DMS625: Introduction to stochastic processes and their applications

Sourav Majumdar

Markov Chains

September 4, 2024

1 Stochastic Processes

A sequence of random variables $X_i, i \in J$ is said to be a stochastic process, where J is some indexing set. For e.g., if $J = \mathbb{N}$, then the sequence of random variables is $X_1, X_2, X_3, \ldots$.

Exercises

1. Suppose $X_1, X_2, \ldots, X_n$ are iid random variables. Is $X_i, i = 1, 2, \ldots, n$ a stochastic process?

The set J could be discrete, or continuous. Usually J is interpreted to be time. However J can be other kinds of sets. The other most common choice of J is space. So $X_i, i \in J$, say $J \subseteq \mathbb{R}^2$, could be random variables distributed over some spatial domain. For, e.g., the distribution of temperature across geography. These are called spatial processes. Another possible choice for J is a cartesian product of multiple sets. The most commonly seen one is space-time processes, i.e., $J = \mathbb{R}^2 \times [0, t)$. These are called spatio-temporal processes. Extending the earlier example, spatio-temporal processes as an example could study variations in temperature across geography through time.

We refer to the values that X_i takes as the **states** of the stochastic process. The states could be discrete or continuous, countable or uncountable. We shall explore some of these in the course.

2 Markov Chains

Let $\mathcal{S}$ be the set of states. In this course, for Markov Chains, we will assume that $\mathcal{S}$ is discrete. Consider a sequence of random variables $X_n, n \geq 0, n \in \mathbb{N}$. We say that these random variables satisfy the **Markov Property** iff,

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n) \tag{1}$$

$\forall n$ and $x_0, \ldots, x_{n+1} \in \mathcal{S}$. This in words means that the probability of moving to a certain state, only depends on the immediate past state and not the entire history.

A sequence of random variables X_n having the Markov property is said to be a **Markov Chain**. $\mathbb{P}(X_{n+1} = y | X_n = x)$ is said to be the **transition probability** of the Markov Chain.

We say a Markov Chain is **homogeneous** (or **time-homogeneous**, if the indices are some subset of time) if $\mathbb{P}(X_{n+1} = y | X_n = x)$ is independent of n. This implies that,

$$\mathbb{P}(X_1 = y | X_0 = x) = \mathbb{P}(X_2 = y | X_1 = x) = \ldots = \mathbb{P}(X_{n+1} = y | X_n = x) = \ldots$$

This means that the probability of going from state x to y is independent of the time at which it is occurring.

Remark 1. *Homogeneous Markov chains are also alternatively referred to as Markov Chains with* ***stationary transition probabilities****. Note that the notion of a stationary distribution, which shall be introduced later, is different from that of stationary transition probability. It is possible for a Markov Chain to have a stationary transition probabilities but not possess a stationary distribution. These concepts shouldn't be confused.*

Remark 2. *In all subsequent mentions in these notes, unless stated otherwise, it will be assumed that the Markov Chains are time-indexed and that they are time-homogeneous.*

For a Markov chain $X_n, n \geq 0$, we call the function $\mathbf{P}(x, y) = \mathbb{P}(X_1 = y | X_0 = x), x, y \in \mathcal{S}$, the **transition function**. The transition function of a Markov chain satisfies the following properties,

1. $\mathbf{P}(x, y) \geq 0, x, y \in \mathcal{S}$, follows from the fact that probabilities are non-negative.
2. $\sum_y \mathbf{P}(x, y) = 1, \forall x \in \mathcal{S}$ (Prove!)

Exercises

1. Show that $\mathbf{P}(x, y) = \mathbb{P}(X_n = y | X_{n-1} = x), \forall n \in \mathbb{N}$.
2. Show that iid X_i form a Markov Chain.
3. Is it necessarily true that $\mathbf{P}(x, y) = \mathbf{P}(y, x)$?

At $t = 0$, the state of the Markov Chain is given by the **initial distribution**, $\pi_0(x)$. It follows that,

1. $\pi_0(x) \geq 0, x \in \mathcal{S}$, follows from the fact that probabilities are non-negative.
2. $\sum_x \pi_0(x) = 1$, sum of probabilities over all states is 1.

Suppose $\mathcal{S}$ is finite with $d + 1$ states, then we can represent the transition probabilities in the form of a matrix, which is referred to as the **transition matrix**. This is given by,

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ \vdots \\ d \end{array} \begin{array}{c} \begin{array}{ccc} 0 & \cdots & d \end{array} \\ \begin{bmatrix} \mathbf{P}(0,0) & \cdots & \mathbf{P}(0,d) \\ \vdots & \ddots & \vdots \\ \mathbf{P}(d,0) & \cdots & \mathbf{P}(d,d) \end{bmatrix} \end{array}$$

Similarly the corresponding initial distribution could also be represented as a vector,

$$\pi_0 = (\pi_0(0), \ldots, \pi_0(d))$$

Example 2.1 (Humidity in Kanpur)**.** *Let the two states be* $\{0, 1\}$*, with 0 denoting a non-humid day and 1 denoting a humid day. Suppose we could model Humidity in Kanpur as a Markov Chain, and the corresponding transition matrix is,*

$$\begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} \end{array}$$

The way to interpret this is as following. The first entry of 0.9 *states that if today is a non-humid day, the probability that tomorrow is also non-humid is 0.9. Likewise, interpret the rest of the entries of the matrix.*

A question of interest could be, in the long-term what proportion of days in Kanpur are humid days?

