^{N-1} Q_{ij}^{(l+1)} \leq \sum_{k=1}^{N-1} Q_{ik}^{(l)} \sum_{j=1}^{N-1} Q_{kj}^{(1)} \leq \sum_{j=1}^{N-1} Q_{ij}^{(l)}.$$

Thus $\sum_{j=1}^{N-1} Q_{ij}^{(l)}$ is a non-increasing function of l. ♣

Proof of proposition 0.1.4.2 First assume P regular. We have

$$P[T_{iN} > l] = \sum_{j=1}^{N-1} P[T_{iN} > l,\ X_l = j \mid X_\circ = i] = \sum_{j=1}^{N-1} Q_{ij}^{(l)}.$$

By regularity of the Markov chain, P^m is positive for some m. Lemma 0.1.4.3 (or more precisely its proof) implies that $\sum_{j=1}^{N-1} Q_{ij}^{(mn)} < \rho^n < 1$. By lemma 0.1.4.4 for $nm \leq l < (n+1)m$ we have

$$\sum_{j=1}^{N-1} Q_{ij}^{(l)} \leq \sum_{j=1}^{N-1} Q_{ij}^{(mn)} < (\rho^{\frac{n}{L}})^L < \lambda^L$$

for some $\lambda < 1$ and we need c to take care of the first m terms. This completes the proof of the proposition for P regular.

For the general case of almost regular it is only necessary to make minor adjustments to the above argument. Let $P = \begin{pmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{pmatrix}$ be the matrix of a almost regular Markov chain. Construct a new Markov chain with transition matrix

$$Q = \begin{pmatrix} P_{11} & q \\ 0 & 1 \end{pmatrix}$$

where q is a column vector of dimension m with entries $q_1, \cdots, q_m$. The quantities q_j are uniquely determined by P_{11} and the requirement that Q is a stochastic matrix. The Markov chain with transition matrix Q has one recurrent state, namely $m+1$, and m transient states. The matrix Q is the transition matrix of the Markov chain where all the recurrent states of P are collapsed into one state. This process facilitates the analysis. Write Q^l in the form

$$Q^l = \begin{pmatrix} P_{11}^l & q^{(l)} \\ 0 & 1 \end{pmatrix}$$

where $q^{(l)}$ is a column vector. It follows from regularity of the matrix P_{22} that for l sufficiently large all entries of the vector $q^{(l)}$ are positive. Then the arguments of leading to the proof of the proposition are applicable to show that

$$P[T'_{i\ m+1} > l] < c\lambda^l \tag{0.1.4.5}$$

where $T'_{i\ m+1}$ is the first hitting time of state $m+1$ starting at i for the Markov chain defined by Q. In other words the probability of the set of paths starting at the transient state i and first hitting a recurrent state at time $> l$ approaches zero like $c\lambda^l$ for some constants c and $\lambda < 1$ as $l \to \infty$. Combining (0.1.4.5) with the proposition for the case of regular P we immediately complete the proof of the proposition after possibly replacing c by a larger constant. ♣

EXERCISES

Exercise 0.1.4.1 *Consider three boxes* $\mathbf{1}, \mathbf{2}, \mathbf{3}$ *and three balls* A, B, C*, and the Markov chain whose state space consists of all possible ways of assigning three balls to three boxes such that each box contains one ball, i.e., all permutations of three objects. For definiteness, number the states of the Markov chain as follows:*

$$1 : ABC, \quad 2 : BAC, \quad 3 : ACB, \quad 4 : CAB, \quad 5 : BCA, \quad 6 : CBA$$

A Markov chain is described by the following rule:

- *A pair of boxes* $(\mathbf{23})$, $(\mathbf{13})$ *or* $(\mathbf{12})$ *is chosen with probabilities* p_1, p_2 *and* p_3 *respectively* $(p_1 + p_2 + p_3 = 1)$ *and the balls in the two boxes are interchanged.*

1. *Exhibit the* 6×6 *transition matrix* P *of this Markov chain.*
2. *Determine the recurrence, periodicity and transience of the states.*
3. *Show that for* $p_j > 0$ *this Markov chain has a unique stationary distribution. Is the long term behavior of this Markov chain reflected accurately in its stationary distribution? Explain.*
4. *Find a permutation matrix*[4] S *such that*

$$SPS^{-1} = \begin{pmatrix} 0 & Q_1 \\ Q_2 & 0 \end{pmatrix},$$

where Q_j*'s are* 3×3 *matrices.*

Exercise 0.1.4.2 *Consider the Markov chain with state space as in exercise 0.1.4.1, but modify the rule* $\bullet$ *as follows:*

- *Assume* $p_j > 0$ *and* $p_1 + p_2 + p_3 < 1$. *Let* $q = 1 - (p_1 + p_2 + p_3) > 0$. *Interchange the balls in boxes according to probabilities* p_j *as in problem 1, and with probability* q *make no change in the arrangement of balls.*

1. *Exhibit the* 6×6 *transition matrix* P *of this Markov chain.*
2. *Determine the recurrence, periodicity and transience of the states.*

[4] A matrix with entries 0 or 1 and exactly one 1 in every row and column is called a *permutation matrix*. It is the matrix representation of permuting n letters or permuting the basis vectors.

3. *Does this Markov chain have a unique stationary distribution? Is the long term behavior of the Markov chain accurately reflected by the stationary distribution? Explain.*

Exercise 0.1.4.3 *Consider ten boxes $\mathbf{1}, \cdots, \mathbf{10}$ and ten balls $A, B, \cdots, J$, and the Markov chain whose state space consists of all possible ways of assigning ten balls to ten boxes such that each box contains one ball, i.e., all permutations of ten objects. Let $p_1, \cdots, p_{10}$ be positive real numbers such that $\sum p_j = 1$, and define the transition matrix of the Markov chain by the following rule:*

- *With probability p_j, $j = 1, \cdots, 9$, interchange the balls in boxes $\mathbf{j}$ and $\mathbf{j+1}$, and with probability p_{10} make no change in the arrangement of the balls.*

1. *Show that this Markov chain is recurrent, aperiodic and all states communicate. (Do not attempt to write down the transition matrix P. It is a $10! \times 10!$ matrix.)*
2. *What is the unique stationary distribution of this Markov chain?*
3. *Show that all entries of the matrix P^{45} are positive.*
4. *Exhibit a zero entry of the matrix P^{44}?*

Exercise 0.1.4.4 *Consider three state Markov chains $X_1, X_2, \cdots$ and $Y_1, Y_2, \cdots$ with the same transition matrix $P = (P_{ij})$. What is the transition matrix of the coupled chain $(X_1, Y_1), (X_2, Y_2), \cdots$? What is the underlying probability space?*

Exercise 0.1.4.5 *Consider the cube with vertices at (a_1, a_2, a_3) where a_j's assume values 0 and 1 independently. Let $A = (0, 0, 0)$ and $H = (1, 1, 1)$. Consider the random walk, initially at A, which moves with probabilities p_1, p_2, p_3 parallel to the coordinate axes.*

1. *Exhibit the transition matrix P of the Markov chain.*
2. *For $\pi = (\pi_1, \cdots, \pi_8)$, does*

$$\pi P = \pi, \quad \pi_j > 0, \quad \sum \pi_j = 1$$

   have a unique solution?

3. *Let Y be the random variable denoting the number of times the Markov chain hits H before its first return to A. Show that $\mathsf{E}[Y] = 1$.*

Exercise 0.1.4.6 *Find a stationary distribution for the infinite state Markov chain described of exercise 0.1.2.7. (You may want to re-number the states in a more convenient way.)*

Exercise 0.1.4.7 *Consider an 8×8 chess board and a knight which from any position can move to all other legitimate positions (according to the rules of chess) with equal probabilities. Make a Markov chain out of the positions of the knight (see exercise 0.1.2.1) and let P denote its matrix of transition probabilities. Classify the states of the Markov chain determined by P^2. From a given position compute the average time required for first return to that position.* (You may make intelligent use of the computer to solve this problem, but do not try to simulate the moves of a knight and calculate the expected return time by averaging from the simulated data.)

Exercise 0.1.4.8 *Consider two boxes* **1** *and* **2** *containing a total N balls. After the passage of each unit of time one ball is chosen randomly and moved to the other box. Consider the Markov chain with state space $\{0, 1, 2, \cdots, N\}$ representing the number of balls in box* **1**.

1. *What is the transition matrix of the Markov chain?*
2. *Determine periodicity, transience, recurrence of the Markov chain.*

Exercise 0.1.4.9 *Consider two boxes* **1** *and* **2** *each containing N balls. Of the $2N$ balls half are black and the other half white. After passage of one unit of time one ball is chosen randomly from each and interchanged. Consider the Markov chain with state space $\{0, 1, 2, \cdots, N\}$ representing the number of white balls in box* **1**.

1. *What is the transition matrix of the Markov chain?*
2. *Determine periodicity, transience, recurrence of the Markov chain.*
3. *What is the stationary distribution for this Markov chain?*

Exercise 0.1.4.10 *Consider the Markov chain with state space the set of integers $\mathbb{Z}$ and (doubly infinite) transition matrix given by*

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

where p_i, q_i are positive real numbers satisfying $p_i + q_i = 1$ for all i. Show that if this Markov chain has a stationary distribution $\pi = (\cdots, \pi_j, \cdots)$, then

$$\pi_j = p_{j-1}\pi_{j-1} + q_{j+1}\pi_{j+1}.$$

Now assume $q_\circ = 0$ and the Markov chain is at origin at time 0 so that the evolution of the system takes place entirely on the non-negative integers. Deduce that if the sum

$$\sum_{n=1}^{\infty} \frac{p_1 p_2 \cdots p_{n-1}}{q_1 q_2 \cdots q_{n-1} q_n}$$

converges then the Markov chain has a stationary distribution.

Exercise 0.1.4.11 *Let $\alpha > 0$ and consider the random walk X_n on the non-negative integers with a reflecting barrier at 0 (that is, $P_{\circ 1} = 1$) defined by*

$$p_{i\ i+1} = \frac{\alpha}{1+\alpha}, \quad p_{i\ i-1} = \frac{1}{1+\alpha}, \quad \text{for } i \geq 1.$$

1. *Find the stationary distribution of this Markov chain for $\alpha < 1$.*
2. *Does it have a stationary distribution for $\alpha \geq 1$?*

Exercise 0.1.4.12 *Consider a region D of space containing N paricles. After the passage of each unit of time, each particle has probability q of leaving region D, and assume that k new particles enter the region D following a Poisson distribution with parameter λ. The exit/entrance of all the particles are assumed to be indpendent. Consider the Markov chain with state space $\mathbb{Z}_+ = \{0, 1, 2, \cdots\}$ representing the number of particles in the region. Compute the transition matrix P for the Markov chain and show that*

$$P_{jk}^{(l)} \longrightarrow e^{-\frac{\lambda}{q}} \frac{\lambda^k}{q^k k!},$$

as $l \to \infty$.

