0 RECALL:

From last week:

DEFINITION 0.1. *Let μ be any probability on S. Let $\mathbf{P} = [p_{ij}]_{i,j\in S}$ be a matrix for which the entries satisfiy $p_{ij} \in [0,1]$ for all $i, j \in S$ and $\sum_{k=1}^{\infty} p_{ik} = 1$ for each $i \in S$.*

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and for $n \geq 0$, let $X_n : \Omega \to S$ be a sequence of random variables whose joint distribution is given by

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_k = i_k) = \mu(\{i_0\})p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{k-1} i_k}, \tag{0.1}$$

for all $k \geq 0$. Then X_n is said to be a Markov chain on state space S, with initial distribution μ, and transition matrix $\mathbf{P}$.

For the precise construction of the probability space and Markov chain in Definition 0.1, we refer the reader to [AS09, Theorem 4.3.5] when S is finite and [AS09, Theorem 4.4.4] when S is countably infinite. We will assume the result and shall now detail some of the properties of a Markov chain X_n.

(a) (Distribution of X_n) Using the definition of Probability, (0.1), and properties of the transition matrix $\mathbf{P}$ we have

$$\begin{aligned}
\mathbb{P}(X_0 = i_0) &= \sum_{j\in S} \mathbb{P}(X_0 = i_0, X_1 = j) \\
&= \sum_{j\in S} \mu(\{i_0\})p_{i_0,j} \\
&= \mu(\{i_0\}) \sum_{j\in S} p_{i_0,j} \\
&= \mu(\{i_0\}).
\end{aligned}$$

Hence $X_0 \sim \mu$, i.e. the distribution of X_0 is given by μ which justifies describing μ as the initial distribution. To calculate the probability X_n is found in a state j we can decompose the event taking the disjoint union over all possible initial states and all

possible paths from those initial states that finish at state j at time n. Specifically, for $n \geq 1$, using the definition of Probability and (0.1),

$$\begin{aligned}
\mathbb{P}(X_n = j) &= \mathbb{P}(\cup_{i_0,i_1,\ldots i_{n-1}\in S}\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1} X_n = j\}) \\
&= \sum_{i_0,i_1,\ldots i_{n-1}\in S} \mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = j) \\
&= \sum_{i_0,i_1,\ldots i_{n-1}\in S} \mu(\{i_0\}) \left(\prod_{k=1}^{n-1} p_{i_{k-1}i_k}\right) p_{i_{n-1},j}. \qquad (0.2)
\end{aligned}$$

In particular, note that the distribution of X_n is determined solely by μ and the transition matrix $\mathbf{P}$.

(b) (Markov Property) Every Markov chain has an underlying memoryless property. For $n \in \mathbb{N}$ let $i_0, i_1, \ldots, i_{n-2}, i, j \in S$ is such that

$$\mathbb{P}(X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) > 0,$$

then

$$\begin{aligned}
&\mathbb{P}(X_n = j | X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) \\
&= \frac{\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0)}{\mathbb{P}(X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, X_0 = i_0)} \\
&= \frac{\mu(\{i_0\}) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i_{n-1}} p_{ij}}{\mu(\{i_0\}) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i_{n-1}}} \\
&= p_{ij} \qquad (0.3)
\end{aligned}$$

and as $\mathbb{P}(X_{n-1} = i) \geq \mathbb{P}(X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) > 0$, we may calculate the conditional probability $\mathbb{P}(X_n = j | X_{n-1} = i) = \frac{\mathbb{P}(X_n = j, X_{n-1} = i)}{\mathbb{P}(X_{n-1} = i)}$ by using

(0.2) and decomposing the probability into all possible paths for which $X_{n-1} = i$ and $X_n = j$. Specifically,

$$\begin{aligned}
&\mathbb{P}(X_n = j | X_{n-1} = i) \\
&= \frac{\mathbb{P}(X_n = j, X_{n-1} = i)}{\mathbb{P}(X_{n-1} = i)} \\
&= \frac{\mathbb{P}(\cup_{i_0,i_1,\ldots i_{n-2}\in S}\{X_n = j, X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_1 = i_1, X_0 = i_0\})}{\mathbb{P}(\cup_{i_0,i_1,\ldots i_{n-2}\in S}\{X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_1 = i_1, X_0 = i_0\})} \\
&= \frac{\sum_{i_0,i_1,\ldots,i_{n-2}\in S} \mu(\{j_0\}) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i} p_{ij}}{\sum_{i_0,i_1,\ldots,i_{n-2}\in S} \mu(\{j_0\}) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i}} \\
&= p_{ij} \cdot \frac{\sum_{i_0,i_1,\ldots,i_{n-2}\in S} \mu(\{i_0\}) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i}}{\sum_{i_0,i_1,\ldots,i_{n-2}\in S} \mu(\{i_0\}) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i}} \\
&= p_{ij} \qquad (0.4)
\end{aligned}$$

Hence from (0.3) and (0.4) we have

$$\mathbb{P}(X_n = j | X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) = p_{ij} = \mathbb{P}(X_n = j | X_{n-1} = i) \qquad (0.5)$$

So, it does not matter how the chain arrived at state i for deciding how likely it is that it will jump to state j in the next step. Thus for a Markov chain, the past and the future are independent given the present. This is referred to as the Markov property of X_n and can be rephrased below

(c) (n-th step Transition Probability Matrix) From (0.5), we have for $n \geq 1, i, j \in S$

$$\mathbb{P}(X_n = j | X_{n-1} = i) = p_{ij}$$

if $\mathbb{P}(X_{n-1} = i) > 0$. This is why the entries of $\mathbf{P}$ are called the (one-step) transition probabilities and the matrix $\mathbf{P}$ the (one-step) transition probability matrix of the Markov chain X_n.

