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 $P^l = (P_{ij}^{(l)})$. Then for $r + s = l$ (r and s non-negative integers) we have

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

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 $e^{\frac{2\pi ij}{n}}$ 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 $j \leftrightarrow e^{\frac{2\pi ij}{n}}$, 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

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

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 $\mathbb{Z}/n$. ♡

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 0.1.2.1:

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 0.1.2.1 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 0.1.2.1, 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 0.1.2.1. 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 0.1.2.1. ♣

We use corollary 0.1.2.1 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 (0.1.2.1). 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 (0.1.2.1) that

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

In other words, corollary 0.1.2.3 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 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

$$\pi P = \pi, \quad \text{with} \ \ \pi_j \geq 0, \ \text{and} \ \sum_{j=1}^{\infty} \pi_j = 1. \tag{0.1.3.1}$$

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 0.1.3.1 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 (0.1.3.1) 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 (0.1.3.1) 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 0.1.3.1 (Brouwer Fixed Point Theorem) - *Let $K \subset \mathbf{R}^n$ be a convex compact*[2] *set, and $F : K \to K$ 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 $v \to vP$. 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

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

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 0.1.3.1 *Assume a solution to (0.1.3.1) 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

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

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

$$\mu_{ij} + \delta_{ij}\mu_j = 1 + \sum_k P_{ik}\mu_{kj}, \quad \text{where } \delta_{ij} = \begin{cases} 1 & \text{if } i = j\ ; \\ 0 & \text{otherwise.} \end{cases} \tag{0.1.3.2}$$

Multiplying (0.1.3.2) by π_i and summing over i (j is fixed) we obtain

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

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 (0.1.3.1) 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.

According to Proposition 0.1.3.1 the stationary distribution $\pi = (\pi_1, \pi_2, \ldots)$ depends only on μ_i's. Therefore it is reasonable to inquire when we can remove the assumption $\mu_{ij} < \infty$ from the hypotheses of the proposition and only retain $\mu_i < \infty$. The following lemma answers this question:

Lemma 0.1.3.1 *Assume the Markov chain $X_\circ, X_1, X_2, \ldots$ is irreducible. Then $\mu_i < \infty$ for all states i implies $\mu_{ji} < \infty$ for all states i, j.*

Proof - Fix states i and j and decompose the underlying probability space Ω into

$$\Omega = \Omega_i \cup \Omega_j, \quad \text{disjoint union,}$$

where Ω_i is the set of paths that return to i without hitting j and Ω_j is its complement, i.e., the set of paths that visit j prior to return to i. Define

$$T_i'(\omega) = I_j T_i,$$

where I_j is the indicator function of the set Ω_j. Clearly $\mathsf{E}[T_i'] \leq E[T_i] < \infty$. Consequently

$$\mathsf{E}[T_i'] = \mathsf{E}[\mathsf{E}[T_i'|T_{ij}]] = \sum_l P[T_{ij} = l](l + \mathsf{E}[T_{ji}])$$

is finite. By irreducibility $P[T_{ij} = l] > 0$ for some l and therefore $\mathsf{E}[T_{ji}] < \infty$. ♣

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

$$\lim_{l\to\infty} P_{ij}^{(l)} = p_j. \tag{0.1.3.3}$$

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 $j^{\rm th}$ 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 $p_j = \pi_j$.

The class of Markov chains for which we prove limiting behavior is that of finite state, aperiodic and irreducible. It is clear that without the assumptions of irreducibility and aperiodicity the theorem below is not valid. The issue of finiteness is more subtle. If the transition matrix P of a Markov chain has the property that all entries of P^l for some l are positive, then we say P or the Markov chain is *regular*. It is a simple argument that regular finite state regular Markov chains are aperiodic and irreducible and conversely (see Corollary 0.1.3.1 below). We prove the following theorem:

Theorem 0.1.3.2 *Let P be the transition matrix of a finite state aperiodic and irreducible Markov chain. Then*

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

The proof of Theorem 0.1.3.2 requires some preparation. First we need to introduce the notion of coupling.