Example 2.2 (Simple Random walk)**.** *Let $\mathcal{S} = \mathbb{Z}$, suppose,*

$$S_i = \sum_{j=1}^{i} X_j$$

where $\mathbb{P}(X_j = 1) = p$ and $\mathbb{P}(X_j = -1) = 1 - p$, and X_j's are iid. Verify that S_i is a Markov Chain. If $p = 1/2$, it is said to be a simple symmetric random walk.

What would be the limiting behavior of a simple random walk, when the time between each jump is small?

Example 2.3 (Brand Preference)**.** *Suppose there are three toothpaste brands in the market. Customers either switch or buy the same brand again. Assume that this behavior could be modelled as a Markov Chain,*

$$\begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.3 & 0.3 & 0.4 \end{bmatrix} \end{array}$$

Would the market shares of the three brands stabilize, or would it keep switching? If it would what would be their resulting proportions?

Example 2.4 (Disability insurance)**.** *The disability insurer provides insurance to their client, when they are disabled and collects premium from them when they are healthy. There are three primary states of interest to the disability insurer: Healthy, Disabled, and Deceased. Suppose the movement of their clients follow a Markov chain. Can you identify some entries of the transition matrix without any additional information?*

Example 2.5 (Creating a Markov Chain)**.** *Suppose the rain today depends on whether or not it rained through the last two days. If it has rained for the last two days, then it will rain again tomorrow with the probability 0.7; if it rained today but not yesterday, then it will rain tomorrow with probability 0.5; if it rained yesterday but not today, then it will rain tomorrow with probability 0.4; if it has not rained in the past two days, then it will rain tomorrow with probability 0.2. Let $X_n = 1$ denote the event that there is rain on n-th day.*

$$\mathbb{P}(X_n | X_{n-1}, X_{n-2}, \ldots, X_0) = \mathbb{P}(X_n | X_{n-1}, X_{n-2})$$

We immediately notice that this is not a Markov Chain. However, we can create a Markov chain by defining it alternatively. We define a new random variable $Y_n \equiv (X_n, X_{n-1})$. Now see that Y_n is a Markov chain,

$$\mathbb{P}(Y_n | Y_{n-1}, Y_{n-2}, \ldots, Y_1) = \mathbb{P}(X_n, X_{n-1} | X_{n-1}, X_{n-2}, \ldots, X_0) = \mathbb{P}(X_n, X_{n-1} | X_{n-1}, X_{n-2}) = \mathbb{P}(Y_n | Y_{n-1})$$

The four states of this new Markov Chain are,

1. *Rained both today and yesterday (RR)*
2. *Rained today but not yesterday (RN)*

3. *Rained yesterday but not today (NR)*
4. *Didn't rain either yesterday or today (NN)*

$$\begin{array}{c} \\ RR \\ RN \\ NR \\ NN \end{array} \begin{array}{c} \begin{array}{cccc} RR & RN & NR & NN \end{array} \\ \begin{bmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{bmatrix} \end{array}$$

Proposition 2.1. $\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = \pi_0(x_0)\mathbf{P}(x_0, x_1)\ldots\mathbf{P}(x_{n-1}, x_n)$

Proof. Note that by the definition of conditional probability,

$$\begin{aligned} \mathbb{P}(X_0 = x_0, X_1 = x_1) &= \mathbb{P}(X_1 = x_1 | X_0 = x_0)\mathbb{P}(X_0 = x_0) \\ &= \mathbf{P}(x_0, x_1)\pi(x_0) \end{aligned}$$

given $\mathbb{P}(X_0 = x_0) > 0$, likewise, it follows that,

$$\begin{aligned} \mathbb{P}(X_0 = x_0, X_1 = x_1, X_2 = x_2) &= \mathbb{P}(X_2 = x_2 | X_1 = x_1, X_0 = x_0)\mathbb{P}(X_1 = x_1 | X_0 = x_0)\mathbb{P}(X_0 = x_0) \\ &= \mathbb{P}(X_2 = x_2 | X_1 = 1)\mathbb{P}(X_1 = x_1 | X_0 = x_0)\mathbb{P}(X_0 = x_0) \\ &\text{(Markov Property)} \\ &= \mathbf{P}(x_1, x_2)\mathbf{P}(x_0, x_1)\pi(x_0) \end{aligned}$$

given $\mathbb{P}(X_0 = x_0) > 0$ and $\mathbb{P}(X_0 = x_0, X_1 = x_1) > 0$.

We shall apply the principle of mathematical induction. Assume that for some $m < n$,

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_{m-1} = x_{m-1}) = \prod_{j=1}^{m-1} \mathbf{P}(x_{j-1}, x_j)\pi(x_0) \tag{2}$$

given $\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_k = x_k) > 0, \forall k = 0, 1, 2, \ldots, m-1$.