Let $\mathbf{P}^n$ be the n-th power of the matrix $\mathbf{P}$ and $p_{ij}^{[n]}$ denote the (i,j)-th element of $\mathbf{P}^n$. Suppose $i \in S$ is such that $\mathbb{P}(X_0 = i) > 0$. For any $j \in S$, we have

$$\begin{aligned}
\mathbb{P}(X_n = j | X_0 = i) &= \frac{1}{\mathbb{P}(X_0 = i)} P(X_n = j, X_0 = i) \\
&= \frac{1}{\mu(\{i\})} \sum_{i_1,\ldots,i_{n-1} \in S} \mathbb{P}(X_n = j, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i) \\
&= \frac{1}{\mu(\{i\})} \sum_{i_1,\ldots,i_{n-1} \in S} \mu(\{i\}) p_{ii_1} p_{i_1 i_2} \cdots p_{i_{n-2} i_{n-1}} p_{i_{n-1} j} \\
&= \sum_{i_1,\ldots,i_{n-1} \in S} p_{ii_1} p_{i_1 i_2} \cdots p_{i_{n-2} i_{n-1}} p_{i_{n-1} j}.
\end{aligned}$$

Now by definition of matrix multiplication we have that

$$p_{ij}^{[n]} = \sum_{i_1,\ldots,i_{n-1} \in S} p_{ii_1} p_{i_1 i_2} \cdots p_{i_{n-2} i_{n-1}} p_{i_{n-1} j}.$$

Thus,

$$\mathbb{P}(X_n = j \mid X_0 = i) = p_{ij}^{[n]} \tag{0.6}$$

Therefore the entries $p_{ij}^{[n]}$ are called the n-step transition probabilities for the chain since they are the probabilities that a chain in state i at time 0 will be in state j after exactly n steps. Likewise the matrix $\mathbf{P}^n$ is referred to as the n-step transition probability matrix of the chain.

This notation even makes sense when $n = 0$. The matrix $\mathbf{P}^0$ is the identity matrix. Therefore $p_{ij}^{[0]}$ will be 1 if $i = j$ and will be 0 if $i \lneq j$.

1 CONTINUED

(d) (Chapman-Kolmogrov Equation) For $r, s \geq 0$ and $r + s = n$, we have

$$\mathbf{P}^n = \mathbf{P}^{r+s} = \mathbf{P}^r \mathbf{P}^s.$$

Standard rules governing multiplication of matrices implies that

$$p_{ij}^{[n]} = \sum_{k \in S} p_{ik}^{[r]} p_{kj}^{[s]}. \tag{1.7}$$

By (0.6), we can conclude

$$\mathbb{P}(X_n = j | X_0 = i) = \sum_{k \in S} \mathbb{P}(X_r = k | X_0 = i) \mathbb{P}(X_s = j | X_0 = k) \tag{1.8}$$

where the sum is understood to include only those terms for which the conditional probabilities are defined. This is referred to as the Chapman-Kolmogorov equation. It may be interpreted probabilistically as a calculation of the n-step transition probability $p_{ij}^{[n]}$ by first decomposing the probability according to the location of the chain at a specific point in time s and then using the Markov property to restart the chain at that point in time.

EXAMPLES OF MARKOV CHAINS

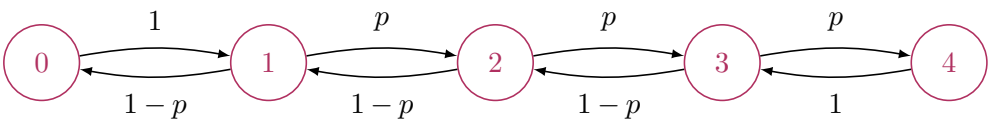
As we have seen above, to define a Markov chain X_n it is enough to specify its state space S, the initial distribution μ of X_0, and transition matrix $\mathbf{P}$. We now present some examples of Markov chains.

EXAMPLE 1.1. (*Simple Random walk*) Let X_n be a Markov chain on S with transition matrix $\mathbf{P}$ and initial distribution μ. Assume that μ is any probability on S.

(a) Let $S = \{0, 1, 2, 3, 4\}$ and the associated transition matrix to be

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

The above dynamics of the chain can be understood via a directed graph with S as the vertex set.