Given two Markov chains $X_\circ, X_1, X_2, \ldots$ and $Y_\circ, Y_1, Y_2, \ldots$ with the state spaces S and transition probabilities $P = (P_{ij})$ and $Q = (Q_{ab})$ we define the *product Markov chain* as one with state space $S\times T$ and transition probability from (i, a) to (j, b) given by $P_{ij}Q_{ab}$. In other words, the product chain is given by the sequence of random variables $(X_\circ, Y_\circ), (X_1, Y_1), \ldots$ with each coordinate evolving in time independently of the other. Now assume $X_\circ, X_1, X_2, \ldots$ and $Y_\circ, Y_1, Y_2, \ldots$ are the same Markov chain except that their initial distributions $X_\circ$ and $Y_\circ$ may be different. In particular the two Markov chains have the same matrix of transition probabilities P. The *coupled chain* is, by definition, the Markov chain with state space $S \times S$ but with transitions defined by the following rule:

$$P_{(a,b)(c,d)} = \begin{cases} P_{ac}P_{bd} & \text{if } a \neq b; \\ P_{ac} & \text{if } a = b \text{ and } c = d; \\ 0 & \text{if } a = b \text{ and } c \neq d. \end{cases}$$

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.

The argument leading to the proof of the existence of $\lim_l P^l$ relies on the following facts:

- For the coupled chain the state space has the decomposition $S \times S = T \cup D$ where T, the set of transient states, consists of non-diagonal states (a,b), $(a \neq b)$, D, the diagonal states, is precisely is the set of recurrent states, and with probability 1 every path enters D.

It is clear from the irreducibility all states in D communicate. The notion of aperiodicity (i.e., $d(i) = 1$) plays an important role in the theory of Markov chains. There is a basic fact from elementary number theory which relates aperiodicity to the theory of Markov chains, namely

Lemma 0.1.3.2 *Let $l_1, l_2, \dots$ be positive integers with gcd$= 1$. Then there is an integer L such that for all $l \geq L$ there are non-negative integers $\alpha_1, \alpha_2, \dots$ such that*

$$l = \alpha_1 l_1 + \alpha_2 l_2 + \dots$$

The proof of this lemma is elementary and irrelevant to to our context and is therefore omitted. Applications of this lemma will be given shortly.

Lemma 0.1.3.3 *With the above notation and hypotheses, for every $(a,b) \in T$ the set of paths starting at (a,b) and terminating in (a,a) has positive probability.*

Proof - Since the Markov chain $X_\circ, X_1, \dots$ is irreducible, there is m such that $P^{(m)}_{ba} > 0$. Aperiodicity of the Markov chain and Lemma 0.1.3.2 imply the existence of L such that for all $l \geq L$ we have $l = \sum \alpha_j l_j$, $\alpha_j \geq 0$, and

$$P^{(l)}_{aa} \geq P^{(\alpha_1 l_1)}_{aa} P^{(\alpha_2 l_2)}_{aa} \dots > 0$$

Therefore for all $l, l' \geq L$ we have

$$P^{(l)}_{aa} > 0, \quad P^{(m+l')}_{ba} > 0.$$

The required result follows. ♣

Lemma 0.1.3.4 *With the above notation and hypotheses, non-diagonal states are transient.*

Proof - Let $(a, b) \in T$. It follows from Lemma 0.1.3.3 that there is a smallest positive integer l such that the set of paths Ω'_{ab} of length l starting at (a, b) and terminating in D has positive probability. Paths in Ω'_{ab} do not visit (a, b) since the minimality assumption on l precludes the possibility of visiting (a, b) prior to hitting D and once a path enters D it never leaves it. This implies (a, b) is necessarily transient. ♣

To complete the proof of • recall that in a finite state Markov chain there is a recurrent state and by the irreducibility hypothesis all states are recurrent. Therefore the diagonal is precisely the set of recurrent states. A transient state is visited only finitely many times with probability 1. Therefore the set of paths that eventually enter the diagonal has probability 1.

The above argument also implies the following general fact:

Corollary 0.1.3.1 *The transition matrix of a finite state, aperiodic and irreducible Markov chain is regular.*

Proof - By aperiodicity $P_{aa}^{(l)} > 0$ for all sufficiently large and $P_{ab}^{(m)} > 0$ for some m. Therefore $P(m + l)_{ab} > 0$ for all l sufficiently large. ♣

Proof of Theorem 0.1.3.2 - 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 •, with probability 1 paths of the coupled chain enter D. We have