It follows from (2) that,

$$\begin{aligned} \mathbb{P}(X_0 = x_0, \ldots, X_m = x_m) =& \mathbb{P}(X_m = x_m | X_{m-1} = x_{m-1}, \ldots, X_0 = x_0) \\ & \mathbb{P}(X_{m-1} = x_{m-1}, \ldots, X_0 = x_0) \\ =& \mathbf{P}(x_{m-1}, x_m) \prod_{j=1}^{m-1} \mathbf{P}(x_{j-1}, x_j)\pi(x_0) \\ =& \pi(x_0)\mathbf{P}(x_0, x_1)\ldots\mathbf{P}(x_{m-1}, x_m) \end{aligned}$$

given $\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_k = x_k) > 0, \forall k = 0, 1, 2, \ldots, m$. Therefore, by the principle of mathematical induction,

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = \pi_0(x_0)\mathbf{P}(x_0, x_1)\ldots\mathbf{P}(x_{n-1}, x_n) \tag{3}$$

given,

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_k = x_k) > 0, \forall k = 0, 1, 2, \ldots, n. \tag{4}$$

However (4) could fail to hold for some k. Suppose we denote the smallest k for which it fails to hold by k^*. Formally this is expressed as,

$$k^* = \inf\{k \geq 0 : \mathbb{P}(X_0 = x_0, \ldots, X_k = x_k) = 0\}$$

Then from (3) it follows that,

$$\mathbb{P}(X_0 = x_0, \dots, X_{k^*-1} = x_{k^*-1}) = \pi(x_0)\mathbf{P}(x_0, x_1)\dots\mathbf{P}(x_{k^*-2}, x_{k^*-1}) > 0$$

and,

$$\begin{aligned}
\mathbf{P}(x_{k^*-1}, x_{k^*}) =& \mathbb{P}(X_{k^*} = x_k^* | X_{k^*-1} = x_{k^*-1}) \\
=& \mathbb{P}(X_{k^*} = x_k^* | X_{k^*-1} = x_{k^*-1}, X_{k^*-2} = x_{k^*-2}, \dots, X_0 = x_0) \\
& \text{(Markov Property)} \\
=& \frac{\mathbb{P}(X_{k^*} = x_{k^*}, X_{k^*-1} = x_{k^*-1}, \dots, X_0 = x_0)}{\mathbb{P}(X_{k^*-1} = x_{k^*-1}, X_{k^*-2} = x_{k^*-2}, \dots, X_0 = x_0)} = 0 \\
& \text{(Numerator is 0, by the definition of } k^*)
\end{aligned}$$

We note that the claim of the theorem still holds, because if $\mathbb{P}(X_0 = x_0, \dots, X_{k^*} = x_k^*) = 0$, then $\mathbb{P}(X_0 = x_0, \dots, X_k^* = x_k^*, \dots, X_n = x_n) = 0$ and $\mathbf{P}(x_{k^*-1}, x_{k^*})$ is a factor on the RHS, setting the RHS as well to 0. ☐

Higher-order transitions

$\mathbf{P}(x, y)$ gives the probability of a Markov Chain going from state x to y in one time-step. We may be interested in knowing the probability of going from x to y in $m \geq 1$ time-steps. In other words, if at present it is at x, what is the probability of it being in y after m time steps? The knowledge of $\mathbf{P}$ and π is sufficient to do these calculations for a Markov Chain.

Proposition 2.2. $\mathbb{P}(X_{n+1} = x_{n+1}, \dots, X_{n+m} = x_{n+m} | X_0 = x_0, \dots, X_n = x_n) = \mathbf{P}(x_n, x_{n+1})\dots\mathbf{P}(x_{n+m-1}, x_{n+m})$

Proof. Left as an exercise. ☐

Suppose we are interested in the probability that the chain is at a particular state after two-time steps, this probability is called **2-step transition probability**. Let us assume that the chain is at time step n and we are interested in the probability of an event at time step $n + 2$, i.e.

$$\begin{aligned}
\mathbb{P}(X_{n+2} = y | X_n = x) =& \mathbb{P}(X_{n+2} = y, X_{n+1} \in \mathcal{S} | X_n = x) \\
& \text{(This denotes that at } X_{n+1} \text{ the chain could be at any possible state)} \\
=& \sum_{s \in \mathcal{S}} \mathbb{P}(X_{n+2} = y, X_{n+1} = s | X_n = x) \\
& \text{(Since the states of the chain are disjoint,} \\
& \text{the probability of union of all states } \mathcal{S} \text{ can be decomposed into their sum)} \\
=& \sum_{s \in \mathcal{S}} \mathbf{P}(x, s)\mathbf{P}(s, y) \\
& \text{(From Proposition (2.2))}
\end{aligned}$$

We will denote this probability as $\mathbf{P}^{(2)}(x, y) = \mathbb{P}(X_{n+2} = y | X_n = x) = \sum_{s \in \mathcal{S}} \mathbf{P}(x, s)\mathbf{P}(s, y)$. We note that $\mathbf{P}^{(2)}(x, y)$ is independent of n, therefore the expression that we calculated holds for any n.

In a similar fashion suppose we are interested in the **m-step transition probability**. We shall denote it by $\mathbf{P}^{(m)}(x, y) = \mathbb{P}(X_{n+m} = y | X_n = x)$.

$$
\begin{aligned}
&\mathbf{P}^{(m)}(x,y)\\
=&\mathbb{P}(X_{n+m}=y|X_n=x)\\
=&\mathbb{P}(X_{n+m}=y, X_{n+m-1}\in\mathcal{S}, X_{n+m-2}\in\mathcal{S},\dots,X_{n+1}\in\mathcal{S}|X_n=x)\\
=&\sum_{s_1\in\mathcal{S}}\sum_{s_2\in\mathcal{S}}\dots\sum_{s_{m-1}\in\mathcal{S}}\mathbb{P}(X_{n+m}=y, X_{n+m-1}=s_{m-1}, X_{n+m-2}=s_{m-2},\dots,X_{n+1}=s_1|X_n=x)\\
=&\sum_{s_1\in\mathcal{S}}\sum_{s_2\in\mathcal{S}}\dots\sum_{s_{m-1}\in\mathcal{S}}\mathbf{P}(x,s_1)\mathbf{P}(s_1,s_2)\dots\mathbf{P}(s_{m-1},y)
\end{aligned}
$$

Again note that the m-step transition probability is also independent of n.