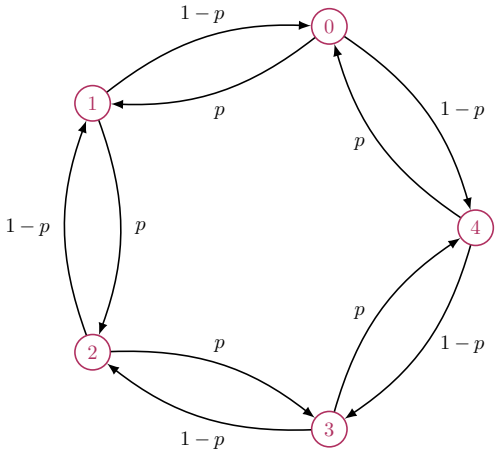


The Markov chain chooses its nearest neighbours with equal probability for vertices $1, 2, 3$ and gets reflected at 0 to 1 and at 4 to 3. The Markov chain is called the *simple random walk on S reflected at* $0, 4$.

(b) Let $S = \{0, 1, 2, 3, 4\}$ and the associated transition matrix to be

$$\mathbf{P} = \begin{bmatrix} 0 & p & 0 & 0 & 1-p \\ 1-p & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & p \\ p & 0 & 0 & 1-p & 0 \end{bmatrix}.$$

The above dynamics of the chain can be understood via a directed graph with S as the vertex set.

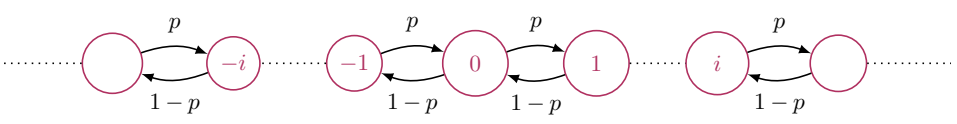


The Markov chain chooses its nearest neighbours with equal probability and is called the *simple random walk on the circle* S.

(c) Let $S = \mathbb{Z}$ and the associated transition matrix to be $\mathbf{P} = [p_{ij}]_{i,j\in\mathbb{Z}}$ having entries

$$p_{ij} = \begin{cases} p & \text{if } j = i+1 \\ 1-p & \text{if } j = i-1 \\ 0 & \text{otherwise} \end{cases}$$

The above dynamics of the chain can be understood via a directed graph with S as the vertex set.



The Markov chain is called a *simple random walk on the integers.*

■

The above example of random walks are fundamental in nature and are widely used in many areas including Statistical Physics and Biology. The Ehrenfest model was originally proposed by Tatiana and Paul Ehrenfest to explain the second law of thermodynamics. It can be formulated as Markov chain. We present that next.

EXAMPLE 1.2. (*Ehrenfest Model*) Suppose there are totally $2N$ balls distributed equally across two urns. At each time step one of the $2N$ balls is chosen at random and moved from the urn it was in to the other urn. Let X_n be the number of balls in urn 1 at time n. Then X_n is a Markov chain on state space $S = \{0, 1, \ldots, 2N\}$ with $X_0 = N$ and transition matrix $\mathbf{P} = [p_{ij}]_{0 \le i,j \le 2N}$ given by

$$p_{ij} = \begin{cases} \frac{i}{2N} & \text{if } j = i-1, i \neq 0 \\ 1 - \frac{i}{2N} & \text{if } j = i+1, i \neq 2N \\ 1 & \text{if } (i,j) \in \{(0,1), (2N, 2N-1)\} \\ 0 & \text{otherwise} \end{cases}$$

The above example is referred to as the Ehrenfest model for diffusion of gases. ■

We will discuss queue and birth-death chains in greater generality in subsequent sections. Here we discuss a special case that is motivated by population genetics.

EXAMPLE 1.3. (*Moran Model*) Assume that a population consists of m individuals of two types, a and A. At each time $n \ge 1$: we choose randomly one individual from the population and add a new individual of the same type; secondly we choose randomly one individual from the population and remove it. We denote by X_n the number of type A individuals in the population at time n. Then X_n is a Markov chain on $S = \{0, 1, 2, \ldots, m\}$ with transition matrix $\mathbf{P} = [p_{ij}]_{0 \le i,j \le m}$ given by .

$$p_{ij} = \begin{cases} \frac{i(m-i)}{m^2} & \text{if } j = i-1, i \neq 0 \\ \frac{i^2+(m-i)^2}{m^2} & \text{if } j = i \\ \frac{i(m-i)}{m^2} & \text{if } j = i+1, i \neq m \\ 0 & \text{otherwise} \end{cases}$$

The above Markov chain is referred to as the Moran Model and is a particular birth and death chain on S.

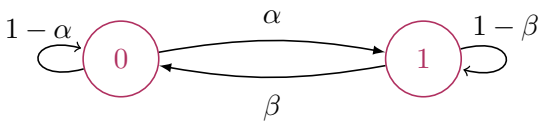
■

In the next example we consider a two state Markov chain where we shall explicitly compute the distribution of X_n and its limit.

EXAMPLE 1.4. Let $0 < \alpha, \beta < 1$. Consider a Markov chain X_n on state space $S = \{0, 1\}$,with $P(X_0 = 0) = \mu_0 \in (0, 1), \mathbb{P}(X_0 = 1) = \mu_1 \in (0, 1)$, and transition matrix

$$\mathbf{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}.$$

The above dynamics of the chain can be understood via a directed graph with S as the vertex set.