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

Since with probability 1 paths enter D, for every $\epsilon > 0$ we have $P[T > l] < \epsilon$ for l sufficiently large. For such l we therefore have

$$|P[X_l = j,\ T > l] - P[Y_l = j,\ T > l]| < \epsilon.$$

On the other hand the events $\{X_l = j,\ T \leq l\}$ and $\{Y_l = j,\ T \leq l\}$ at both identical with the event $\{X_l = j, Y_l = j\}$ and therefore

$$P[X_l = j,\ T \leq l] - P[Y_l = j,\ T \leq l] = 0.$$

Therefore for l sufficiently large $|P_{ij}^{(l)} - \pi_j| < \epsilon$. ♣

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, assume the Markov chain $X_\circ, X_1, \ldots$ has a unique stationary distribution (e.g. hypothesis of Proposition **??** are fulfilled). 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 0.1.3.5 *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 0.1.3.5 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 0.1.3.2 *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. ♣

In general for a Markov chain $\mathsf{E}[T_i]$ may be ∞. In fact in the next section we will show that while all states are recurrent for the simple symmetric random walk on $\mathbb{Z}$, $\mathsf{E}[T_i] = \infty$. For a finite state aperiodic irreducible Markov chain not only the expectations, but all moments of T_i and T_{ij} are finite. This follows from the following proposition:

Proposition 0.1.3.2 *Let $X_\circ, X_1, \ldots$ be an irreducible, aperiodic and finite state Markov chain*[3]. *Then for all states i, j there is $\gamma < 1$ and c such that*

$$P[T_{ij} > l] < c\gamma^l,$$

for all l. Similar statement is valid for T_i. In particular all moments of T_{ij} and T_i exist.

The proof of Proposition 0.1.3.2 requires some preliminaries. We need some preliminary considerations for the proof of proposition **??**. The hypotheses imply that the matrix of transition probabilities P is a regular $N \times N$ matrix. For definiteness set $j = N$. Define $(N-1) \times (N-1)$ matrices $Q^{(l)} = (Q^{(l)}_{ij})$, where $1 \leq i, j \leq N-1$ by

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

[3] The assumption of aperiodicity is inessential. It is made here for simplicity of exposition. For a general finite state Markov chain $\mathsf{E}[T_{ij}]$ may be infinite if i is recurrent and j is transient or does not communicate with i. The modification of the statement of the theorem for general finite state Markov chains is straightforward.

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

$$Q_{ij}^{(l)} = \sum_{j_1 \neq N} \sum_{j_2 \neq N} \cdots \sum_{j_{l-1} \neq N} P_{ij_1} P_{j_1 j_2} \cdots P_{j_{l-1} N}. \tag{0.1.3.4}$$

We need two simple technical lemmas.

Lemma 0.1.3.6 *If P is positive then there is $\rho < 1$ such that*

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

and consequently $\sum_{l=1}^{\infty} \sum_{j=1}^{N-1} Q_{ij}^{(l)}$ 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 (0.1.3.4) that

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

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

Lemma 0.1.3.7 *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 0.1.3.2 Since P is 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 0.1.3.6 (or more precisely its proof) implies that $\sum_{j=1}^{N-1} Q_{ij}^{(mn)} < \rho^n < 1$. By lemma 0.1.3.7 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. ♣

EXERCISES

Exercise 0.1.3.1 *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 0.1.3.2 *Consider the Markov chain with state space as in exercise 0.1.3.1, but modify the rule* • *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 0.1.3.3 *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 0.1.3.4 *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 0.1.3.5 *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 0.1.3.6 *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 0.1.3.7 *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 0.1.3.8 *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 0.1.3.9 *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 0.1.3.10 *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 0.1.3.11 *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 0.1.3.12 *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 0.1.3.13 *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 0.1.3.14 *Let* P *be the* 6×6 *matrix of the Markov chain chain in exercise 0.1.3.2. 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 0.1.3.2.*

Exercise 0.1.3.15 *Assume we are in the situation of exercise 0.1.3.3 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 0.1.3.2.*

0.1.4 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. We have already seen in the subsection "Classification of States" that the generating functions P_{ij} and F_{ij} and the relationship between them provided important implications about Markov chains.

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

$$\mathsf{F}_X(\xi) = \sum_{n=\circ}^{\infty} f_X(n)\xi^n = \mathsf{E}[\xi^X]. \tag{0.1.4.1}$$

This infinite series converges for $|\xi| < 1$ since $0 \le f_X(n) \le 1$ 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

$$\mathsf{E}[\mathsf{E}[Z \mid Y]] = \mathsf{E}[Z], \tag{0.1.4.2}$$

to evaluate $\mathsf{E}[Z]$ directly, and then draw conclusions about the random variable X. Examples 0.1.4.2 and 0.1.4.4 are simple demonstrations of this point.