Exercise 0.1.4.13 *Let $f_1, f_2, \cdots$ be a sequence of positive real numbers such that $\sum f_j = 1$. Let $F_n = \sum_{i=1}^n f_i$ and consider the Markov chain with state space $\mathbb{Z}_+$ defined by the transition matrix $P = (P_{ij})$ with*

$$P_{i\circ} = \frac{f_{i+1}}{1 - F_i}, \quad P_{i\ i+1} = 1 - p_{i\circ} = \frac{1 - F_{i+1}}{1 - F_i}$$

for $i \geq 0$. Let q_l denote the probability that the Markov chain is in state 0 at time l and $T_\circ$ be the first return time to 0. Show that

1. *$P[T_\circ = l] = f_l$.*

2. *For* $l \geq 1$, $q_l = \sum_k f_k q_{l-k}$. *Is this the re-statement of a familiar relation?*

3. *Show that if* $\sum_j (1 - F_j) < \infty$, *then the equation* $\pi P = \pi$ *can be solved to obtain a stationary distribution for the Markov chain.*

4. *Show that the condition* $\sum_j (1-F_j) < \infty$ *is equivalent to the finiteness of the expectation of first return time to 0.*

Exercise 0.1.4.14 *Let* P *be the* 6×6 *matrix of the Markov chain chain in exercise 0.1.4.2. Let* $p_1 = p_2 = p_3 = \frac{2}{7}$ *and* $q = \frac{1}{7}$. *Using a computer (or otherwise) calculate the matrices* P^l *for* $l = 2, 5$ *and* 10 *and compare the result with the conclusion of theorem 0.1.4.2.*

Exercise 0.1.4.15 *Assume we are in the situation of exercise 0.1.4.3 except that we have 4 boxes instead of 10. Thus with probability* p_j, $j = 1, 2, 3$ *the balls in boxes* $\mathbf{j}$ *and* $\mathbf{j+1}$ *are interchanged, and with probability* p_4 *no change is made. Set*

$$p_1 = \frac{1}{5}, \quad p_2 = \frac{1}{4}, \quad p_3 = \frac{1}{5}, \quad p_4 = \frac{13}{60}.$$

Exhibit the 24×24 *matrix of the Markov chain. Using a computer, calculate the matrices* P^l *for* $l = 3, 6, 10$ *and 20 and compare the result with the conclusion of theorem 0.1.4.2.*

0.2 Examples of Markov Chains

0.2.1 Patterns in Coin Tossing

In this subsection we consider the problem of the first appearance of certain patterns in coin tossing. To put the problem in a more concrete framework suppose two individuals are playing a coin tossing game where a random number generator, starting at time $l = 1$, produces a 0 or 1 (or equivalently an H or T), with equal probability, each unit of time. Player A chooses a pattern of some length n. Then player B chooses another pattern of the same length. The random number generator is turned on and it produces a stream of 0's and 1's which we denote by $X_1, X_2, \cdots, X_l, \cdots$, where X_l denotes the integer generated at time l. Our computer has a window which at time l displays the portion $X_{l-n+1}, X_{l-n+2}, \cdots, X_l$ of the sequence. the player whose pattern appears first wins the game. The interesting conclusion is that regardless of what pattern player A chooses, B can choose a pattern that beats it (we assume $n \geq 3$). Since there are only finitely many patterns of a given length and each pattern of length n has the same probability of appearance in n tosses of a fair coin, this may appear very surprising. Two points will clarify this matter.

The fact that all patterns of a given length have the same probability of appearance is not relevant. For example assume player A chooses 000, then player B by choosing the pattern 100 will beat player A although both patterns have the same probability of appearance. In fact if of the first three numbers generated are not 000 then necessarily player B wins because in that case the first time 000 appears it is preceded by 1. Therefore probability of A winning is only $\frac{1}{8}$. The point is that because we have a moving window which displays the sequence $X_{l-n+1}, X_{l-n+2}, \cdots, X_l$, the appearance of various patterns are no longer independent events.

Note that if A beats B and B beats C we cannot conclude that A beats C. To demonstrate this point let us consider a different context. Assume we have three students A, B and C who take 100 tests. Assume

1. A scored 70 on every test.

2. B scored 50 on tests 1 through 60, and scored 90 on tests 61 through 100.

3. C scored 45 on tests 1 through 40 and scored 80 on tests 41 through 100.

It follows that 60% of the times A scored higher than B; B scored higher than C 80% of the times; but 60% of the times C scored higher than A. One can summarize this observation by the statement that *stochastic order relations, unlike standard ones, are* **not** *transitive.*

Example 0.2.1.1 Before considering the general case let us analyze the case of patterns of length $n = 3$ in greater detail. We have already shown that the pattern $\beta = 100$ beats

$\alpha = 000$ seven out of eight times. Let $\gamma = 110$ and A denote the event that β beats γ given that $X_1 = 1$. Since $P[A \mid X_2 = 1] = 0$ we obtain

$$P[A] = \frac{1}{4}P[A \mid X_2 = 0, X_3 = 0] + \frac{1}{4}P[A \mid X_2 = 0, X_3 = 1].$$

It is clear that $P[A \mid X_2 = 1] = 0$, $P[A \mid X_2 = 0, X_3 = 0] = 1$ and $P[A \mid X_2 = 0, X_3 = 1] = P[A]$. Therefore $P[A] = \frac{1}{3}$. Let $\beta\top\gamma$ denote the event that β beats γ. Then

$$P[\beta\top\gamma] = \frac{1}{2}P[A] + \frac{1}{2}P[\beta\top\gamma|X_1 = 0].$$

Since $P[\beta\top\gamma|X_1 = 0] = P[\beta\top\gamma]$ and $P[A] = \frac{1}{3}$ we obtain

$$P[\beta\top\gamma] = \frac{1}{3},$$

or two out of three times pattern $\gamma = 110$ beats $\beta = 100$. How can we beat γ? Pattern $\beta' = 011$ beats γ three out of four times since unless $X_1 = 1, X_2 = 1$, in which case $\gamma\top\beta'$, 011 beats 110. By symmetry pattern $\gamma' = 001$ beats $\beta' = 011$ two out of three times. Consider the pattern $\epsilon = 101$. We proceed as before by conditioning on X_1:

$$\begin{aligned} P[\epsilon\top\gamma] &= \tfrac{1}{2}P[\epsilon\top\gamma \mid X_1 = 0] + \tfrac{1}{2}P[\epsilon\top\gamma \mid X_1 = 1] \\ &= \tfrac{1}{2}P[\epsilon\top\gamma] + \tfrac{1}{2}P[\epsilon\top\gamma \mid X_1 = 1]. \end{aligned}$$

Let C be the event $[\epsilon\top\gamma \mid X_1 = 1]$. Then $P[C|X_2 = 1] = 0$ and consequently by conditioning on X_3 we obtain

$$P[C] = \frac{1}{2}P[C|X_2 = 0] = \frac{1}{4}P[\epsilon\top\gamma] + \frac{1}{4}.$$

Substituting, we obtain after a simple calculation

$$P[\epsilon\top\gamma] = \frac{1}{3}.$$

We summarize our calculations as follows:

1. $\beta = 100$ beats $\alpha = 000$ with probability $\frac{7}{8}$.

2. $\gamma = 110$ beats $\beta = 100$ with probability $\frac{2}{3}$.

3. $\beta' = 011$ beats $\gamma = 110$ with probability $\frac{3}{4}$.

4. $\gamma = 110$ beats $\epsilon = 101$ with probability $\frac{2}{3}$.

In view of the symmetry between 0 and 1 this completes our claim that for every pattern of length 3 there is pattern of the same length that beats it. ♠

While for short patterns one can completely analyze the situation by conditioning, it is necessary to invoke some more sophisticated techniques to obtain formulas for relevant probabilities. We invoke our knowledge of finite Markov chains for this purpose. Consider the Markov chain whose state space S consists of $\emptyset$ and all patterns of length $\leq n$. Thus S has cardinality $2^{n+1}-1$. Initially the Markov chain is in state $\emptyset$, at time 1 it moves to one of the states 0 or 1; at time 2 to one of the four states $00, 01, 10, 11$ etc. until time n when it enters one of the 2^n patterns of length n. For $l > n$ it remains in one of the states corresponding to patterns of length n. Thus the states corresponding to $\emptyset$ and patterns of length $< n$ are transient while those of length n are recurrent.

Let $\alpha_1, \cdots, \alpha_k$ be patterns of length n of 0's and 1's, and set $R = \{\alpha_1, \cdots, \alpha_k\} \subset S$. We consider the problem of calculating the probability that the first visit to R is by hitting state α_1. Let $a \in S$ but $a \notin R$, T_{aR} denote the first hitting time of R starting at a, and μ_{aR} be its expectation. For $\alpha \in R$ let R_α denote the event that the Markov chain enters R for the first time through $\alpha \in R$, i.e., the set of paths whose first visit to R is at α. Set $\psi_{a\alpha} = P[R_\alpha]$ where we are assuming that paths start at a. We have

$$\begin{aligned} \mu_{a\beta} &= \mathsf{E}[T_{a\beta}] \\ &= \mathsf{E}[T_{a\beta} - T_{aR}] + \mathsf{E}[T_{aR}] \\ &= \sum_{\alpha \in R} \mathsf{E}[T_{a\beta} - T_{aR} \mid R_\alpha]\psi_{a\alpha} + \mu_{aR}. \end{aligned}$$

Looking at the paths, it is immediate that

$$\mathsf{E}[T_{a\beta} - T_{a\alpha} \mid R_\alpha] = \mu_{\alpha\beta}.$$

Therefore

$$\mu_{a\beta} = \mu_{aR} + \sum_{\alpha \in R} \mu_{\alpha\beta}\psi_{a\alpha} \tag{0.2.1.1}$$

The quantities $\psi_{a\alpha}$ satisfy the linear relation

$$\sum_{\alpha \in R} \psi_{a\alpha} = 1. \tag{0.2.1.2}$$

Here we are using the assumption that the Markov chain enters the set R with probability 1 (recurrence, see below for further discussion). If R has cardinality r, then the equations (0.2.1.1) and (0.2.1.2) form a linear system of $r+1$ equations in $r+1$ unknowns μ_{iR}, ψ_{ia} which can be solved to obtain the desired quantities. Note that in these equations $\mu_{\alpha\alpha} = 0$.

Let us go back to our Markov chain. After time $l = n$ a moving window exhibits the latest pattern of length n which is a recurrent state, and states of length $< n$ are transient.

To demonstrate the applicability of the above analysis, we consider the special case where R consists of two patterns of length n which we denote by **1** and **2**. Let $a = \emptyset$. Then solving equations (0.2.1.1) and (0.2.1.2) yields

$$\psi_{\emptyset 1} = \frac{\mu_{\emptyset 1} - \mu_{\emptyset 2} + \mu_{21}}{\mu_{12} + \mu_{21}}, \quad \psi_{\emptyset 2} = \frac{\mu_{\emptyset 2} - \mu_{\emptyset 1} + \mu_{12}}{\mu_{12} + \mu_{21}}. \tag{0.2.1.3}$$

The equations (0.2.1.3) enable one in general to compare the relative merit of two patterns of length $n = 3$. Note that the numerator depends on the difference $\mu_{\emptyset 1} - \mu_{\emptyset 2}$ which shows that the time $n - 1$ required to get to patterns of length n is in fact irrelevant. The fact that the solutions do not involve μ_{aR} is not limited to the case $n = 3$ since by taking the differences of the equations in (0.2.1.1) and (0.2.1.2) we obtain a set of r equations in r unknowns which does not involve μ_{aR}. Of course one should address the issue of solvability of this system, however we will ignore this matter.