Suppose we wish to determine the n-step transition probabilities $p_{ij}^{[n]}$. From the Chapman-Kolmogorv equations applied to the relation $\mathbf{P}^n = \mathbf{P}^{n-1}\mathbf{P}$, we obtain

$$\begin{aligned} p_{00}^{[n]} &= p_{00}^{[n-1]}(1-\alpha) + p_{01}^{[n-1]}\beta \\ &= p_{00}^{[n-1]}(1-\alpha-\beta) + \beta, \end{aligned}$$

w as $p_{00}^{[n-1]} + p_{01}^{[n-1]} = 1$. This formula provides a recursive method for determining $p_{00}^{[n]}$ from $p_{00}^{[n-1]}$ and there is a standard technique to solve such recursions explicitly. As, $\mathbf{P}^0 = I$ we have $p_{00}^{[0]} = 1$. First we check if there is a constant solution, say $p_{00}^{[n]} = p$. Then p would have to satisfy,

$$p = p(1-\alpha-\beta) + \beta \qquad \Longrightarrow \qquad p = \frac{\beta}{\alpha+\beta}$$

Set $x_n = p_{00}^{[n]} - \frac{\beta}{\alpha+\beta}$. So $x_0 = 1 - \frac{\beta}{\alpha+\beta} = \frac{\alpha}{\alpha+\beta}$, and x_n can be expressed as

$$\begin{aligned} x_n &= p_{00}^{[n]} - \frac{\beta}{\alpha+\beta} \\ &= p_{00}^{[n-1]}(1-\alpha-\beta) + \beta - \frac{\beta}{\alpha+\beta} \\ &= x_{n-1}(1-\alpha-\beta). \end{aligned}$$

This implies that

$$x_n = (1-\alpha-\beta)^n x_0 = \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n.$$

Therefore

$$p_{00}^{[n]} = \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n + \frac{\beta}{\alpha+\beta}. \tag{1.9}$$

Imitating the above we can also obtain,

$$p_{10}^{[n]} = \frac{\beta}{\alpha+\beta} - \frac{\beta}{\alpha+\beta}(1-\alpha-\beta)^n. \tag{1.10}$$

With these transition probabilities we can then determine how likely it is that the chain is in any given state at time n.

Note,

$$\begin{aligned} P(X_n = 0) &= P(X_n = 0, X_0 = 0) + P(X_n = 0, X_0 = 1) \\ &= P(X_n = 0 \mid X_0 = 0)P(X_0 = 0) + P(X_n = 0 \mid X_0 = 1)P(X_0 = 1) \\ &= p_{00}^{[n]}\mu(\{0\}) + p_{10}^{[n]}\mu(\{1\}) \\ &= (p_{00}^{[n]} - p_{10}^{[n]})\mu(\{0\}) + p_{10}^{[n]}, \end{aligned} \tag{1.11}$$

where we used $\mu(\{0\}) + \mu(\{1\}) = 1$.

Substituting (1.9) and (1.10) into (1.11) we have

$$\mathbb{P}(X_n = 0) = \frac{\beta}{\alpha+\beta} + (1-\alpha-\beta)^n(\mu(\{0\}) - \frac{\beta}{\alpha+\beta}), \tag{1.12}$$

and using $\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0)$ we have

$$\mathbb{P}(X_n = 1) = \frac{\alpha}{\alpha+\beta} + (1-\alpha-\beta)^n(\mu(\{1\}) - \frac{\alpha}{\alpha+\beta}), \tag{1.13}$$

With a clear explicit expression for the probabilities two observations are immediate:

(a) As $\mid 1-\alpha-\beta \mid < 1$ we have,

$$\lim_{n\to\infty} \mathbb{P}(X_n = 0) = \frac{\beta}{\alpha+\beta} \text{ and } \lim_{n\to\infty} \mathbb{P}(X_n = 1) = \frac{\alpha}{\alpha+\beta}.$$

Thus regardless of the initial distribution μ the distribution of the chain on $\{0,1\}$ a $(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta})$. This is called the *limiting distribution.*

(b) Secondly, suppose $\mu(\{0\}) = \frac{\beta}{\alpha+\beta}$ then $\mu(\{1\}) = \frac{\alpha}{\alpha+\beta}$ and this forces,

$$\mathbb{P}(X_n = 0) = \frac{\beta}{\alpha+\beta} \text{ and } \mathbb{P}(X_n = 1) = \frac{\alpha}{\alpha+\beta}$$