Example 0.1.4.1 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

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

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

Proposition 0.1.4.1 *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 0.1.4.2 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 0.1.4.3 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

$$P[X_1 = k_1, X_2 = k_2] = \binom{n}{k_1, k_2} p_1^{k_1} p_2^{k_2} p_3^{n-k_1-k_2}. \qquad (0.1.4.3)$$

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 (0.1.4.3) we obtain

$$\mathsf{F}_{X_1,X_2}(\xi, \eta) = (p_1\xi + p_2\eta + p_3)^n, \qquad (0.1.4.4)$$

for the generating function of the trinomial random variable. ♠

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

$$\mathsf{E}[X] = (\frac{d\mathsf{F}_X(\xi)}{d\xi})_{\xi=1^-}, \quad \text{if } P[X = \infty] = 0. \qquad (0.1.4.5)$$

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

$$\mathsf{Var}[X] = \left[\frac{d^2\mathsf{F}_X(\xi)}{d\xi^2} + \frac{d\mathsf{F}_X(\xi)}{d\xi} - \left(\frac{d\mathsf{F}_X(\xi)}{d\xi}\right)^2\right]_{\xi=1^-}. \qquad (0.1.4.6)$$

Another useful relation involving generating functions is

$$\sum_n P[X > n]\xi^n = \frac{1 - \mathsf{E}[\xi^X]}{1 - \xi}. \qquad (0.1.4.7)$$

The identities are proven by simple and formal manipulations. For example to prove (0.1.4.7), 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 (0.1.4.7). The coefficient $P[X>n]$ of ξ^n on left hand side of (0.1.4.7) is often called *tail probabilities*. We will see examples of tail probabilities later.

Example 0.1.4.4 As an application of (0.1.4.5) 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 (0.1.4.5) 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 $1\leq j\leq m$

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

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

$$\mathsf{F}_X(\xi)=\mathsf{E}[\xi^X]=\frac{p^m\xi^m(1-p\xi)}{1-\xi+qp^m\xi^{m+1}}. \tag{0.1.4.8}$$

Using (0.1.4.5), 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

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

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

$$1 + Y(H+T) = X + Y. \tag{0.1.4.9}$$

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

$$YA \;=\; S(1 + A^1\Delta_{n-1} + A^2\Delta_{n-2} + \ldots + A^{n-1}\Delta_1). \tag{0.1.4.10}$$

For example in this case we get

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

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 (0.1.4.9) and (0.1.4.10) 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

$$A(1 - X) = X\Big[A + \big(1 + \sum_{j=1}^{n-1} A^j \Delta_j\big)\big(1 - H - T\big)\Big] \tag{0.1.4.11}$$

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

$$\mathsf{F}(\xi) = \frac{A(\xi)}{A(\xi) + \big(1 + \sum_{j=1}^{n-1} A^j(\xi)\Delta_{n-j}\big)\big(1 - \xi\big)}. \tag{0.1.4.12}$$

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 (0.1.4.11) one easily obtains the expectation and variance of τ. In fact we obtain

$$\mathsf{E}[\tau] = \frac{1 + \sum_{j=1}^{n-1} A^j(1)\Delta_{n-j}}{A(1)}, \quad \mathsf{Var}[\tau] = \mathsf{E}[\tau]^2 - \frac{1 + \sum_{j=1}^{n-1}(2j-1)A^j \Delta_{n-j}}{A(1)}. \tag{0.1.4.13}$$

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

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

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

$$1 + Y(H+T) = X + Y. \qquad (0.1.4.14)$$

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

$$YA \;=\; S(1 + A^1\Delta_{n-1} + A^2\Delta_{n-2} + \ldots + A^{n-1}\Delta_1). \qquad (0.1.4.15)$$

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 (0.1.4.14) and (0.1.4.15) 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

$$A(1-X) = X\Big[A + \big(1 + \sum_{j=1}^{n-1} A^j\Delta_j\big)\big(1 - H - T\big)\Big] \qquad (0.1.4.16)$$