Although the term μ_{aR} does not appear in the calculation of $\psi_{\emptyset\alpha}$, nevertheless it is reasonable to compute it. Let $R = \{\alpha_1, \cdots, \alpha_k\}$ consist of k patterns of length n, and $\mathsf{F}_R(\xi)$ denote the generating function for the first hitting time of the set R. Let $P^{(l)}_{\emptyset R}$ denote the probability the the window shows one of the patterns in R at time l, and $\mathsf{P}_R(\xi) = \sum P^{(l)}_{\emptyset R}\xi^l$ be its probability generating function. Then $\mathsf{P}_{\emptyset R}(\xi)$ can be easily expressed in terms of $\mathsf{P}_{\emptyset\alpha_j}(\xi)$'s. It is clear that

$$\mathsf{P}_{\emptyset R}(\xi) = \sum_{j=1}^{k} \mathsf{P}_{\emptyset\alpha_j}(\xi). \tag{0.2.1.4}$$

This follows from the fact the events $X_l = \alpha_i$ and $X_l = \alpha_j$ for $i \neq j$ are mutually exclusive events. Since we already know, in principle, how to calculate $\mathsf{P}_{\emptyset\alpha_j}(\xi)$, the calculation of $\mathsf{P}_{\emptyset R}(\xi)$ is straightforward. We also have the relation

$$\mathsf{P}_{\emptyset R}(\xi) = \mathsf{F}_R(\xi)\mathsf{P}_{RR}(\xi), \tag{0.2.1.5}$$

where $\mathsf{P}_{RR}(\xi)$ is the probability generating function for being in R given that initially the MArkov chain is in R. Equation 0.2.1.5 is similar to the second equation in theorem 0.1.3.1. In principle we also know how to calculate $\mathsf{P}_{RR}(\xi)$, and therefroe we can calculate $\mathsf{F}_R(\xi)$ and calculate the required expectation. The generating function calculation is perhaps laborious and it is not necessary for computing the quantities $\psi_{\emptyset a}$.

We already know how to calculate the expectation $\mu_{\emptyset\alpha}$ of the first hitting time to state $\alpha \in S$ for example by making use of the generating function for time of first appearance of the pattern α. It is not difficult to calculate the expectation $\mu_{\alpha\beta}$ of the hitting time of β given that we start in $\alpha \neq \beta$. To do so we make use of generating function $\mathsf{F}_{\alpha\beta}$. We have

$$\mu_{\alpha\beta} = \lim_{\xi \to 1^-} \frac{1 - \mathsf{F}_{\alpha\beta}(\xi)}{1 - \xi}. \tag{0.2.1.6}$$

The quotient on the right hand side is the derivative except that since the derivative may not exist at $\xi = 1$, we prefer to use this expression. To evaluate the limit we multiply numerator and denominator by $\mathsf{P}_{\beta\beta}(\xi)$ and use the fundamental relations of theorem 0.1.3.1 to obtain

$$\mu_{\alpha\beta} = \lim_{\xi\to 1^-} \frac{\mathsf{P}_{\beta\beta}(\xi) - \mathsf{P}_{\alpha\beta}(\xi)}{(1-\xi)\mathsf{P}_{\beta\beta}(\xi)}. \tag{0.2.1.7}$$

Notice that for a recurrent state β we have $\lim_{\xi\to 1^-}\mathsf{P}_{\beta\beta}(\xi) = \infty$, so it is necessary to determine the limit in the denominator.

Lemma 0.2.1.1 *With the above notation we have*

$$\lim_{\xi\to 1^-}(1-\xi)\mathsf{P}_{\beta\beta}(\xi) = \frac{1}{\mu_\beta},$$

where μ_β denotes the expectation of the first return time to β.

Proof - Let N be a large integer and write

$$\mathsf{P}_{\beta\beta}(\xi) = \sum_{l=0}^{N} P_{\beta\beta}^{(l)}\xi^l + \sum_{l=N+1}^{\infty} P_{\beta\beta}^{(l)}\xi^l.$$

Clearly for every fixed $N < \infty$,

$$\lim_{\xi\to 1^-}(1-\xi)\sum_{l=0}^{N} P_{\beta\beta}^{(l)}\xi^l = 0.$$

On the other hand, from theorem 0.1.4.2 we know that for l sufficiently large, $P_{\beta\beta}^{(l)}$ is approximately $\frac{1}{\mu_\beta}$ which is the β^{th} component of the vector representing the stationary distribution. Therefore

$$\lim_{\xi\to 1^-}(1-\xi)\sum_{l=N+1}^{\infty} P_{\beta\beta}^{(l)}\xi^l = \lim_{\xi\to 1^-}[(1-\xi)\frac{\xi^{N+1}}{1-\xi}\frac{1}{\mu_\beta}] = \frac{1}{\mu_\beta},$$

completing the proof of the lemma. ♣

Note that lemma 0.2.1.1 did not make any use of the particular Markov chain under discussion and is of a general nature. In evaluating the numerator of 0.2.1.7 we make use of an observation about our particular Markov chain. Since we are interested in patterns of length n and our random number generator produces a 0 or a 1 at each unit of time independently, we have

$$P^{(l)}_{\alpha\beta} = P^{(l)}_{\beta\beta} \quad \text{for} \quad l \geq n.$$

It follows that the numerator of 0.2.1.7 is the finite sum

$$1 + \sum_{l=1}^{n-1} [P^{(l)}_{\beta\beta} - P^{(l)}_{\alpha\beta}]. \tag{0.2.1.8}$$

Therefore

$$\mu_{\alpha\beta} = \big(1 + \sum_{l=1}^{n-1} [P^{(l)}_{\beta\beta} - P^{(l)}_{\alpha\beta}]\big)\mu_\beta. \tag{0.2.1.9}$$

It should be pointed out the expectations $\mu_{\emptyset\alpha}$ depend on the particular recurrent state α. This point is discussed in exercise 0.2.1.3 below.

In the Markov chain of the above discussion, patterns of length n are the recurrent states and those of length $< n$ are transient. A related issue is whether as the window moves and our computer generates 0's and 1's with probability $\frac{1}{2}$ ad infinitum, every pattern (of fixed length n) is generated infinitely often. We can answer this question more or less in the same way that we showed that with probability 1 a transient state is visited only finitely many times but with one important modification. If we are looking at patterns of length n then we consider, for some starting point $k = 1, \cdots, n-1$, the set

$$\Omega_\circ = \bigcap_{l=1}^{\infty} \bigcup_{m=l}^{\infty} \Omega^{(k+mn)},$$

where $\Omega^{(k+mn)}$ is the set of paths which have the desired pattern at time $k+mn, k+mn+1, \cdots, k+mn+n-1$. Notice that here we are only looking at the sets $\Omega^{(k+mn)}$ with k and n fixed. It is clear the events $\Omega^{(k+n)}$, $\Omega^{(k+2n)}$, $\Omega^{(k+3n)}$, $\cdots$ are independent. To prove that every pattern of length appears infinitely often with probability 1, it suffices to show that for some k the complement of $\Omega_\circ$ has probability 0. The complement of $\Omega_\circ$ is

$$\Omega'_\circ = \bigcup_{l=1}^{\infty} \bigcap_{m=l}^{\infty} A_m,$$

where A_m denote the complement of $\Omega^{(k+mn)}$. Thus it suffices to show that for all l

$$P[\bigcap_{m=l}^{\infty} A_m] = 0. \tag{0.2.1.10}$$

The independence of $\Omega^{(k+mn)}$'s as m varies implies that of $A_1, A_2, A_3, \cdots$ and therefore we have

$$P[\bigcap_{m=i}^{i+j} A_m] = \prod_{m=i}^{i+j} \big(1 - P[\Omega^{(k+mn)}]\big).$$

Since $1 - x \leq e^{-x}$ we obtain

$$P[\bigcap_{m=i}^{i+j} A_m] \leq \prod_{m=i}^{i+j} e^{-P[\Omega^{(k+mn)}]}. \tag{0.2.1.11}$$

Now we make use of the fact each pattern is a recurrent state to conclude that, for some k, $\sum_m \Omega^{(k+mn)} = \infty$. Substituting in (0.2.1.11) and letting $j \to \infty$ we conclude the validity of (0.2.1.10). Therefore every pattern is visited infinitely often with probability 1. The general issue of infinite visits to a recurrent state of a Markov chain will be discussed in a later subsection.

The algebraic method developed earlier for the calculation of the generating function of the time of the appearance of a pattern can be extended to this case of comparing the relative merits of two patterns. Let A and B be two pattern of the same length n let X_A be the formal sum of all sequences of H's T's terminating in pattern A and such that pattern B does not appear anywhere. In other words, it is the formal sum of all possible sequences with pattern A beating pattern B. Let Y be the formal sum of all finite sequences of H's T's not containing either pattern A or pattern B. Just as in the case of a single pattern it is a simple matter to verify the validity of the following identity:

$$1 + Y(H + T) = Y + X_A + X_B. \tag{0.2.1.12}$$

Introduce the quantities Δ_j^{AB} by

$$\Delta_j^{BA} = \begin{cases} 1 & \text{if } \ B^j = A_j; \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\Delta_j^{BA} \neq \Delta_j^{AB}$. It is straightforward to verify the validity of the following equations which are patterned after a similar identity for the case of a single pattern:

$$YA = X_A \sum_{j=\circ}^{n-1} \Delta_{n-j}^{A} A^j + X_B \sum_{j=\circ}^{n-1} \Delta_{n-j}^{BA} A^j, \tag{0.2.1.13}$$

and

$$YB = X_B \sum_{j=\circ}^{n-1} \Delta_{n-j}^{AB} A^j + X_B \sum_{j=\circ}^{n-1} \Delta_{n-j}^{B} B^j. \qquad (0.2.1.14)$$

Let $X_A(\xi)$ denote the series obtained from substituting $p\xi$ for H and $q\xi$ for T in X_A. If we furthermore set $\xi = 1$ in $X_A(\xi)$ we obtain the probability that when pattern A appears for the first time pattern B has not yet appeared. Now assume that the coin is fair so that $p = \frac{1}{2} = q$. Then calculations simplify considerably and we obtain the neat formula (0.2.1.15) given below describing the relative merits of patterns A and B. Note that to quantify the relative merits of the patterns A and B we calculate $\frac{X_A(\frac{1}{2})}{X_B(\frac{1}{2})}$. This quotient being > 1 (resp. < 1) means pattern A (resp. pattern B) is superior. It is straightforward algebra to show that the substitution $H = \frac{1}{2} = T$ in (eq:BoCan32) and (eq:BoCan32A) yields

$$X_A(\frac{1}{2})\wp(A,A) + X_B(\frac{1}{2})\wp(B,A) = X_A(\frac{1}{2})\wp(A,B) + X_B(\frac{1}{2})\wp(B,B),$$

where

$$\wp(A,B) = \sum_{j=\circ}^{n-1} 2^i \Delta_j^{AB}.$$

Consequently

$$\frac{X_A(\frac{1}{2})}{X_B(\frac{1}{2})} = \frac{\wp(B,B) - \wp(B,A)}{\wp(A,A) - \wp(A,B)}. \qquad (0.2.1.15)$$

From this formula it is a simple matter to examine the relative merits of any two given patterns. A general formula for beating a pattern is given in exercise 0.2.1.5.