Remark 3. *Note that $\mathbf{P}^{(m)}(x,y)\neq(\mathbf{P}(x,y))^m$. We shall see a simple way of calculating this through the Chapman-Kolmogorov Equation.*

Exercises

1. Show that $\mathbb{P}(X_{n+m}=y|X_0=x, X_n=z)=\mathbf{P}^{(m)}(z,y)$

Example 2.6 (Gambler's Ruin)**.** *A gambler at a Casino is playing each round where she can win 1 unit of wealth with probability 0.4 and lose 1 unit of wealth with probability 0.6. She stops playing once she has reached 4 units of wealth and once she reaches 0 units of wealth, she again stops playing, as she is ruined. Let X_n denote the wealth of the gambler after n rounds, assume that X_n is a Markov chain, then the transition probabilities are,*

$$
\begin{array}{c c}
 & \begin{array}{ccccc} \mathit{0} & \mathit{1} & \mathit{2} & \mathit{3} & \mathit{4} \end{array} \\
\begin{array}{c} \mathit{0} \\ \mathit{1} \\ \mathit{2} \\ \mathit{3} \\ \mathit{4} \end{array} &
\begin{bmatrix}
1 & 0 & 0 & 0 & 0\\
0.6 & 0 & 0.4 & 0 & 0\\
0 & 0.6 & 0 & 0.4 & 0\\
0 & 0 & 0.6 & 0 & 0.4\\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
$$

Notice that $\{1\}$ and $\{4\}$ are absorbing states as $\mathbf{P}(1,1)=\mathbf{P}(4,4)=1$.

Exercises

1. For the gambler's ruin problem compute the following $\mathbb{P}(X_0=1, X_1=2, X_2=3, X_3=2, X_4=3)$.

2. For the gambler's ruin problem compute $\mathbb{P}(X_5=4|X_0=1)$. What is $\mathbb{P}(X_{10}=4|X_5=1)$?

Chapman-Kolmogorov Equation

Theorem 2.1. $\mathbf{P}^{(n+m)}(x,y)=\sum_z\mathbf{P}^{(n)}(x,z)\mathbf{P}^{(m)}(z,y)$

Proof.

$$\begin{aligned}\mathbf{P}^{(n+m)}(x,y) &= \mathbb{P}(X_{n+m}=y|X_0=x)\\ &= \sum_z \mathbb{P}(X_n=z|X_0=x)\mathbb{P}(X_{n+m}=y|X_0=x,X_n=z)\\ &= \sum_z \mathbf{P}^{(n)}(x,z)\mathbb{P}(X_{n+m}=y|X_0=x,X_n=z)\\ &= \sum_z \mathbf{P}^{(n)}(x,z)\mathbf{P}^{(m)}(z,y)\end{aligned}$$

□

Note a consequence of the Chapman-Kolmogorov equation, upon setting $n=m=1$, we obtain,

$$\mathbf{P}^{(2)}(x,y) = \sum_z \mathbf{P}(x,z)\mathbf{P}(z,y)$$

This is nothing but the (x,y)-th entry of the square of the transition matrix $\mathbf{P}$. Extending the logic, we get that the probability of going from x to y in m-steps is the (x,y)-th entry in the m-th power of the transition matrix, $\mathbf{P}$.

Proposition 2.3. $\mathbb{P}(X_n=y) = \sum_x \pi_0(x)\mathbf{P}^{(n)}(x,y)$

Proof.

$$\mathbb{P}(X_n=y) = \sum_x \mathbb{P}(X_0=x,X_n=y) = \sum_x \mathbb{P}(X_0=x)\mathbb{P}(X_n=y|X_0=x) = \sum_x \pi_0(x)\mathbf{P}^{(n)}(x,y)$$

□

To calculate $\mathbf{P}^{(3)}(1,2)$, i.e., starting at $\{1\}$ and reaching $\{2\}$ after 3 steps in the gambler's ruin problem, we first evaluate $\mathbf{P}^{(3)} = \mathbf{P}\times\mathbf{P}\times\mathbf{P}$,

$$\begin{aligned}\mathbf{P}^{(3)} &= \mathbf{P}\times\mathbf{P}\times\mathbf{P}\\ &= \begin{array}{c} \\ 0\\1\\2\\3\\4\end{array}\begin{array}{c} \begin{array}{ccccc} 0 & \quad 1 & \quad 2 & \quad 3 & \quad 4\end{array}\\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0\\ 0.714 & 0 & 0.192 & 0 & 0.064\\ 0.360 & 0.288 & 0 & 0.192 & 0.160\\ 0.216 & 0 & 0.288 & 0 & 0.496\\ 0 & 0 & 0 & 0 & 1\end{bmatrix}\end{array}\end{aligned}$$

Therefore, $\mathbf{P}^{(3)}(1,2) = 0.192$.

Hitting times

Note that the m-step transition probability gives us the probability of going from x to y in m-steps. However, it is possible that the Markov chain already visits y once or more before the m-th step. Suppose we are interested in understanding when does the Markov chain visit y for the first time, for this we will study the notion of hitting times.

Let $A \subseteq \mathcal{S}$. The **hitting time** T_A of A is defined by,

$$T_A = \min(n > 0 : X_n \in A)$$

if $X_n \in A$ for some $n > 0$, and by $T_A = \infty$ if $X_n \notin A, \forall n > 0$.