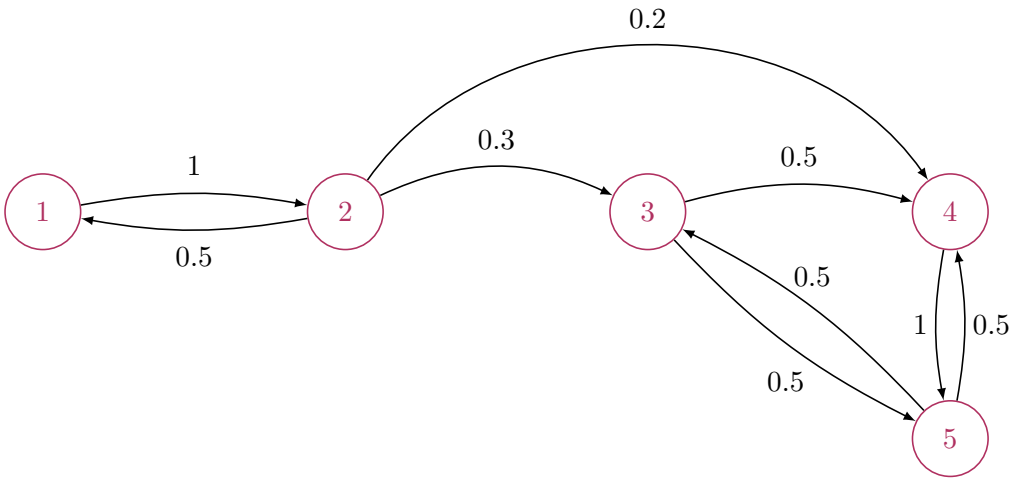
for all $n \geq 1$. Thus if you start with this particular initial distribution then the law of X_n does not change for any $n \geq 1$. Such a distribution is called *stationary distribution.*

■

EXAMPLE 1.5. Consider a Markov chain X_n on state space $S = \{1, 2, 3, 4, 5\}$, $X_0 \sim$ Uniform $\{1, 2, 3, 4, 5\}$, and transition matrix

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

The above dynamics of the chain can be understood via a directed graph with S as the vertex set.



From the graph we can make the following elementary observations. Once the chain reaches a state in the set $\{3, 4, 5\}$ it remains within that set of states. (The states $\{1, 2\}$ are not accessible after this point in time). Intuitively, the fraction of time the chain spends in the states $\{1, 2\}$ should tend to zero as time goes to infinity. This does not seem to be the case with $\{3, 4, 5\}$. In fact, one can show that the fraction of time spent at 5 is twice that of 3 which is twice that of 4. Alternatively, using Linear Algebra, one could try to show that

$$\lim_{n\to\infty} p_{11}^{[n]} = 0, \qquad \lim_{n\to\infty} p_{12}^{[n]} = 0, \qquad \lim_{n\to\infty} p_{13}^{[n]} = \frac{2}{7}, \qquad \lim_{n\to\infty} p_{14}^{[n]} = \frac{1}{7}, \qquad \lim_{n\to\infty} p_{15}^{[n]} = \frac{4}{7}.$$

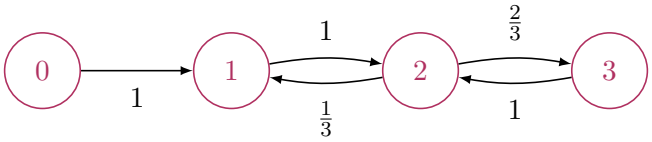
■

In Example 1.5, if the chain ever reaches one of the states 3, 4, or 5, then it may never again visit states 1 or 2. Hence some states in S are not always (i.e. at all times) accessible by the chain with positive probability. In this section we shall identify and classify states based on the extent of their accessibility from other states.

2 RECURRENCE AND TRANSIENCE

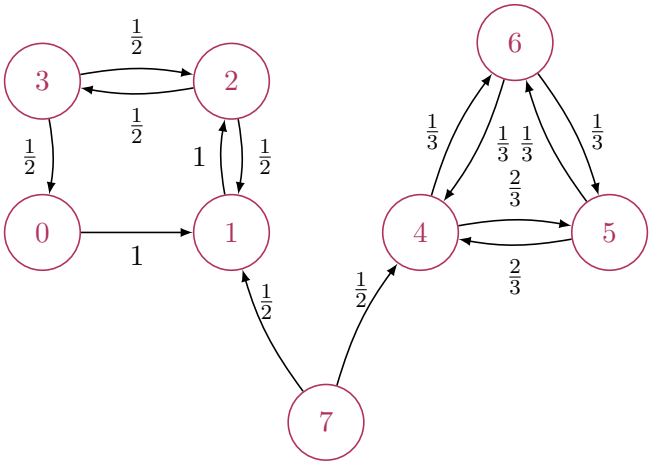
A quantitative way to understand long time behaviour of chain is to decide how many times does it visit a site. We consider a few situations in the next example.

EXAMPLE 2.1. First we consider, X_n, to be a Markov chain on $S = \{0, 1, 2, 3\}$ with the dynamics desribed via the graph below.



It is immediate that the number of vists to $\{0\}$ (counting from time $n = 0$) is either 1 (when the chain starts at state 0) or0 (since if the chain does not begin at state 0 is can never visit that state in the future. Consequently the set $\{1, 2, 3\}$ will be visited infinitely often by the chain as $n \to \infty$. Though it may not be as apparent, it is a fact that since $1 \leftrightarrow 2 \leftrightarrow 3$, the chain will visit $\{1\}, \{2\}$, and $\{3\}$ infinitely often.