EXERCISES

Exercise 0.2.1.1 *Verify the conclusion of example 0.2.1.1 on the basis of (0.2.1.3).*

Exercise 0.2.1.2 *Assume the initial state is a pattern α of length n. Let $T^{(r)}$ denote the time of the r^{th} appearance of α. What is the normal approximation to $T^{(r)}$ as $r \to \infty$? How should this modified if we start with initial state $\emptyset$?*

Exercise 0.2.1.3 *Use the generating functions for patterns 000 and 100 to show that $\mu_{\emptyset\alpha}$ depends on α.*

Exercise 0.2.1.4 *Analyze patterns of length four (as in example 0.2.1.1.*

Exercise 0.2.1.5 *Let $A = A_1A_2\ldots A_n$ be a pattern of length $n \geq 3$ with each $A_j = H$ or T. Let A'_j be the reverse of A_j. Show that the pattern $A'_2A_1A_2\ldots A_{n-1}$ beats pattern A.*

0.2.2 A Branching Process

It is natural to see if we can derive a formula for the generating function of sums of iid random variables. In fact, more generally, we let N be a random variable with values in $\mathbf{N}$ and X a random variable variable with values in $\mathbb{Z}_+$. Let $X_1, X_2, \cdots$ be a sequence of iid random variables with with the same distribution as X. The random variable N may not be independent from $X_1, X_2, \cdots$. Set $Y = X_1 + X_2 + \cdots + X_N$. We have

Lemma 0.2.2.1 *With the above notation and hypothesis, we have*

$$\mathsf{F}_Y(\xi) = \mathsf{F}_N(\mathsf{F}_X(\xi)).$$

In particular, if N is a fixed integer, then $\mathsf{F}_Y(\xi) = \mathsf{F}_X(\xi)^N$.

Proof - We have

$$\begin{array}{rcl} \mathsf{E}[\xi^Y] & = & \mathsf{E}[\mathsf{E}[\xi^Y \mid N]] \\ \text{(by independence)} & = & \mathsf{E}[\mathsf{E}[\xi^{X_1}] \cdots \mathsf{E}[\xi^{X_N}]] \\ & = & \mathsf{E}[\mathsf{E}[\xi^X]^N] \\ & = & \mathsf{F}_N(\mathsf{E}[\xi^X]) \\ & = & \mathsf{F}_N(\mathsf{F}_X(\xi)), \end{array}$$

proving the lemma. ♣

As another example of a Markov chain and application of generating functions we consider a simple *Branching Process*. Assume at time $l = 0$ there is one particle. Let X be a random variable with values in $\mathbb{Z}_+$ and $f(k) = P[X = k]$ be its density function. At time $l = 1$ there are X particles. At time $l = 2$ each particle that has appeared at time $l = 1$ is replaced by X particles independently of each other. Thus if Z_l is the number of particles at time l, (so that $Z_\circ = 1$), then

$$Z_{l+1} = \sum_{j=1}^{Z_l} X_j \tag{0.2.2.1}$$

where X_j are iid with the same distribution as X. We are interested in evaluating $\lim_{l\to\infty} P[Z_l = 0]$, i.e., the probability that the branching process (or chain) dies. The answer to this question is quite simple and elegant in terms of the generating function $\mathsf{F}_X(\xi)$. In fact, we show that if ζ is smallest non-negative root of the equation $\mathsf{F}_X(\xi) = \xi$, then

$$\lim_{l\to\infty} P[Z_l = 0] = \zeta. \tag{0.2.2.2}$$

To eliminate some trivial cases we assume

$$0 < f(0) < 1.$$

Let $\zeta_n = P[Z_n = 0]$. In view of the the inclusion of events $\{Z_n = 0\} \subset \{Z_{n+1} = 0\}$, we have

$$\zeta_n \leq \zeta_{n+1} \leq 1,$$

and therefore $\lim_{n\to\infty} \zeta_n$ exists. We denote this limit by ζ. Set $\mathsf{G}_n(\xi) = \mathsf{F}_{Z_n}(\xi)$. Then lemma 0.2.2.1 implies

$$\mathsf{G}_{n+1}(\xi) = \mathsf{G}_n(\mathsf{F}_X(\xi)) = \mathsf{G}_{n-1}(\mathsf{F}_X(\mathsf{F}_X(\xi))) = \cdots = \mathsf{F}_X(\mathsf{G}_n(\xi)).$$

Clearly $P[Z_n = 0]$ is the constant term of $\mathsf{F}_n(\xi)$ and so

$$\zeta_{n+1} = \mathsf{F}_{n+1}(0) = \mathsf{F}_X(\mathsf{F}_n(0)) = \mathsf{F}_X(\zeta_n).$$

Let $n \to \infty$ to obtain

$$\zeta = \mathsf{F}_X(\zeta).$$

Now $\mathsf{F}_X(\xi)$ in an increasing function of ξ since $\mathsf{F}_X(\xi) = \mathsf{E}[\xi^X]$. Let $\eta > 0$, $\mathsf{F}_X(\eta) = \eta$ and η be the smallest positive root of the equation $\mathsf{F}_X(\xi) = \xi$. If $\zeta_n < \eta$, then (draw a picture and use increasing property of F_X)

$$\zeta_{n+1} = \mathsf{F}_X(\zeta_n) \leq \mathsf{F}_X(\eta) = \eta.$$

Since $0 < f(0) < 1$ we have $\eta > \mathsf{F}_X(0) = \zeta_\circ$, and it follows that ζ is the smallest positive root of $\mathsf{F}_X(\xi) = \xi$.

Except when the density function f_X of X has special forms, it is not possible to give closed form expressions for the generating functions $\mathsf{F}_X(\xi)$ and $\mathsf{F}_l(\xi)$. An example where such a calculation is possible is given in exercise 0.2.2.3.

It is intuitively clear that as as $l \to \infty$ one should expect Z_l either go to ∞ or 0. This follows easily from

Lemma 0.2.2.2 *Assume $f(1) \neq 1$. The states $1, 2, 3, \cdots$ of the branching process $Z_\circ = 1, Z_1, Z_2, \cdots$ are transient.*

Proof - Let $k \geq 1$ be an integer and $\mathsf{F}_n = \sum_{l\geq 1} P[Z_{n+l} = k \mid Z_n = k]$. If $f(0) = 0$ then it is clear that

$$\mathsf{F}_n = f(1)^n < 1.$$

If $f(0) > 0$, then we have

$$\mathsf{F}_n \leq 1 - f(0)^n < 1.$$

Since F_n is the probability of return to state k given that the branching process is in state k at time n, we conclude that every $k = 1, 2, 3, \cdots$ is a transient state. ♣

Corollary 0.2.2.1 *With the above notation, assuming $f(1) \neq 1$ we have*

$$\lim_{l\to\infty} Z_l = \begin{cases} \infty, & \text{withprobability } 1-\zeta; \\ 0, & withprobability\ \zeta. \end{cases}$$

Proof - Since every state $k = 1, 2, \cdots$ is transient, and with probability 1 a transient is visited only finitely many times, the possible limiting values for Z_l are 0 and ∞ only. The required result follows since ζ is the probability of extinction. ♣

Since the limiting values of Z_l are only 0 and ∞, it is reasonable to try to understand the limiting behavior of a branching process with $\mu = \mathsf{E}[X] > 1$ by looking at the random variables $\frac{Z_l}{\mu^l}$. This will be discussed in the context of martingales. The case $\mu \leq 1$ is simpler (see exercise 0.2.2.1).

EXERCISES

Exercise 0.2.2.1 *With the above notation assume $f(1) \neq 1$, $\mu \leq 1$ and consider the branching process $Z_\circ = 1, Z_1, Z_2, \cdots$ as above. Show that with probability 1*

$$\lim_{l\to\infty} Z_l = 0.$$

Exercise 0.2.2.2 *Let Z_n be a branching process with $Z_\circ = 1$ and $\mathsf{E}(Z_1) = \mu$. Show that for $m \leq n$ we have*

$$\mathsf{E}(Z_m Z_n) = \mu^{n-m}\mathsf{E}(Z_m^2).$$

Exercise 0.2.2.3 *Consider the branching process $Z_\circ = 1, Z_1, Z_2, \cdots$ where $f(k) = P[X = k] = \frac{a}{b^{k-1}}$ for $k = 1, 2, 3, \cdots$, and $f(0) = 1 - \sum f(k)$ where $a, b > 0$ and $ab + 1 < b$. Show that*

$$\mathsf{F}_X(\xi) = 1 - \frac{ab}{b-1} + \frac{ab\xi}{b-\xi}, \quad \mu = \mathsf{E}[X] = \frac{ab^2}{(b-1)^2}.$$

Therefore if $\mathsf{E}[X] \neq 1$ then the non-negative solution ζ of $\xi = \mathsf{F}_X(\xi)$ giving the probability of extinction is

$$\zeta = b\frac{b - ab - 1}{b-1}.$$

Calculate the generating functions $\mathsf{F}_l(\xi)$ for $l = 2, 3$. Draw the curves $\mathsf{F}_l(\xi)$ for $l = 1, 2, 3$ and $a = b^{-1} = \frac{1}{3}$.

Exercise 0.2.2.4 (Continuation of exercise 0.2.2.3) *Deduce for $\mu \neq 1$*

$$\mathsf{F}_l(\xi) = 1 - \mu^l\left(\frac{1-\zeta}{\mu^l - \zeta}\right) + \frac{\mu^l\left(\frac{1-\zeta}{\mu^l-\zeta}\right)^2\xi}{1 - \frac{\mu^l-1}{\mu^l-\zeta}\xi}$$

For $\mu = 1$, F_l reduces to

$$\mathsf{F}_l(\xi) = \frac{lb - (lb + b - 1)\xi}{1 - b + lb - lb\xi}.$$

0.2.3 Fluctuations of the Random Walk on $\mathbb{Z}$

In this subsection we analyze the simple symmetric random walk on $\mathbb{Z}$ by introducing the geometric idea known as the *Reflection Principle*. Througout this section we assume $p = q = \frac{1}{2}$ unless explicitly stated to the contrary. While some of the conclusions we derive from the reflection principle can also be obtained by calculations with generating functions, the application of the former method is conceptually more appealing. As usual we let the horizontal axis denote the (discrete) time variable and the vertical axis the states. Let $\mathcal{N}(l,n)$ denote the set of paths starting at $(0,0)$ and ending at (l,n). Note that since we looking at the paths of a simple random walk, we only allow paths where the point (l',m) is joined only to $(l'+1,m+1)$ or $(l'+1,m-1)$. The subset of $\mathcal{N}(l,n)$ consisting of paths such that for some $l'<l$ the path reaches the point (l',k), where $k>n$, is denoted by $\mathcal{N}(l,n,k)$. Denote the cardinalities of $\mathcal{N}(l,n)$ and $\mathcal{N}(l,n,k)$ by $N(l,n)$ and $N(l,n,k)$ respectively. We have

Lemma 0.2.3.1 *$\mathcal{N}(l,n)$ is empty unless $l \geq |n|$ and l and n have the same parity (both odd or both even). In such a case we have*

$$N(l,n) = \binom{l}{\frac{l+n}{2}}.$$

Proof - The first statement is clear. A path joining $(0,0)$ to (l,n) is of the form

$$(0,0),(1,\epsilon_1),(2,\epsilon_1+\epsilon_2),\cdots,(l,\sum_{j=1}^{l}\epsilon_j), \tag{0.2.3.1}$$

where ϵ_j's are ±1 and $\sum_{j=1}^{l}\epsilon_j = n$. Therefore the paths joining $(0,0)$ to (l,n) are precisely those for which the number of $\epsilon_j = +1$ exceeds those for which $\epsilon_j = -1$ by n. The required result follows. ♣