We will define a new notation, $\mathbb{P}_x(A) = \mathbb{P}(A|X_0 = x)$, for some event A, to denote that the Markov chain starts at x, i.e. $X_0 = x$.

To make the definition more explicit, consider $T_y = 5$, i.e., the event where chain reaches (hits) the state y for the first time at the 5-th step, so,

$$\mathbb{P}_x(T_y = 5) = \mathbb{P}_x(X_1 \neq y, X_2 \neq y, X_3 \neq y, X_4 \neq y, X_5 = y)$$

Hitting times play a prominent role in both theoretical and applied analysis of Markov chains. We will explore the application of the notion to classification of states subsequently. In finance, continuous state space, continuous time generalizations of the Markov chain, which are called Markov processes, are used to study stock prices. Notion of hitting times play an important role there. Particularly, the question of when a stock price would hit a particular value for the first time can be studied through hitting times. The answer to this question has implications in drawdown calculations, risk management, trading, etc.

Proposition 2.4. $\mathbf{P}^{(n)}(x,y) = \sum_{m=1}^{n} \mathbb{P}_x(T_y = m)\mathbf{P}^{(n-m)}(y,y)$

Proof. The events $\{T_y = m, X_n = y\}, 1 \leq m \leq n$ are disjoint. To see this observe that for some $1 \leq k < l < n$,

$$\{T_y = k, X_n = y\} = \{X_0 \neq y, X_1 \neq y, \ldots, X_{k-1} \neq y, X_k = y, X_n = y\}$$

and

$$\{T_y = l, X_n = y\} = \{X_0 \neq y, X_1 \neq y, \ldots, X_k \neq y, \ldots, X_{l-1} \neq y, X_l = y, X_n = y\}$$

and therefore $\{T_y = k, X_n = y\}$ and $\{T_y = l, X_n = y\}$ are disjoint. Therefore, $\{T_y = m, X_n = y\}, 1 \leq m \leq n$ are disjoint.

It also follows that,

$$\{X_n = y\} = \cup_{m=1}^{n}\{T_y = m, X_n = y\}$$

$$\begin{aligned}
\mathbf{P}^{(n)}(x,y) &= \mathbb{P}_x(X_n = y) \\
&= \sum_{m=1}^{n} \mathbb{P}_x(T_y = m, X_n = y) \text{ (2nd axiom of probability)} \\
&= \sum_{m=1}^{n} \mathbb{P}_x(T_y = m)\mathbb{P}(X_n = y|X_0 = x, T_y = m) \\
&= \sum_{m=1}^{n} \mathbb{P}_x(T_y = m)\mathbb{P}(X_n = y|X_0 = x, X_1 \neq y, \ldots, X_{m-1} \neq y, X_m = y) \\
&= \sum_{m=1}^{n} \mathbb{P}_x(T_y = m)\mathbf{P}^{(n-m)}(y,y)
\end{aligned}$$

□

Proposition 2.5.

$$\mathbb{P}_x(T_y = n+1) = \sum_{z \neq y} \mathbf{P}(x,z)\mathbb{P}_z(T_y = n)$$

Proof. Note that when the hitting time is 1, this is equivalent to the 1-step transition probability,

$$\mathbb{P}_x(T_y = 1) = \mathbb{P}_x(X_1 = y) = \mathbf{P}(x,y)$$

and,

$$\begin{aligned}
\mathbb{P}_x(T_y = 2) &= \mathbb{P}_x(X_1 \neq y, X_2 = y) \\
&= \sum_{z \neq y} \mathbb{P}_x(X_1 = z, X_2 = y) \\
&= \sum_{z \neq y} \mathbf{P}(x, z)\mathbf{P}(z, y)
\end{aligned}$$

and by induction it can be shown that,

$$\mathbb{P}_x(T_y = n + 1) = \sum_{z \neq y} \mathbf{P}(x, z)\mathbb{P}_z(T_y = n)$$

□

Thanks to Manan Kabra for an alternate proof,

Proof.

$$\begin{aligned}
\mathbb{P}_x(T_y = n + 1) &= \mathbb{P}(X_{n+1} = y, X_n \neq y, \ldots, X_1 \neq y, X_0 = x) \\
&= \mathbb{P}(X_{n+1} = y, X_n \neq y, \ldots, X_2 \neq y | X_1 \neq y, X_0 = x)\mathbb{P}(X_1 \neq y, X_0 = x) \\
&= \sum_{z \neq y} \mathbb{P}(X_{n+1} = y, X_n \neq y, \ldots, X_2 \neq y | X_1 = z, X_0 = x)\mathbb{P}(X_1 = z, X_0 = x) \\
&= \sum_{z \neq y} \mathbb{P}_z(T_y = n)\mathbf{P}(x, z)
\end{aligned}$$

□

For the gambler's ruin problem, $\mathbb{P}_1(T_3 = 1) = \mathbf{P}(1, 3) = 0$, i.e., the probability of hitting $\{3\}$ in one step from $\{1\}$. Similarly,

$$\begin{aligned}
\mathbb{P}_1(T_3 = 2) &= \mathbf{P}(1, 0)\mathbb{P}_0(T_3 = 1) + \mathbf{P}(1, 2)\mathbb{P}_2(T_3 = 1) \\
&= 0.6 \times 0 + 0.4 \times 0.4 = 0.16
\end{aligned}$$

Next,

$$\mathbb{P}_1(T_3 = 3) = \mathbf{P}(1, 0)\mathbb{P}_0(T_3 = 2) + \mathbf{P}(1, 2)\mathbb{P}_2(T_3 = 2)$$

and so on.