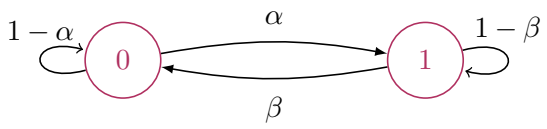
Let us now consider another Markov chain on Y_n on $T = \{0, 1, 2, 3, 4, 5, 6, 7\}$ whose dynamics is given by the graph below.



It is immediate that the number of vists to $\{7\}$ (counting from time $n = 0$) is either 1 (when the chain starts at state 7) or 0 (if the chain stats in any other state). Consequently the

sets $\{0,1,2,3\}$ or $\{4,5,6\}$ will be visited infinitely often by the chain as $n \to \infty$. Further, a little bit of thought will yield that for some starting points only one of them is visited infintely often and the other set is never visited.

Our final example is the Markov chain from Example 1.4 given by dynamics below



If $0 < \alpha, \beta < 1$ then one can show that both 0 and 1 are visited infinitely often by the chain as $n \to \infty$. There are no states in the chain that are visited finitely many times. ■

We shall now formalise the observations from the above example and proceed to establish a dichotomy property for a state in terms of the number of visits to the state by the chain. Let X_n be a Markov chain on S with transition matrix $\mathbf{P}$ and initial distribution μ. . Define number of visits by time n to the state $i \in S$ by

$$N_n(i) = \sum_{m=0}^{n} 1_{\{X_m = i\}} \tag{2.14}$$

and the total number of visits to the state $i \in S$ by the chain as

$$N(i) = \sum_{m=0}^{\infty} 1_{\{X_m = i\}}. \tag{2.15}$$

A state $i \in S$ is called recurrent if, with probability one, the total number of visits to state i is infinite and is called transient if, with probability one, the total number of visits to state i is finite.

DEFINITION 2.2. *A state $i \in S$ is said to be*

1. *recurrent if $\mathbb{P}(N(i) = \infty \mid X_0 = i) = 1$.*
2. *transient if $\mathbb{P}(N(i) < \infty \mid X_0 = i) = 1$.*

The Markov chain X_n is said to be recurrent (resp. transient) if every state is recurrent (resp. transient).

The dichotomy referred to earlier is the fact that a state must be be either recurrent or transient. To establish this fact we will need to introduce the notion of return times. For any $A \subset S$, the return time to A is defined by

$$T_A = \inf\{n \geq 1 : X_n \in A\}. \tag{2.16}$$

Note that if $A = \{j\}$ is a single-state set, then $T_{\{j\}} = T_j$ is simply the hitting time of state $\{j\}$.

Suppose a chain starts at state $i \in S$ and visits the i, m times. Then the time length of each of the m excursions between each visit are i.i.d. under P_i by the Markov Property. Thus, the probability that the chain makes at least m visits to a site $i \in S$ is the same as the probabilty that the return time to site i is finite for m independent copies of the chain. This notion is made rigorous in the next lemma where we derive a relationship between $N(i)$ and T_i for $i \in S$.

LEMMA 2.3. *Let X_n be a Markov chain on S with transition matrix $\mathbf{P}$ and initial distribution μ. Let $i \in S$. Then for all $m \geq 1$*

$$\mathbb{P}(N(i) \geq m \mid X_0 = i) = [\mathbb{P}(T_i < \infty \mid X_0 = i)]^m \tag{2.17}$$

Proof : Let $m \in \mathbb{N}$. First note that

$$\{N(i) \geq m\} = \{N(i) \geq m, T_i < \infty\} = \bigcup_{k=1}^{\infty} \{N(i) \geq m, T_i = k\}.$$

For $k \geq 1$, if $\mathbb{P}(T_i = k \mid X_0 = i) > 0$ then using definition of conditional probability and $N(i)$ we have

$$\begin{aligned}
&\mathbb{P}(N(i) \geq m, T_i = k \mid X_0 = i) && (2.18)\\
&= \mathbb{P}(N(i) \geq m \mid T_i = k, X_0 = i)\mathbb{P}(T_i = k \mid X_0 = i)\\
&= \mathbb{P}(N(i) \geq m | X_k = i, X_{k-1} \neq i, \ldots, X_1 \neq i, X_0 = i)\mathbb{P}(T_i = k \mid X_0 = i)\\
&= \mathbb{P}(N(i) \geq m - 1)\mathbb{P}(T_i = k \mid X_0 = i). && (2.19)
\end{aligned}$$

The last equality is a consequence of the Markov Property (using definition of $N(i)$). Intuitively, the memoryless property of the Markov chain implies that the proability of having at least m visits to the state $i \in S$ given that there is already one visit to state i by time k is the same as probability of starting at i and having at least $m-1$ visits to state $i \in S$.