The quantities $N(l,n)$ and $P_{on}^{(l)}$ are easy to calculate, but it is not immediately clear how to calculate $N(l,n,k)$. The reflection principle allows us to calculate $N(l,n,k)$ by relating it to $N(l,n')$. For a path $\omega \in \mathcal{N}(l,n,k)$ let $l'<l$ be the largest integer such that $\omega(l') = k$. Now define the mapping $\mathcal{R}: \mathcal{N}(l,n,k) \rightarrow \mathcal{N}(l,2k-n)$ by reflecting the portion of the path to the right of (l',k) relative to the horizontal line through k. More precisely, if the path ω is given as

$$\omega: \ (0,0) \rightarrow (1,\omega(1)) \rightarrow (2,\omega(2)) \rightarrow \cdots \rightarrow (l,\omega(l)),$$

then $\mathcal{R}(\omega)$ is the path

$$\mathcal{R}(\omega)(j) = \begin{cases} \omega(j) & \text{if } j \leq l'; \\ 2k-\omega(j) & \text{if } l'+1 \leq j \leq l. \end{cases}$$

The proof of the following basic lemma is straightforward:

Lemma 0.2.3.2 *Let $0 < n \geq k$. Then $\mathcal{R} : \mathcal{N}(l,n,k) \rightarrow N(l, 2k-n)$ is a bijection (i.e., establishes one to one correspondence between the two sets).*

In particular ,for $k = n+1$ we obtain

Corollary 0.2.3.1 *The map $\mathcal{R} : \mathcal{N}(l,n,n+1) \rightarrow \mathcal{N}(l,n+2)$ is a bijection*

A fundamental consequence of corollary 0.2.3.1 is

$$N(l,n,n+1) = N(l,n+2) = \binom{l}{\frac{l+n}{2}+1}. \tag{0.2.3.2}$$

Now we apply the reflection principle or more precisely (0.2.3.2) to compute the density function of the random variable $T_{\circ n}$ (first hitting time of state n for a simple random walk on $\mathbb{Z}$). We want to determine the set of paths, in the range $0 \leq l' \leq l$, for which $T_{\circ n} = l$. For a path ω in this set we have

$$\omega(l-1) = n-1, \quad \text{and} \quad \omega(j) \leq n-1 \quad \text{for all} \quad j \leq l-1.$$

It follows that set of such paths is precisely the complement of $\mathcal{N}'(l-1,n-1)$ in $\mathcal{N}(l-1,n-1)$. Note that we are assuming $\omega(l) = n$. Therefore the number of such paths in the range $0 \leq l' \leq l$ is the cardinality of $\mathcal{N}(l-1,n-1) \backslash \mathcal{N}(l-1,n-1,n)$, i.e.,

$$N(l-1,n-1) - N(l-1,n+1)$$

by (0.2.3.2). Representing such paths in the form (0.2.3.1), we note that for every such path $\frac{l+n}{2} - 1$ of ϵ_j's are $+1$ and $\frac{l-n}{2}$ of ϵ_j's are -1. Therefore

$$F^l_{\circ n} = P[T_{\circ n} = l] = \frac{1}{2}\left[\binom{l-1}{\frac{l+n}{2}-1} - \binom{l-1}{\frac{l+n}{2}}\right]\frac{1}{2^l} = \frac{n}{l}\binom{l}{\frac{l+n}{2}}\frac{1}{2^l}. \tag{0.2.3.3}$$

Here we are assuming l and n have the same parity for otherwise $F^l_{\circ n} = P[T_{\circ n} = l] = 0$. Note that the key step in the calculation of the density function for $T_{\circ n}$ was the application of the reflection principle which has other applications as well (see for example exercise 0.2.3.2). In example 0.1.3.8 of subsection Generating Functions we derived the density function for $T_{\circ n}$. It is instructive to consider relative merits of the two methods.

Remark 0.2.3.1 For the simple random walk where $p \neq q$ the above argument is applicable to give

$$p\left[\binom{l-1}{\frac{l+n}{2}-1} - \binom{l-1}{\frac{l+n}{2}}\right] p^{\frac{l+n}{2}} q^{\frac{l-n}{2}} = \frac{n}{l}\binom{l}{\frac{l+n}{2}} p^{\frac{l+n}{2}} q^{\frac{l-n}{2}}$$

for $F^l_{\circ n} = P[T_{\circ n} = l]$. ♡

Now we apply the reflection principle to the simple symmetric random walk on $\mathbb{Z}$ and derive the arc-sine law. Consider the simple symmetric random walk $S_\circ = 0, S_1, S_2, \cdots$ on $\mathbb{Z}$. Let V_{2l} be the random variable denoting the time prior to $2l+1$ of the last visit to state 0. The computation of the density function of V_{2k} is substantially similar to that of $T_{\circ n}$ but involves an additional observation. Clearly

$$P^{(2m-1)}_{\circ 1} = \binom{2m-1}{m} \frac{1}{2^{2m-1}}. \tag{0.2.3.4}$$

To calculate $F^{2m}_{\circ\circ}$ we consider the set $\mathcal{B}$ of all paths ω from $(0,0)$ to $(2m,0)$ such that $\omega(k) \neq 0$ for all $1 \leq k \leq 2m-1$. This set is divided into two disjoint subsets

$$\mathcal{B} = \mathcal{B}_+ \cup \mathcal{B}_-$$

according as $\omega(k)$ is positive or negative for $1 \leq k \leq 2m-1$. Paths in $\mathcal{B}_\pm$ are necessarily at ± 1 at time $l = 2m-1$. Clearly

$$F^{2m}_{\circ\circ} = \frac{1}{2}P[\mathcal{B}_+] + \frac{1}{2}P[\mathcal{B}_-],$$

and $P[\mathcal{B}_+] = P[\mathcal{B}_-$. By reversing the paths in we see that the set of paths in $\mathcal{B}_-$ are in bijection with the paths from $(0,0)$ to $(2m-1,1)$ such that $\omega(k) \leq 0$ for $1 \leq k \leq 2m-1$. Therefore applying (0.2.3.3) we obtain

$$F^{2m}_{\circ\circ} = \frac{1}{2}P[\mathcal{B}_+] + \frac{1}{2}P[\mathcal{B}_-] = \frac{1}{2m-1}\binom{2m-1}{m}\frac{1}{2^{2m-1}}. \tag{0.2.3.5}$$

Now we can prove the important fact

Lemma 0.2.3.3 *For the simple symmetric random walk on $\mathbb{Z}$ we have*

$$P[T_{\circ\circ} > 2l] = P^{(2l)}_{\circ\circ}.$$

Proof - From the identity

$$\frac{1}{2m-1}\binom{2m-1}{m}\frac{1}{2^{2m-1}}=\binom{2m-2}{m-1}\frac{1}{2^{2m-2}}-\binom{2m}{m}\frac{1}{2^{2m}},$$

and (0.2.3.5) we deduce[5]

$$F_{\circ\circ}^{2m}=P_{\circ\circ}^{(2m-2)}-P_{\circ\circ}^{(2m)}. \tag{0.2.3.6}$$

Since

$$P[T_{\circ\circ}>2l]=1-\sum_{m=1}^{l}F_{\circ\circ}^{2m},$$

we obtain the required result from (0.2.3.6). ♣

Remark 0.2.3.2 It is interesting to compare the above proof of lemma 0.2.3.3 with one based on generating functions which in fact is simpler. Recall that for the simple symmetric random walk we have

$$\mathsf{F}_{\circ\circ}(\xi)=1-\sqrt{1-\xi^2},\quad \mathsf{P}_{\circ\circ}(\xi)=\frac{1}{\sqrt{1-\xi^2}}.$$

Substituting for $\mathsf{F}_{\circ\circ}$ we obtain the identity

$$\frac{1-\mathsf{F}_{\circ\circ}(\xi)}{1-\xi^2}=\mathsf{P}_{\circ\circ}(\xi). \tag{0.2.3.7}$$

We know from (0.1.3.9) that the left hand side of (0.2.3.7) is the generating function for the tail probabilities

$$\text{Coefficient of } \xi^{2m} \text{ is } \ P[T_{\circ\circ}>2m]=F_{\circ\circ}^{2m+2}+F_{\circ\circ}^{2m+4}+\cdots$$

This implies $P_{\circ\circ}^{2l}=P[T_{\circ\circ}>2m]$ which is the conclusion of lemma 0.2.3.3. ♡

A consequence of lemma 0.2.3.3 is

Lemma 0.2.3.4 *The density function of V_{2l} is given by*

$$P[V_{2l}=2k]=P[S_{2k}=0]P[S_{2l-2k}=0].$$

[5]It would be interesting to give a direct probabilistic interpretation of this identity in terms of cardinalities of sets of paths. For the similar statement regarding $F_{\circ n}^{l}$ we had shown, using the reflection principle, the analogous formula $F_{\circ n}^{l}=\frac{1}{2}\left[P_{\circ n-1}^{(l-1)}-P_{\circ n+1}^{(l-1)}\right]$. Therefore this identity is not surprising.

Proof - The event $[V_{2l} = 2k]$ means

$$S_{2k} = 0, \ S_{2k+1} \neq 0, \ \cdots, \ S_{2l} \neq 0,$$

which has the same probability as the event $[T_{\circ\circ} > 2l - 2k]$. Applying lemma 0.2.3.3 we obtain the required result. ♣

It follows from lemma 0.2.3.4 that if l is odd then with probability $\frac{1}{2}$ the simple symmetric random walk on $\mathbb{Z}$ does not visit 0 during the interval $[l, 2l]$. This may appear rather surprising at first.

In view of lemma 0.2.3.4 we have

$$P[V_{2l} = 2k] = \frac{1}{2^{2l}} \binom{2k}{k} \binom{2l-2k}{l-k}. \tag{0.2.3.8}$$

The arc-sine law is an asymptotic approximation to the probability $P[V_{2l} \leq 2k]$ when both l and k become large. To obtain the approximation we let $x = \frac{k}{l}$, then

$$P[V_{2l} \leq 2k] = \frac{1}{2^{2l}} \sum_{j=0}^{[2lx]} \binom{2j}{j} \binom{2l-2j}{l-j}.$$

To put this in more manageable form we make use of the Stirling approximation

$$n! \simeq \sqrt{2\pi} e^{-n} n^{n+\frac{1}{2}}, \tag{0.2.3.9}$$

to obtain

$$P[V_{2l} \leq 2k] \simeq \frac{1}{\pi} \sum_j \frac{1}{l} \frac{1}{\sqrt{\frac{j}{l}(1-\frac{j}{l})}}.$$

Substituting $x = \frac{k}{l}$ and converting the sum into an integral we obtain

$$P[V_{2l} \leq 2k] \simeq \frac{1}{\pi} \int_\circ^x \frac{dt}{\sqrt{t(1-t)}}.$$

The change of variable $\sqrt{t} = u$ immediately yields

Corollary 0.2.3.2 *The distribution function for last visit to the origin is approximated by*

$$P[V_{2l} \leq 2k] \simeq \frac{2}{\pi} \sin^{-1} \sqrt{x}.$$

This is the *arc-sine law* approximating the probability that there are no crossings of zero in the period $2k+1, \cdots 2l$. The density function for V_{2l} as given by lemma 0.2.3.4 will also be refered to as the arc-sine law. A random variable with a similar distribution is that of the time spent above or below the horizontal axis. If fact let $U_{2l}(\omega)$ be the time spent by the path above the horizontal axis. Then we have

Corollary 0.2.3.3 *The density function for U_{2l} is*

$$P[U_{2l} = 2k] = P[S_{2k} = 0]P[S_{2l-2k} = 0].$$

Proof - The proof is by induction on l. The validity of the assertion for small l (e.g., $l = 1, 2$) is easily verified. We assume the assertion is valid for $\leq l-1$. Conditioning on the first return time to the origin and noting that up to the first hitting time a path is entirely above or below the horizontal axis, we obtain

$$P[U_{2l} = 2k] = \frac{1}{2}\sum_{j=1}^{k} F_{\circ\circ}^{2j} P[U_{2l-2j} = 2k-2j]\frac{1}{2}\sum_{j=1}^{l-k} F_{\circ\circ}^{2j} P[U_{2l-2j} = 2k].$$

Applying the induction hypothesis the right hand side becomes

$$P[U_{2l} = 2k] = \frac{1}{2}P_{\circ\circ}^{(2l-2k)}\sum_{j=1}^{k} F_{\circ\circ}^{2j} P_{\circ\circ}^{(2k-2j)} + \frac{1}{2}P_{\circ\circ}^{(2k)}\sum_{j=1}^{l-k} F_{\circ\circ}^{2j} P_{\circ\circ}^{(2l-2k-2j)}.$$

The required result follows immediately. ♣

The reflection principle (or the generating function) allows to understand other properties of the simple symmetric random walk on $\mathbb{Z}$. For example, by applying lemma 0.2.3.2 we prove

Lemma 0.2.3.5 *The probability that the maximum of a path of length l is exactly n is equal to the larger of the numbers $P_{\circ n}^{(l)} + P_{\circ n+1}^{(l)}$.*(One of the two quantities $P_{\circ r}^{(l)}$ and $P_{\circ r+1}^{(l)}$ is zero for parity reasons.)