Transient and Recurrent States

Define $\rho_{xy} = \mathbb{P}_x(T_y < \infty)$. ρ_{xy} is the probability of a Markov Chain at x to be in y in finite time. ρ_{yy} denotes the probability that the chain starting at y will return to y. A state y is called **recurrent** if $\rho_{yy} = 1$ and **transient** if $\rho_{yy} < 1$. This implies that a chain will always return to a recurrent state in finite time, however there is a positive probability of a chain never returning to a transient state. Recall the disability insurance example, the deceased state which was an absorbing state is also a recurrent state. We will develop tools to be able to identify these states.

Example 2.7 (Absorbing states are recurrent).

$$
\begin{array}{c c}
 & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\end{array}
$$

Example 2.8 (All states are recurrent, but none are absorbing).

$$
\begin{array}{c c}
 & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}
\end{array}
$$

Example 2.9 (Transient state).

$$
\begin{array}{c c}
 & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.2 & 0.4 & 0.4 \\ 0 & 0.3 & 0.7 \\ 0 & 0.7 & 0.3 \end{bmatrix}
\end{array}
$$

Notice $\{1\}$ is a transient state, since it can leave $\{1\}$ with a positive probability and once it leaves $\{1\}$, it can never reach $\{1\}$ again. $\{2\}$ and $\{3\}$ are recurrent states, but not absorbing states.

Let $I_y(z), z \in \mathcal{S}$, denote the indicator function, where $I_y(z) = 1, z = y$ and $I_y(z) = 0, z \neq y$.
Let $N(y)$ be the number of times $n \geq 1$ that the chain visits y. We can write it as,

$$N(y) = \sum_{n=1}^{\infty} I_y(X_n)$$

where if the chain is at y at time n, $I_y(X_n) = 1$ and otherwise 0. Note that $\mathbb{P}_x(N(y) \geq 1) = \mathbb{P}_x(T_y < \infty) = \rho_{xy}$.

The probability of a chain first visiting y after starting from x in time m and then visits y again after another time n is $\mathbb{P}_x(T_y = m)\mathbb{P}_y(T_y = n)$. Therefore,

$$\begin{aligned}
\mathbb{P}_x(N(y) \geq 2) &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \mathbb{P}_x(T_y = m)\mathbb{P}_x(T_y = n) \\
&= \left(\sum_{m=1}^{\infty} \mathbb{P}_x(T_y = m)\right)\left(\sum_{n=1}^{\infty} \mathbb{P}_y(T_y = n)\right) \\
&= \rho_{xy}\rho_{yy}
\end{aligned}$$

It follows that,

$$\mathbb{P}_x(N(y) \geq m) = \rho_{xy}\rho_{yy}^{m-1} \tag{5}$$

Exercises

1. Show that $\mathbb{P}_x(N(y) = m) = \rho_{xy}\rho_{yy}^{m-1}(1 - \rho_{yy})$
2. Show that $\mathbb{P}_x(N(y) = 0) = 1 - \rho_{xy}$

We denote $\mathbb{E}_x$ to denote the expectation of an event of a Markov chain starting at x.

Exercises

1. Show that $\mathbb{E}_x(I_y(X_n)) = \mathbf{P}^{(n)}(x,y)$

Now proceeding,

$$\begin{aligned}\mathbb{E}_x(N(y)) &= \mathbb{E}_x\left(\sum_{n=1}^{\infty} I_y(X_n)\right)\\ &= \sum_{n=1}^{\infty} \mathbb{E}_x(I_y(X_n))\\ &= \sum_{n=1}^{\infty} \mathbf{P}^{(n)}(x,y)\end{aligned}$$

We set $G(x,y) = \mathbb{E}_x(N(y)) = \sum_{n=1}^{\infty} \mathbf{P}^{(n)}(x,y)$. $G(x,y)$ denotes the expected number of times the chain visits y from x.

Theorem 2.2. *Let y be a transient state. Then,*

$$\mathbb{P}_x(N(y) < \infty) = 1$$

and

$$G(x,y) = \frac{\rho_{xy}}{1-\rho_{xy}}$$

$G(x,y)$ is finite.

Proof. Since y is a transient state, $\rho_{yy} < 1$.

$$\mathbb{P}_x(N(y) = \infty) = \lim_{m\to\infty} \mathbb{P}_x(N(y) \geq m) = \lim_{m\to\infty} \rho_{xy}\rho_{yy}^{m-1} = 0$$

Therefore the first part follows.

$$\begin{aligned}G(x,y) &= \mathbb{E}_x(N(y))\\ &= \sum_{m=1}^{\infty} m\mathbb{P}_x(N(y) = m)\\ &= \sum_{m=1}^{\infty} m\rho_{xy}\rho_{yy}^{m-1}(1-\rho_{yy})\\ &= \frac{\rho_{xy}}{1-\rho_{yy}}\end{aligned}$$

For the last step you require, $\sum_{m=1}^{\infty} mt^{m-1} = \frac{1}{(1-t)^2}, t < 1$. Show it! □

Since y is a transient state, and $\sum_{n=1}^{\infty} \mathbf{P}^{(n)}(x,y) = G(x,y) < \infty$. Therefore, $\lim_{n\to\infty} \mathbf{P}^{(n)}(x,y) = 0, x \in \mathcal{S}$.