For $\mathbb{P}(T_i = k \mid X_0 = i) = 0$ then $\mathbb{P}(N(i) \geq m, T_i = k \mid X_0 = i) = 0$. So, even in this case trivially,

$$\mathbb{P}(N(i) \geq m, T_i = k \mid X_0 = i) = \mathbb{P}(N(i) \geq m-1 \mid X_0 = i)\mathbb{P}(T_i = k \mid X_0 = i).$$

Using this and (2.18) we have,

$$\begin{aligned}
\mathbb{P}(N(i) \geq m \mid X_0 = i) &= \sum_{k=1}^{\infty} \mathbb{P}(N(i) \geq m, T_i = k \mid X_0 = i) && (2.20)\\
&= \sum_{k=1}^{\infty} \mathbb{P}(N(i) \geq m-1 \mid X_0 = i)\mathbb{P}(T_i = k \mid X_0 = i) && (2.21)\\
&= \mathbb{P}(N(i) \geq m-1 \mid X_0 = i) \sum_{k=1}^{\infty} \mathbb{P}(T_i = k \mid X_0 = i) && (2.22)\\
&= \mathbb{P}(N(i) \geq m-1 \mid X_0 = i)\mathbb{P}(T_i < \infty \mid X_0 = i) && (2.23)
\end{aligned}$$

An inductive argument then shows that, for all $m \geq 1$,

$$\mathbb{P}(N(i) \geq m \mid X_0 = i) = [\mathbb{P}(T_i < \infty \mid X_0 = i)]^m$$

■

We are now ready to state and prove the dichotomy result mentioned at the start of this section.

PROPOSITION 2.4. *Let X_n be a Markov chain on S with transition matrix $\mathbf{P}$ and initial distribution μ.*

(a) A state $i \in S$ is recurrent if and only if $\mathbb{P}(T_i < \infty \mid X_0 = i) = 1$.

(b) A state $i \in S$ is transient if and only if $\mathbb{P}(T_i < \infty \mid X_0 = i) < 1$.

In particular a state $i \in S$ is either transient or recurrent.

Proof: The proof follows readily from Lemma 2.3. The event

$$\{N(i) = \infty\} = \bigcap_{m=1}^{\infty} \{N(i) \geq m\}.$$

Since the right hand side above is an intersection of decreasing sequence of events.

$$\mathbb{P}(N(i) = \infty \mid X_0 = i) = \lim_{m \to \infty} \mathbb{P}(\{N(i) \geq m\} \mid X_0 = i)$$

Furtehr note that,

$$\lim_{m\to\infty}[\mathbb{P}(T_i < \infty \mid X_0 = i)]^m = \begin{cases} 1 & \text{if } P(T_i < \infty \mid X_0 = i) = 1 \\ 0 & \text{if } P(T_i < \infty \mid X_0 = i) < 1 \end{cases}$$

Now using (2.17) from Lemma 2.3 and the above we have Since the right hand side above is an intersection of decreasing sequence of events, using (2.17) we have

$$\mathbb{P}(N(i) = \infty \mid X_0 = i) = \begin{cases} 1 & \text{if } \mathbb{P}(T_i < \infty \mid X_0 = i) = 1 \\ 0 & \text{if } \mathbb{P}(T_i < \infty \mid X_0 = i) < 1 \end{cases} \tag{2.24}$$

By definition of recurrence and transience and (2.24) we have the result. ∎

REMARK 2.5. (*Criteria for Recurrence and Transience*) As a consequence of the preceeding proposition, one way to decide whether a state $i \in S$ is recurrent or not is to check if $P_i(T_i < \infty) = 1$. That is, if a chain beginning in a state is guaranteed to return to that state (at some later point in time) the state is recurrent. If there is some chance the chain will not return to that state, the state is transient. Checking this condition may sometimes prove difficult so we shall next try to propose another equivalent criteria.

Claim: $i \in S$ is recurrent if and only if

$$\sum_{n=0}^{\infty} p_{ii}^{[n]} = \infty. \tag{2.25}$$

Proof of Claim: First we observe that a state $i \in S$ is transient is equivalent to $\mathbb{E}[N(i) \mid X_0 = i] < \infty$, since

$$\begin{aligned} \mathbb{P}(N(i) < \infty \mid X_0 = i) = 1 &\iff \mathbb{P}(T_i < \infty \mid X_0 = i) < 1 \\ &\iff \sum_{m=1}^{\infty} [\mathbb{P}(T_i < \infty \mid X_0 = i)]^m < \infty \\ &\iff \sum_{m=1}^{\infty} \mathbb{P}(N(i) \geq m \mid X_0 = i) < \infty \qquad (2.26) \\ &\iff \mathbb{E}[N(i) \mid X_0 = i] < \infty. \qquad (2.27) \end{aligned}$$

Next we will invoke a result on convergence of expectations of non-negative random variables, and obtain

$$\begin{aligned}\mathbb{E}[N(i) \mid X_0 = i] &= \mathbb{E}[\sum_{n=1}^{\infty} 1(X_n = i) \mid X_0 = i] \\ &= \sum_{n=1}^{\infty} \mathbb{E}[1(X_n = i) \mid X_0 = i] \\ &= \sum_{n=1}^{\infty} \mathbb{P}(X_n = i \mid X_0 = i) \\ &= \sum_{n=1}^{\infty} p_{ii}^{[n]}. \end{aligned} \tag{2.28}$$

Thus (2.26) and (2.28) establish (2.25). ■.