Proof - From the reflection principle (lemma 0.2.3.2) we know that $\mathcal{N}(l, n, k)$ has the same cardinality as $\mathcal{N}(l, 2n-k)$. It follows that the probability of paths joining $(0, 0)$ to (l, n) with maximum exactly equal to k is

$$P_{\circ 2k-n}^{(l)} - P_{\circ 2k+2-n}^{(l)}.$$

Adding these quantities for all n (all but finitely many are zero) we obtain the desired result. ♣

With another geometric idea we can relate F^l_{on} to the n^{th} return time to the origin. Assume l is even which no loss of generality. Consider the set $\mathcal{B}_{l,n}$ of paths which have their n^{th} return to the origin at time l. Any such path consists of n segments with end points on the time (horizontal) axis. Two paths in $\mathcal{B}_{l,n}$ are considered equivalent ($\sim$) if their visits to the origin occur at the same times and between consecutive visits they can only differ by a sign. In each equivalence class there are 2^n paths which are obtained from each other by changing signs in the segments joining consecutive returns to the origin. From each equivalence class choose the unique representative ω for which $\omega(j) \leq 0$ for all $0 \leq j \leq l$, and denote this set of representatives for equivalence classes by $\mathcal{B}^{\circ}_{l,n}$. We establish a bijection between $\mathcal{B}^{\circ}_{l,n}$ and the set of paths ω for which $T_{\circ n}(\omega) = l - n$. The bijection is best described by looking at an example. Consider, e.g., the path

$\omega(1) = -1,\quad \omega(2) = 0,\quad \omega(3) = -1,\quad \omega(4) = -2,\quad \omega(5) = -1,$
$\omega(6) = 0,\quad \omega(7) = -1,\quad \omega(8) = -2,\quad \omega(9) = -3,\quad \omega(10) = -2,$
$\omega(11) = -1,\quad \omega(12) = 0,$

for which $n = 3, l = 12$. We eliminate all portions of the path joining two consecutive integers k and $k + 1$ for which $\omega(k) = 0$. This will give us a new path of length $l - n$. For example, the above path becomes (draw pictures of the paths)

$\omega(1) = 1,\quad \omega(2) = 0,\quad \omega(3) = 1,\quad \omega(4) = 2,\quad \omega(5) = 1,$
$\omega(6) = 0,\quad \omega(7) = 1,\quad \omega(8) = 2,\quad \omega(9) = 3\quad ,$

which is a path whose first hitting time of state 3 is 9. It is easy to see that this procedure gives the claimed bijection. According to corollary **??** the number of such paths (or the cardinality of $\mathcal{B}^{\circ}_{l,n}$) is

$$\frac{n}{l-n}\binom{l-n}{\frac{l}{2}}.$$

Therefore the probability of the event $\mathcal{B}^{\circ}_{l,n}$ is $\frac{n}{l-n}\binom{l-n}{\frac{l}{2}}\frac{1}{2^l}$ and

$$P[\mathcal{B}_{l,n}] = \frac{n}{l-n}\binom{l-n}{\frac{l}{2}}\frac{1}{2^{l-n}}$$

which is the same as F^{l-n}_{on}. Therefore we have shown

Corollary 0.2.3.4 *The probability that n^{th} return to the origin occurs at time l is equal to $F^{l-n}_{\circ n}$.*

Another interesting aspect of the simple symmetric random walk on $\mathbb{Z}$ is the determination of the distribution of the number of crossings. A path ω has a crsooing at $k \geq 2$ if $\omega(k-1)\omega(k+1) < 0$. Let Z_{2l-1} be the random variable which assigns to a path ω of length

$2l-1$ the number of its crossings. It is convenient to look only at paths of odd length to ensure that the terminal point of the path does not lie on the horizontal axis for otherwise there is an ambiguity as whether the end point should be counted as a crossing.We have

Proposition 0.2.3.1 *The density function for Z_{2l-1} is given by*

$$P[Z_{2l-1}=r]=2P[S_{2l-1}=2r+1],$$

where $r\leq l$.

Proof - We prove the assertion by induction on l. For $l=2$ the assertion is proven easily by inspecting the possibilities. We assume the assertion is true for odd integers $\leq 2l-1$. By conditioning on the first return time to the origin we obtain

$$P[Z_{2l+1}=r]=2\sum_{k=1}^{l}F_{\circ\circ}^{2k}\big(P[Z_{2l-2k+1}=r-1]+P[Z_{2l-2k+1}=r]\big).$$

The induction hypothesis and a comparison of the events $S_{2l-2k+1}=2r-1$, $S_{2l-2k}=2r$ etc. implies

$$P[Z_{2l+1}=r]=2\sum_{k=1}^{l}F_{\circ\circ}^{2k}P[S_{2l-2k+2}=2r]. \tag{0.2.3.10}$$

The sum on right hand side of (0.2.3.10) is the probability of reaching the point $(2l+2,2r)$ after at least one return to the origin. This quantity was essentially computed in exercise 0.2.3.2. We now compute without refernce to the exercise. Since $X_1=\pm1$, we divide the paths from $(0,0)$ to $(2l+2,2r)$ into two subsets $\mathcal{B}_\pm$ according as $X_1=\pm1$. The event $\mathcal{B}_-$ has the same cardinality as the set of paths joining $(0,0)$ to $(2l+1,2r+1)$. We should compute the cardinality of the subset $\mathcal{B}'_+\subset\mathcal{B}_+$ which consists of paths with a return to the origin. To calculate the cardinality of $\mathcal{B}'_+$ we make use of the reflection principle. Let l' be the smallest integer such that $\omega(l')=0$, where $\omega\in\mathcal{B}_+$. Reflecting the portion of the path ω between the points $(1,1)$ and $(l',0)$ with respect to the horizontal axis we see that $\mathcal{B}'_+$ has the same cardinality as the set of paths from $(1,-1)$ to $(2l,2r)$, or equivalently the set of paths joining $(0,0)$ to $(2l-1,2r+1)$. Putting these observations together we see that (0.2.3.10) becomes

$$P[Z_{2l+1}=r]=2P[S_{2l+1}=2r+1]$$

completing the induction. ♣

Remark 0.2.3.3 Proposition 0.2.3.1 shows that for fixed l the probability of $r \leq l$ crossings is a decreasing function of r. While this maybe counter-intuitive, it is compatible with the conclusion of the arc-sine law about the density the density function of V_{2l}. ♡

Another geometric idea which is useful in analyzing the random walk on $\mathbb{Z}$ is a duality transformation which we now describe. Assume we are looking at the random walk from time 0 to time l, so that we have

$$S_\circ = 0,\ S_1 = X_1,\ S_2 = X_1 + X_2,\ \cdots,\ S_l = X_1 + \cdots + X_l$$

where X_k's are iid random variables taking values ± 1 with equal probability $\frac{1}{2}$. To a path $\omega(k)$, $k = 0, \cdots, l$ assign the path

$$\omega^\star(0) = 0, \omega^\star(1) = X_l,\ \omega^\star(2) = X_l + X_{l-1},\ \cdots,\ \omega^\star(l) = X_1 + \cdots + X_l.$$

The transformation $\omega \rightarrow \omega^\star$ is a bijection of the set paths starting at $(0,0)$ up to time l, onto itself. It is called the *dualizing transformation*. The importance of the dualizing transformation lies in the fact that expressing a condition on ω's in terms of $\omega^\star$'s sometimes simplifies the calculation of probabilities. The following example demonstrates this fact:

Example 0.2.3.1 Consider the familiar $T_{\circ n}$ which is the first hitting time of state $n > 0$. The event $T_{\circ n} = l$ is the set of paths ω such that

$$\omega(l) = n,\ \omega(l) > \omega(k) \quad \text{for} \quad k = 1, \cdots, l-1.$$

Applying the dualizing transformation this condition becomes

$$\omega^\star(k) > 0,\ \omega^\star(l) = r.$$

This latter condition was in fact investigated in the proof of proposition 0.2.3.1. In essence we had to count the number of paths from $(0,0)$ to (l,r) which stayed above the horizontal axis for all time $k = 1, \cdots, l$ (see also exercise 0.2.3.2). This number is $\binom{l-1}{\frac{l+n}{2}-1} - \binom{l-1}{\frac{l+n}{2}}$, and thus we obtain another derivation for the density function of $T_{\circ n}$. ♠

Example 0.2.3.2 For given l let $\mathcal{B}$ be the set of paths of length l such that the terminal point $\omega(l)$ is visited prior to time l. Applying the duality operator we see that the subset $\mathcal{B}$ has the same cardinality as the subset $\mathcal{B}^\star$ of paths satisfying

$$\omega^\star(j) = 0, \quad \text{for some} \quad j > 0.$$

It follows that the distribution of last visit to the terminal point is identical with that of last visit to the origin in simple symmetric random walk. The latter follows the (discrete) arc sine law and therefore the density function of last visit follows the same law. This in particular implies the surprising fact that the last visit to the terminal point is more likely to occur near the beginning or the end rather than in the middle. ♠

EXERCISES

Exercise 0.2.3.1 *Consider the simple symmetric random walk on* $\mathbb{Z}^2$ *with* $X_\circ = (0,0)$*. Let* T *be the first hitting time of the line* $x + y = m$*.*

1. *Show that the generating function for* T *is*

$$\mathsf{F}_T(\xi) = [\xi^{-1}(1 - \sqrt{1-\xi^2})]^m.$$

2. *Show that* $\mathsf{E}[T] = \infty$*.*

3. *Let* (X_n, Y_n) *denotes the coordinates of of the random walk at time* n*. Prove that*

$$\mathsf{E}[X_n^2 + Y_n^2] = n.$$

Exercise 0.2.3.2 *Consider the simple symmetric random walk* $S_1, S_2, \cdots$ *on* $\mathbb{Z}$ *where* $X_\circ = 0$*,* $X_j = \pm 1$ *with probabilities* $\frac{1}{2}$*, and* $S_l = X_\circ + X_1 + \cdots + X_l$*. Show that for* $n \leq l$ *we have*

$$P[S_1 \neq 0, S_2 \neq 0, \cdots, S_{2l} \neq 0 \mid S_{2l} = 2n] = \frac{n}{l}.$$

(Hint - Use the reflection principle.)