Theorem 2.3. *Let y be a recurrent state.*

1. *Then $\lim_{m\to\infty} \mathbb{P}_y(N(y) \geq m) = 1$ and $G(y,y) = \infty$.*
2. *$\lim_{m\to\infty} \mathbb{P}_x(N(y) \geq m) = \mathbb{P}_x(T_y < \infty) = \rho_{xy}$*
3. *If $\rho_{xy} = 0$, then $G(x,y) = 0$, while if $\rho_{xy} > 0$, then $G(x,y) = \infty$*

Proof. Left as an exercise. $\square$

We say that a state x **leads** to y, if $\rho_{xy} > 0$. We will denote it as $x \to y$.

Proposition 2.6. *$x \to y$ if and only if $\mathbf{P}^{(n)}(x, y) > 0$ for some n.*

Proof. Hint: Use Proposition 2.4. $\square$

Proposition 2.7. *If $x \to y$ and $y \to z$, then $x \to z$*

Proof. Hint: Use Proposition 2.6 and the Chapman-Kolmogorov equation. $\square$

Proposition 2.8. *A Markov chain having a finite state space must have atleast one recurrent state.*

Proof. Hint: Prove by contradiction. $\square$

Theorem 2.4. *Let x be a recurrent state and suppose that $x \to y$. Then y is recurrent and $\rho_{xy} = \rho_{yx} = 1$.*

Proof. Let n^* denote the time for which the probability of n^* being a first hitting time at y from x is non-zero. Recall that we saw in the Gambler's ruin example some hitting times could have zero probability. Formally,

$$n^* = \min\{n \geq 1 : \mathbb{P}_x(T_y = n) > 0\}$$

This implies that $\mathbf{P}^{(n^*)}(x, y) > 0$ and that there exists states $y_1, \ldots, y_{n^*-1}$, such that,

$$\mathbb{P}_x(X_1 = y_1, \ldots, X_{n^*-1} = x_{n^*-1}, X_{n^*} = y) = \mathbf{P}(x, y_1) \ldots \mathbf{P}(y_{n^*-1}, y) > 0$$

None of these states $y_1, \ldots, y_{n^*-1}$ equal x or y, otherwise we would contradict the definition of n^*, since then we could go to y from x in lesser than n^* steps.

Now assume that $\rho_{yx} < 1$. The probability that the chain visits $y_1, \ldots, y_{n^*-1}, y$ in the first n^* steps and never returns to x is given by,

$$\mathbf{P}(x, y_1), \ldots, \mathbf{P}(y_{n^*-1}, y)(1 - \rho_{yx})$$

If this probability becomes non-zero, it implies that x would never be visited after time n^*, which contradicts the definition of x being a recurrent state. Therefore, $\rho_{yx} = 1$. By Proposition 2.6, there is a n^{**} such that $\mathbf{P}^{(n^{**})}(y, x) > 0$, since $\rho_{yx} > 0$.

$$\begin{aligned}
\mathbf{P}^{(n^*+n^{**}+n)}(y, y) &= \mathbb{P}_y(X_{n^*+n^{**}+n_0} = y) \\
&\geq \mathbb{P}_y(X_{n^{**}} = x, X_{n^{**}+n} = x, X_{n^{**}+n+n^*} = y) \\
&\text{(Notice it is a $\geq$ because it is only a subset of the event in the previous equation)} \\
&= \mathbf{P}^{(n^{**})}(y, x)\mathbf{P}^{(n)}(x, x)\mathbf{P}^{(n^*)}(x, y)
\end{aligned}$$

We use the above in the following,

$$
\begin{aligned}
G(y,y) &= \sum_{k=1}^{\infty} \mathbf{P}^{(k)}(y,y) \\
&\geq \sum_{k=n^{**}+1+n^{*}} \mathbf{P}^{(k)}(y,y) \\
&= \sum_{k=1}^{\infty} \mathbf{P}^{n^{**}+n^{*}+k}(y,y)
\end{aligned}
$$

(From the above result)

$$\geq \mathbf{P}^{(n^{**})}(y,x)\mathbf{P}^{n^{*}}(x,y)\sum_{k=1}^{\infty} \mathbf{P}^{(k)}(x,x) \to \infty$$

(Since $\sum_{k=1}^{\infty} \mathbf{P}^{(k)}(x,x) = G(x,x)$, and x is a recurrent state. Therefore, from Theorem 2.3 it goes to infinity.)

Therefore y is recurrent and $y \to x$. It follows that $\rho_{xy} = 1$ from the first part of the theorem. □

We say a set C of states to be **closed** if no state inside of C leads to any state outside of C, i.e. $\rho_{xy} = 0, x \in C, y \notin C$. A closed set C is called **irreducible** if x leads to y for all choices of x and y in C.

Example 2.10.

$$
\begin{array}{c c}
 & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}
\end{array}
$$

Consider the set of states $C = \{1,2,3\}, C_1 = \{1\}, C_2 = \{2,3\}$. Verify that C, C_1, C_2 are closed states. Check that C_1, C_2 are also irreducible, but C is not irreducible.

Theorem 2.5. *If C is an irreducible closed set, then every state in C is either recurrent or transient.*

Proof. For some $x \in C$, where x is recurrent, since C is irreducible, therefore x leads to all states in C. Hence, by Theorem 2.4, all states are recurrent.

For some $y \in C$, where y is transient, assume that $\exists x \in C$, where x is recurrent. Since C is irreducible, $x \to y$, therefore, by Theorem 2.4, y is recurrent which leads us to a contradiction.