We conclude with and example on the behaviour of Simple random walk on the integer lattice.

EXAMPLE 2.6. Let X_n be a Markov chain on $S = \mathbb{Z}$ with tranistion matrix $\mathbf{P}$ given by

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

where $0 < p < 1$. Suppose $P(X_0 = 0) = 1$.

We now calculate probabilities at state i for all time n. At each time step the chain moves up by one or moves down by one, consequently the chain cannot return to its starting in odd number of steps. For it to return in any even number of steps, say $2n$ steps, it must have moved up by one in n steps and moved down by one in n steps. These can happen in one of $\binom{2n}{n}$ permutations and each such sequence has probability $p^n(1-p)^n$. So, we have for any $i \in S$,

$$p_{ii}^{[2n]} = \binom{2n}{n} p^n (1-p)^n = \frac{2n!}{n!n!} p^n (1-p)^n.$$

Sterling's formula implies that there is a $c_1 > 0$ such that for all $n \geq 1$

$$\frac{1}{c_1} n^{n+\frac{1}{2}} e^{-n} \leq n! \leq c_1 n^{n+\frac{1}{2}} e^{-n}.$$

Consequently, there is a $c_2 > 0$ such that

$$\frac{1}{c_2} 4^n n^{-\frac{1}{2}} \leq \frac{2n!}{n!n!} \leq c_2 4^n n^{-\frac{1}{2}}.$$

This implies for $n \neq 1$,

$$\frac{1}{c_2} [4p(1-p)]^n n^{-\frac{1}{2}} \leq p_{ii}^{[2n]} \leq c_2 [4p(1-p)]^n n^{-\frac{1}{2}}.$$

When $p = \frac{1}{2}$ then $p_{ii}^{[2n]} \geq \frac{1}{c_2}[4p(1-p)]^n n^{-\frac{1}{2}} = \frac{1}{c_2} n^{-\frac{1}{2}}$ and when $0 < p < 1$, $p \neq \frac{1}{2}$ then $p_{ii}^{[2n]} \leq \frac{1}{c_2} \alpha^n n^{-\frac{1}{2}}$ with $0 < \alpha = 4p(1-p) < 1$. So it is easy to see from common facts about infinite sums that

$$\sum_{k=0}^{\infty} p_{ii}^{[k]} < \infty \iff p \neq \frac{1}{2}.$$

Hence by (2.25), $i \in S$ is a recurrent state if and only if $p = \frac{1}{2}$ or in other words, the Markov chain X_n is recurrent if $p = \frac{1}{2}$ and transient otherwise. ■

BIBLIOGRAPHY

[AS09] Siva, Athreya, Sunder, V.S. *Measure and Probability* CRC Press (Outside India) 2009 ISBN 14 3980 126 6.

[CasBer90] Casella, George; Berger, Roger L.*Statistical inference.* The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990. xviii+650 pp. ISBN: 0-534-11958-1

[Fel68] W. Feller, *An introduction to probability theory and its applications* Vol. I. Third edition John Wiley & Sons, Inc., New York-London-Sydney 1968.

[Fel71] W. Feller, *An introduction to probability theory and its applications* Vol. II. Second edition John Wiley & Sons, Inc., New York-London-Sydney 1971.

[FPP98] T.Freedman, R.Pisani, R.Purves, *Statistics* Third edition W.W.Norton and Company, Inc., New York-London, 1998.

[FG97] B. Fristedt and L. Gray, *A Modern Approach to Probability Thoery* Birkhauser, Boston 1997.

[FW] M.i. Friedlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer-verlag, New York, 1998.

[Ghah00] S. Ghahramani *Fundamentals of probability.* Second edition Prentice Hall, New Jersey, 2000.

[HMC05] R.Hogg, J. McKean, A. Craig, *Introduction to Mathematical Statistics*, Sixth Edition, Pearson Education, Inc., New Jersey, 2005.

[HPS72] P.G. Hoel, S.C. Port, and C.J. Stone, *Introduction to Probability Theory*, Houghton Mifflin Company, 1972.

[Keane] M. Keane, *The Essence of the Law of Large Numbers*, , Pages 125–129, Algorithms, Fractals, and Dynamics, Springer USA, 1995.

[Nor97] J.R. Norris, *Markov Chains,* Cambridge University Press, 1997.

[Pit89] J. Pitman, *Probability,* Springer-verlag, New York, 1989.

[Ram97] Ramasubramanian, *The Normal Distribution 1. From Binomial to Normal* Resonance, Vol 2, No.6, (June 1997), pp 15-14.

[Rao73] C.R.Rao, *Linear Statistical Inference and its Application* , Second Edition, John Wiley, New York, 1973.

[Ross84] S. Ross *A first course in probability*, Second edition Macmillan Publishing Company, New York, 1984.

[Stig84] Stephen M. Stigler, *Kruskal's Proof of the Joint Distribution of* $\overline{X}_n$ *and* s_n^2, The American Statistician, Vol 38, No. 2, (May 1984), pp 134-135.

[LPW] Levin DA, Peres Y. and Wilmer EL, *Markov chains and mixing times.* American Mathematical Soc., Providence. 2009;107.