Exercise 0.2.3.3 *Let* $S_\circ = 0, S_1, S_2, \cdots$ *be the simple symmetric random walk on* $\mathbb{Z}$ *as in exercise 3 above. Let* R_a *denote the number of visits to* $a \in \mathbb{Z}$ *before first return to 0 and* ρ_a *be its expectation. Show that*

$$\rho_a = 1.$$

(Hint - Show that

$$\rho_a = \sum_{l=1}^{\infty} P[S_l = a, S_1 \neq 0, \cdots, S_{l-1} \neq 0].$$

Now using the observation that given $S_l = a$ the conditions

$$X_1 + \cdots + X_j \neq 0 \quad \text{an} \quad X_{j+1} + \cdots + X_l \neq a$$

are equivalent, deduce that $\rho_a = \sum F_{\circ a}^l$.)

Exercise 0.2.3.4 *Let $p < 1-p = q$ and consider the simple random walk $S_\circ = 0, S_1, \cdots$ on $\mathbb{Z}$ where transition from i to $i+1$ (resp. $i-1$) occurs with probability p (resp. q). Let V_a denote the number of visits to state a where for $a = 0$ we include the visit at time $l = 0$.*

1. *Show that*
$$\mathsf{E}[V_\circ] = \frac{1}{q-p}.$$

2. *Show that*
$$\mathsf{E}[V_1] = \frac{p}{q(q-p)}.$$

3. *Show that for a positive integer a we have*
$$\mathsf{E}[V_a] = \frac{1}{q-p}(\frac{p}{q})^a.$$

Exercise 0.2.3.5 *Make a simulation of the simple symmetric random walk on $\mathbb{Z}$ by considering one hundred realizations of the paths from $l = 0$ to $l = 99$ and*

1. *Compute the mean and standard deviation of the number of crossings of the horizontal axis.*

2. *Compute the number Z of changes of sign for the one hundred realizations. How does the empirical probability (on the basis of one hundred realizations) of $P[Z = n]$ for $n = 1, \cdots, 10$ compare with the theoretical prediction?*

Exercise 0.2.3.6 *Make a simulation of the simple symmetric random walk on $\mathbb{Z}$ by considering one hundred realizations of the paths from $l = 0$ to $l = 10,000$. Let U denote the amount of time the random walk is above the horizontal axis. Compute the empirical probability (on the basis of one hundred realizations) of $P[U \geq n]$ for $n = 6000,\ 7,000,\ 8,000$ and 9,000. How does this compare with the theoretical prediction (the arc sine law)?*

0.2.4 An Example from Economics/Game Theory

A simple model in economics, formulated as a finding the or an optimal strategy in a game, can be described as follows: Suppose we have a Markov chain $X_\circ, X_1, \cdots$ with state space S and a real valued function $f : S \rightarrow \mathbb{R}$. The value of the random variable may be thought of as the nominal value of an investment at time l, and the states of the Markov chain reflect various economic parameters such as prices of certain commodities, demand for certain items, consumer confidence etc. Let $\alpha > 0$ and assume that we have the option of cashing in our investment at any time l in which case we receive a sum of $f(X_l)$. Since $f(X_l)$ represents the the nominal value of the investment it is reasonable to consider the quantity $e^{-\alpha l} f(X_l)$ as the real value of the investment at time l. We can make the problem slightly more complicated by introducing a cost function g which is non-negative real valued function on the state space so that real value of the investment at time l would be

$$e^{-\alpha l} f(X_l) - \sum_{j=0}^{l-1} e^{-\alpha j} g(X_j).$$

Since we are allowed to observe the Markov chain at all times, we can devise a strategy for cashing in our investment. For instance one may decide that if the demand for a commodity goes below a certain threshold then we cash in. Or one may use the strategy that if in each of three successive periods the real value of the investment is increased by a certain amount, then it is time to cash in. Since the values of the Markov chain represent the the market fluctuations of the investment, a strategy should be defined as a random variable T which at each time l depends only the values $X_\circ, X_1, \cdots, X_l$ and perhaps other observables until time l. Obviously we do not want T to depend on the knowledge of the future which is not available. Such T is codified as a stopping time for which we will shortly give a precise definition. In order to do so it is useful to first re-examine the notion of conditional probability. The reformulation of the concepts of conditional probability and expectation given below will be useful in other contexts such as martingales.

Recall that we have a probability space Ω on which random variables are defined. For simplicity we may assume the random variables assume only a countable set of values, however, the results are valid for random variables taking values in $\mathbb{R}$. To obtain this greater generality requires a refinement of the notion of conditional probability and expectation which is not of immediate concern to us and is therefore postponed to a later subsection. The assumption that random variables take only countably many values dispenses with the need to introduce the concept of σ-algebras. For random variables X and Y conditional probability $P[X = a|Y]$ is itself a random variable. For every $b \in \mathbb{Z}$ such that $P[Y = b] \neq 0$

we have

$$P[X = a \mid Y = b] = \frac{P[X = a \text{ and } Y = b]}{P[Y = b]}.$$

Similarly conditional expectation $\mathsf{E}[X|Y]$ is a random variable which takes different values for different b's. We can put this in a slightly different language by saying that Y defines a partition of the probability space Ω;

$$\Omega = \bigcup_{b\in\mathbb{Z}} A_b, \quad \text{(disjoint union)} \tag{0.2.4.1}$$

where $A_b = \{\omega \in \Omega \mid Y(\omega) = b\}$. Thus conditional expectation is a random variable which is constant on each piece A_b of the partition defined by Y. With this picture in mind we redefine the notion of conditioning by making use of partitionings of the probability space. So assume we are given a partition of the probability space Ω as in (0.2.4.1), however we do not require that this partition be defined by any specifically given random variable. It is just a partition which somehow has been specified. Each subset A_b is an event and if A_b has non-zero probability, then $P[X|A_b]$ and $\mathsf{E}[X|A_b]$ make sense. It is convenient to introduce a succint notation for a partition. Generally we use $\mathcal{A}$ or $\mathcal{A}_n$ to denote a partition or sequence of partitions which we shall encounter shortly. Notice that each $\mathcal{A}_n$ is a partition of the probability space and the subscript n does not refer to the subsets comprising the partition $\mathcal{A}$. By $\mathcal{A} \prec \mathcal{A}'$ we mean every subset of Ω defined by the partition $\mathcal{A}$ is a union of subsets of Ω defined by the partition $\mathcal{A}'$. In such a case we say $\mathcal{A}'$ is a *refinement* of $\mathcal{A}$. For example if Y and Z are random variables, we define the partition $\mathcal{A}_Y$ as

$$\Omega = \bigcup A_b^Y, \quad \text{where} \quad A_b^Y = \{\omega|Y(\omega) = b\},$$

and similarly for $\mathcal{A}_Z$. Then the collection of intersections $A_b^Y \cap A_c^Z$ defines a partition $\mathcal{A}'$ which is a refinement of both $\mathcal{A}^Y$ and $\mathcal{A}^Z$. The set $A_b^Y \cap A_c^Z$ consists of all $\omega \in \Omega$ such that $Y(\omega) = b$ and $Z(\omega) = c$. For a random variable X notion of $P[X = a|\mathcal{A}]$ simply means that for every subset of positive probability defined by the partition $\mathcal{A}$ we have a number which is the conditional probability of $X = a$ given that subset. Thus $P[X = a|\mathcal{A}]$ itself is a random variable which is constant on each subset A_b of the partition defined by $\mathcal{A}$. Similarly, the conditional expectation $\mathsf{E}[X|\mathcal{A}]$ is a random variable which is constant on each subset defined by the partition $\mathcal{A}$.

Given a partition $\mathcal{A}$ we say a random variable X is $\mathcal{A}$-*admissible* if X is constant on every subset defined by $\mathcal{A}$. For example, the random variables $P[X = a|\mathcal{A}]$ and $\mathsf{E}[X|\mathcal{A}]$ are $\mathcal{A}$-admissible. For an $\mathcal{A}$-admissible random variable X clearly we have

$$\mathsf{E}[X \mid \mathcal{A}] = X. \tag{0.2.4.2}$$

If $\mathcal{A}'$ is a refinement of the partition $\mathcal{A}$, then

$$\mathsf{E}[\mathsf{E}[X \mid \mathcal{A}'] \mid \mathcal{A}] = \mathsf{E}[X \mid \mathcal{A}]. \tag{0.2.4.3}$$

This is a generalization of the statement $\mathsf{E}[\mathsf{E}[X|Y]] = \mathsf{E}[X]$ and its proof is again just a rearrangemet of a series. We will often use the notation $\mathsf{E}_i[Y]$, where Y is a random variable that depends on the Markov chain X_j and $i \in S$, as the conditional expectation

$$\mathsf{E}_i[Y] = \mathsf{E}[Y \mid X_\circ = i].$$

Assume we are given a sequence of partitions $\mathcal{A}_l$ with $\mathcal{A}_l \prec \mathcal{A}_{l+1}$. A random variable $T : \Omega \to \mathbb{Z}_+$ such that the set $\{\omega \in \Omega \mid T(\omega) = l\}$ is a union of subsets of $\mathcal{A}_l$ is called a *Markov time*. If in addition $P[T < \infty] = 1$, then T is called a *stopping time*. In particular, if we let $\mathcal{A}_l$ to be the partition defined by the the random variables $X_\circ, X_1, \cdots, X_l$, then the fact that the strategy or stopping time T depends only on observations X_j up to and including time l is expressed by the Markov time condition that $\{\omega \in \Omega \mid T(\omega) = l\}$ is a union of subsets of $\mathcal{A}_l$. However, by freeing $\mathcal{A}_l$ from the random variables X_j we allow ourselves to incorporate other market conditions into the strategy which may not be directly picked up the Markov chain $X_\circ, X_1, \cdots$. The condition $P[T < \infty] = 1$ is a technical condition which is necessary for finiteness of certain expectations.

With these preliminaries out of the way we go back to our original problem. As a first step, if the cost function is identically 0, then we define the *value of the game* as

$$V(i) = \sup_T \mathsf{E}_i[e^{-\alpha T} f(X_T)],$$

where the supremum is taken over all possible stopping times T. The random variable $e^{-\alpha T} f(X_T)$ represents the the real value of investment if we follow the given strategy T and $V(i)$ is the largest expected value of the investment among all possible strategies. If there is a cost function g which represents the nominal cost of the investment in each time period, then we define the value of the game as

$$V(i) = \sup_T \mathsf{E}_i[e^{-\alpha T} f(X_T) - \sum_{l=0}^{T-1} e^{-\alpha l} g(X_l)],$$

where the summation is over all possible stopping times T. Our problem therefore can be reformulated more precisely in two steps as follows:

1. Compute the value of the game.

2. Can we find a stopping time $T_\circ$ which realizes the value of the game.

The solution to this problem can be stated as the following theorem:

Theorem 0.2.4.1 *Assume f and g are bounded non-negative functions on the state space S. Then there is a minimal solution to the inqualities*

$$\begin{array}{ll} V(i) \geq -g(i) + e^{-\alpha}\sum_j P_{ij}V(j), & i \in S; \\ V(i) \geq f(i), & i \in S. \end{array}$$

The value of the game is this minimal solution. Let $R = \{i \in S \mid V(i) = f(i)\}$. If the state space is finite, then the optimal stopping time T is the time of first visit to R.