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

$$\mathsf{F}(\xi) = \frac{A(\xi)}{A(\xi) + \big(1 + \sum_{j=1}^{n-1} A^j(\xi)\Delta_{n-j}\big)\big(1-\xi\big)}. \qquad (0.1.4.17)$$

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

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

for the generating function of the time of the first appearance of $HHTHH$. From (0.1.4.16) one easily obtains the expectation and variance of τ. In fact we obtain

$$\mathsf{E}[\tau] = \frac{1+\sum_{j=1}^{n-1} A^j(1)\Delta_{n-j}}{A(1)}, \quad \mathsf{Var}[\tau] = \mathsf{E}[\tau]^2 - \frac{1+\sum_{j=1}^{n-1}(2j-1)A^j\Delta_{n-j}}{A(1)}. \qquad (0.1.4.18)$$

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:

$$q_n = \frac{b_1}{\alpha_1^{n+1}} + \frac{b_2}{\alpha_2^{n+1}} + \cdots + \frac{b_m}{\alpha_m^{n+1}} \qquad (0.1.4.19)$$

To see how (0.1.4.19) 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 0.1.4.5 To illustrate the above idea of using partial fractions consider example 0.1.4.4 above. We can write the generating function (0.1.4.8) 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,

$$\alpha f_{j+1} + \beta f_j + \gamma f_{j-1} = 0, \tag{0.1.4.20}$$

it is straighforward to explicitly compute the generating function $\mathsf{F}(\xi)$. In fact, it follows from (0.1.4.20) 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

$$\mathsf{F}(\xi) = \frac{\alpha f_\circ + (\alpha f_1 + \beta f_\circ)\xi}{\alpha + \beta\xi + \gamma\xi^2}. \tag{0.1.4.21}$$

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

$$(j+1)f_{j+1} - j f_j - f_{j-1} = 0, \quad j = 2, 3, \cdots \tag{0.1.4.22}$$

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

$$(1-\xi)\frac{d\mathsf{F}}{d\xi} - \xi\mathsf{F} = f_1 + (f_1 + 2f_2)\xi. \tag{0.1.4.23}$$

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

$$f_{j+1} = P[A_{j+1} \mid B_{j+1}]\frac{j}{j+1}. \tag{0.1.4.24}$$

On the other hand,

$$P[A_{j+1} \mid B_{j+1}] = \frac{1}{j}f_{j-1} + \frac{j-1}{j}P[A_j \mid B_j]. \tag{0.1.4.25}$$

Equations (0.1.4.24) and (0.1.4.25) imply validity of (0.1.4.22) and (0.1.4.23) with

$$f_1 = 0; \quad f_2 = \frac{1}{2}. \tag{0.1.4.26}$$

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

$$f_k = \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^k}{k!}. \tag{0.1.4.27}$$

Thus for k large, the probability of no matches is approximately $\frac{1}{e}$. Of course one can derive (0.1.4.27) by a more elementary (but substantially the same) argument. ♠

Example 0.1.4.7 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

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

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 0.1.2.1 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

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

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 0.1.4.1 *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 0.1.4.8 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

$$\mathsf{F}_{\circ n}(\xi) = [\mathsf{F}_{\circ 1}(\xi)]^n. \qquad (0.1.4.28)$$

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 (0.1.4.28) imply

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

Solving the quadratic equation we obtain

$$\mathsf{F}_{\circ 1}(\xi) = \frac{1 - \sqrt{1 - 4pq\xi^2}}{2q\xi}. \qquad (0.1.4.29)$$

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 0.1.4.9 We shown that the simple symmetric random walk on $\mathbb{Z}$ is recurrent and exercise 0.1.4.11 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 $\sum P_{\circ\circ}^{(l)}$. We have $P_{\circ\circ}^{(2l+1)} = 0$ and

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

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

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

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

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

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

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

Now recall Stirling's formula

$$n! = \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} e^{\rho(n)}, \quad where \frac{1}{12(n+\frac{1}{2})} < \rho(n) < \frac{1}{12n}. \tag{0.1.4.30}$$

Applying Stirling's formula we obtain the bound

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

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

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

Exercise 0.1.4.2 *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 0.1.4.3 *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 $p_l = P^{(l)}_{AA}$ 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 0.1.4.4 *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 0.1.4.5 *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 0.1.4.6 *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 0.1.4.7 *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 0.1.4.8 *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 0.1.4.9 *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*

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

Exercise 0.1.4.10 *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 0.1.4.11 *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 0.1.4.9.)