The proof of this theorem requires introducing some interesting ideas. First let us understand what it says. Assume the state space is finite which is generally the case in practice, although this finite number can be quite large. Nevertheless the value of the game can be computed by looking at it as a problem in linear programming, viz.,

$$\begin{array}{cc} \text{Minimize } \sum_i V(i), & \text{subject to} \\ V(i) \geq -g(i) + e^{-\alpha}\sum_j P_{ij}V(j), & i \in S; \\ V(i) \geq f(i), & i \in S. \end{array}$$

Thus one can make use of a standard linear programming software for solving for the function V. Note that the reason we can formulate the solution in this fashion is the existence of a minimal solution as stated in theorem 0.2.4.1. When the cardinality of S is small it is not difficult to calculate the the value of the game by carrying out the minimization problem. An examination of some special cases (see for example exercises 0.2.4.1 and 0.2.4.2) give a good idea on how f and V are related. *Roughly speaking*, V looks like a function which passes through some of the local maxima of f (namely, the set R) and the graph of f lies below it. Then theorem 0.2.4.1 describes the optimal strategy by comparing V and f. The optimal strategy T does not depend on the initial state i, $X_\circ = i$ which also intuitively makes sense. For infinite state spaces, the optimal stopping time may not exist in the sense that no strategy may have the the same expected return as the the value of the game, however, one can get arbitrarily close to the value by appropriate choice of T.

The proofs of these statements require the introduction some important machinery and is the given later.

EXERCISES

Exercise 0.2.4.1 *Consider the Markov chain with state space* $\{0,1,2,\cdots,9\}$ *and transition matrix* P *defined as*

$$P_{ij} = \begin{cases} \frac{1}{2}, & \text{if } j = i \pm 1 \text{ and } i = 1, \cdots, 8; \\ 1, & \text{if } i = j = 0; \\ \frac{1}{3}, & \text{if } i = 9 \text{ and } j = 0, 8, 9; \\ 0, & \text{otherwise.} \end{cases}$$

Let f *be defined by*

$$f(0) = 0, \quad f(1) = 3, \quad f(2) = 10, \quad f(3) = 12, \quad f(4) = 8,$$
$$f(5) = 7, \quad f(6) = 11, \quad f(7) = 15, \quad f(8) = 12, \quad f(9) = 5.$$

Let $\alpha = 0$ *and* $g \equiv 0$. *Compute the value of the game and exhibit* f *and* V *on the same graph. What is the optimal stopping time?*

Exercise 0.2.4.2 (Continuation of exercise 0.2.4.1) *With the same notation and hypotheses of exercise 0.2.4.1 except that* $\alpha = \log \frac{10}{9}$ *and*

$$g(i) = \begin{cases} 1, & \text{if } i \neq 0; \\ 0 & \text{if } i = 0. \end{cases}$$

Compute the value of the game and exhibit f *and* V *on the same graph. What is the optimal stopping time?*

0.2.5 Generating Functions Revisited

We have already seen that generating functions can be useful in the investigation of problems in Markov chains. The crux of the matter was the following issues:

1. Given a sequence of quantities $\{a_n\}$ satisfying some recursive relation, one seeks a closed form formula for the expression $\sum a_n\xi^n$. The recursion relation, in many cases, is obtained by conditioning on appropriate random variables.

2. Once the generating function is calculated explicitly, then we need analytical tools to extract information from the generating function about the original problem.

In this subsection we elaborate on these themes in several ways. It is not the case that $\sum a_n\xi^n$ is the only way of constructing a general formula (generating function) which in principle contains all the information about a given sequence of quantities $\{a_n\}$. We show by means of examples that, depending on the situation, other infinite sums may be more suitable for obtaining closed form expressions containing all the required information. No attempt will be made to develop a general theory, and the reader is referred to H. Wilf - *Generatingfunctionology* for a more systematic treatment and references to the literature. Regarding the second issue, we explain some mathematical tools with emphasis on their application to concrete problems.

The first problem we consider has the flavor of one from physics or chemistry and models of this general form exist in abundance. Let $L \subset \mathbb{R}^2$ be a lattice and since the exact nature of the lattice makes a difference we limit ourselves to two cases. Identify $\mathbb{R}^2$ with complex numbers $\mathbf{C}$, let $\zeta = e^{\frac{2\pi i}{3}}$ so that ζ is the unit vector making a $120°$ angle with the positive x-axis. Define

$$L_1 = \{a + ib \mid a, b \in \mathbb{Z}h\}, \quad L_2 = \{a + b\zeta \mid a, b \in \mathbb{Z}\frac{h}{\sqrt{3}}\},$$

where $h > 0$ is fixed constant which is introduced for physical reasons and is useful when approximating a continuous domain by a discrete mesh. The nearest neighbors of the origin in each of the two cases are

$$N_1 = \{\pm h, \pm ih\}, \quad N_2 = \{\pm\frac{h}{\sqrt{3}}, \pm\zeta\frac{h}{\sqrt{3}}, \pm(1+\zeta)\frac{h}{\sqrt{3}}\}.$$

By translations we obtain the nearest neighbors of all the points in the lattice L_j. In case L_1 every point has $q_1 = 4$ nearest neighbors, while every point in L_2 has $q_2 = 6$ nearest neighbors. Points which are nearest neighbors of each other are called *adjacent*. The distance between adjacent points is h in case L_1 and $\frac{h}{\sqrt{3}}$ in case L_2. Consider a bounded open set $\Omega \subset \mathbf{C}$

whose boundary $\partial\Omega$ consists of continuous curves with no self intersections and such that each maximal connected boundary curve consists of line segments joining adjacent points. Examples of such regions are shown in Figures XXXXX. Now assume each site/point is occupied by an atom and atoms which are nearest neighbors of each other are joined by an edge. We also assume that the atoms are identical. We obtain a graph Γ, which includes the boundary of the region, and will play an important role in the sequel. Points not on the boundary are called *interior*. Let N be the number of interior points and B the number of boundary points. We assume that non-adjacent atoms do not exert any force on each other but there is interaction between adjacent ones. Interior atoms can move along perpendiculars to the plane of the lattice, but boundary points are required to remain fixed for all time. The vertical motions of the interior atoms are assumed to be governed by

$$m\frac{d^2\zeta_j}{dt^2} = -K\sum(\zeta_j - \zeta_{j_k}), \qquad (0.2.5.1)$$

where the summation is over all neighbors $j_1, \ldots, j_q$ of the nearest neigbors of the site j, and m denotes the mass of the atom. This is similar to Hooke's law in physics and one reasonably expects the system to display oscillatory motion like vibrations of a drum. For this reason we make the substitution

$$\zeta_j(t) = A_j e^{i\omega t}.$$

If ζ_j satisfies (0.2.5.1) then it is a simple calculation that ω is of the form

$$\omega^2 = \frac{qh^2K}{2m}\lambda,$$

where λ is the eigenvalue of the system

$$\frac{2}{qh^2}\sum_{k=1}^{q}(A_{j_k} - A_j) = -\lambda A_j, \quad \text{subject to} \;\; A_j = 0 \text{ for } j \in \partial\Omega, \qquad (0.2.5.2)$$

where $q = q_1 = 4$ or $q_2 = 6$ depending on the lattice under consideration. This problem can be re-stated in the more succint form of an eigenvalue problem

$$Tw = \mu w \qquad (0.2.5.3)$$

where the $N \times N$ matrix T is defined as

$$T_{jk} = \begin{cases} 1, & \text{if } j \text{ and } k\text{areadjacent;} \\ 0, & \text{otherwise.} \end{cases}$$

(In graph theory T is called the *adjacency matrix* of the graph.) The eigenvalues μ and "frequencies" ω are related by

$$\mu = q - \frac{m}{K}\omega^2.$$

In particular only finitely many frequencies $\omega_1, \ldots, \omega_N$ are possible.

The problem we want to investigate is how the frequencies ω_j can be related to the geometry of the domain Ω. It is reasonable to assert the frequencies (at least in the case of a continuous domain) are the harmonics one hears upon playing on the drum of the given shape. Therefore a more general question is whether, in the discrete framework, the knowledge of the eigenvalues (or frequencies) determines the shape of the drum. While the answer is negative, yet there close relationship between the frequencies and the geometry of the drum which we now investigate.

To investigate this problem it is convenient to introduce the generating function

$$\mathsf{H}(\xi) = \sum e^{\mu_j \xi} = p_\circ + p_1\xi + p_2\frac{\xi^2}{2} + p_3\frac{\xi^3}{3!} + \ldots$$

where $p_s = \sum_j \mu_j^s$ is the sum of the $s^{\rm th}$ powers of the eigenvalues of T. Clearly

$$p_s = \mathrm{Tr}(T^s). \tag{0.2.5.4}$$

The key observation in the evaluation of $\mathrm{Tr}(T^s)$ that it is equal the number of paths starting from an interior point j and returning to j in s steps without entering the boundary, and summing over all j. Perhaps the most conceptually transparent way of seeing the validity of this assertion is by considering the Markov chain with state space (of cardinality $N+B$) the vertices of the graph Γ and where one moves from an interior point to any one of its nearest neighbors with probability $\frac{1}{q}$ and the boundary points are absorbing in the sense that once a path hits a boundary point, it remains there for ever. With this interpretation the validity of the assertion about $\mathrm{Tr}(T^s)$ is immediate. Obviously for $s = 1$ no such path exists and

$$p_\circ = N, \quad p_1 = 0. \tag{0.2.5.5}$$

In case of L_1, $p_{2k+1} = 0$. For $s = 2$ the number of such paths is

$$p_2 = qN - B. \tag{0.2.5.6}$$

In general, p_s can be computed recursively from

$$p_s = N\rho_s - \Lambda_{s,1} - \Lambda_{s,2} - \ldots - \Lambda_{s,s-1}, \tag{0.2.5.7}$$

where $\Lambda_{s,l}$ is the number of closed paths of length s whose first hitting time of the boundary is l, and ρ_s is the number closed paths in the lattice L_i starting at the origin and of length s. We can use this observation to compute p_3 for the lattice L_2. Let τ be the number of triangles which have exactly one vertex on the boundary. Then $\Lambda_{3,2} = 2\tau$, and

$$p_3 = N\rho_3 - \frac{\rho_3}{q}B - 2\tau. \tag{0.2.5.8}$$

This formula can be further simplified. Assume Ω consists c disjoint domains $\Omega_1, \ldots, \Omega_c$ and the domain Ω_α has h_α holes so that its boundary consists of $h_\alpha + 1$ disjoint closed curves without self-intersection. Set $h = \sum_\alpha h_\alpha$. Then (0.2.5.8) can be simplified as

$$c - h = \frac{1}{6}p_3 - \frac{1}{2}p_2 + p_\circ \tag{0.2.5.9}$$

for the lattice L_2. The derivation of this formula is by purely geometric considerations and is left to the reader since it is not relevant to the subject of this text.

It should be emphasized that in the above calculation $\mathsf{H}(\xi)$ was defined as a generating function in terms of analytical data (namely, the frequencies or eigenvalues), yet the coefficients have simple geometric interpretations. Looking at the underlying Markov chain on the associated graph facilitated relating the coefficients of the generating function to a problem on counting paths which then revealed the geometric nature of the coefficients p_k at least for small k.