In the above paragraph, we used contradiction to show that there cannot be an irreducible closed set with a mix of transient and recurrent state. However it is not immediately obvious from the contradiction that an irreducible set with all transient states can exist. To show that such a case is possible, we will have to give an example that admits this property. In the Birth and Death chain example subsequently we will construct a chain like this. □

Theorem 2.6. *Let C be a finite irreducible closed set of states. Then every state is recurrent.*

Proof. Hint: Use Proposition 2.8 and Theorem 2.5. □

Proposition 2.9.

$$\rho_{xy} = \mathbf{P}(x,y) + \sum_{z \neq y} \mathbf{P}(x,z)\rho_{zy}$$

Proof. Hint: Use Proposition 2.4. □

Example 2.11.

$$\begin{array}{c|cccccc} & 0 & 1 & 2 & 3 & 4 & 5 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ 2 & 0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & 0 & \frac{1}{5} \\ 3 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\ 4 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 5 & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \end{array}$$

Verify the following: $C = \{0,1,2,3,4,5\}$ forms a set of closed states. $C_1 = \{0\}$ and $C_2 = \{3,4,5\}$ are irreducible closed states. $\mathcal{S}_T = \{1,2\}$ are transient states. $\mathcal{S}_R = \{0,3,4,5\}$ are recurrent states.

Does $C_3 = \{1,2\}$ form a closed state? The answer is no. Verify!

Note that if $x \in C_1 \cup C_2$, then $\rho_{xy} = 1$, for some $y \in C_1 \cup C_2$, from Theorem 2.6. For e.g., $\rho_{34} = 1$

Suppose $x \in S_R \notin C_2$, such as $\{0\}$, then $\rho_{03} = 0$.

Using Proposition 2.9,

$$\rho_{10} = \mathbf{P}(1,0) + \sum_{y \neq \{0\}} \mathbf{P}(1,y)\rho_{y0}$$

which leads to,

$$\rho_{10} = \frac{1}{4} + \frac{1}{2}\rho_{10} + \frac{1}{4}\rho_{20}$$

and similarly,

$$\rho_{20} = 0 + \frac{1}{5}\rho_{10} + \frac{2}{5}\rho_{20} + \frac{1}{5}\rho_{30} + \frac{1}{5}\rho_{50}$$

Note that $\rho_{30} = \rho_{50} = 0$, since C_1 is unreachable from C_2. Solving, we get $\rho_{10} = \frac{3}{5}$ and $\rho_{20} = \frac{1}{5}$.

Birth and Death chain

Birth and death chain is a class of Markov chains that arise very commonly in applications. One would be able to spot that many common types of Markov chains are actually birth and death chains in disguise. Let us first describe the transition probability of the birth and death chain.

Let $\mathcal{S} = \{0, 1, \ldots, d\}$, where d could be finite or infinite. Then the transition probability is given by,

$$\mathbf{P}(x,y) = \begin{cases} q_x & y = x-1 \\ r_x & y = x \\ p_x & y = x+1 \end{cases}$$

where $p_x + q_x + r_x = 1$ and $q_0 = 0$. If d is finite, then $p_d = 0$. Notice at each state $\{x\}$ the chain can either go to a state which is an increment by one $\{x+1\}$, stay at the same state $\{x\}$ or move to a state that is a decrement by one $\{x-1\}$. We saw such a structure in the gambler's ruin problem. The way we defined gambler's ruin, in birth and death chain notation would be expressed as the following, $d = 4, r_0 = 1, p_1 = p_2 = p_3 = 0.4, q_1 = q_2 = q_3 = 0.6, r_4 = 1$. If d is not finite, we can study an extension of the gambler's ruin problem, where the gambler will only stop playing if she has been ruined. Earlier, we said the gambler would stop playing once she has accumulated 4 units of wealth.

Remark 4. *We only consider irreducible Birth and Death chains, all further references to Birth and Death chain imply that they are irreducible.*

The simple random walk that we discussed earlier is also a Birth and Death chain (Check how!). Birth and death chain can be used to model queues. Queuing theory is a major topic of study in Operations Research that looks at the analysis of formation of queues to design optimal layouts for managing traffic. The Birth and Death chain can be thought of as a queue forming, where the queue length at each time point increases or decreases by one, or stays the same. We will study a generalization of this later called the Birth and Death process.

Also note that we describe the transition probabilities p_x, q_x, r_x by a subscript x which denotes that the probability of the transition to another state depends on which state the chain is currently at, so for example it is possible in a birth and death chain for p_{10} and p_{100} to be different.

Proposition 2.10. *For $a, b \in \mathcal{S}$, where $a < b$, define,*

$$u(x) = \mathbb{P}_x(T_a < T_b), a < x < b$$

and set $u(a) = 1, u(b) = 0$. Here $u(x)$ is the probability of the event that a Birth and Death chain starting at $\{x\}$ hits state $\{a\}$ for the first time before it hits state $\{b\}$ for the first time. Then,

$$u(x) = \frac{\sum_{y=x}^{b-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y}$$

where $\gamma_y = \frac{q_1 \ldots q_y}{p_1 \ldots p_y}$

Proof. Notice that we can state $u(x)$ recursively in the following manner, for some $a < y < b$

$$u(y) = q_y u(y-1) + r_y u(y) + p_y u(y+1)$$

where in the next time-step the chain could either decrease, increase or stay at the same state. Since $r_y = 1 - p_y - q_y$,

$$u(y+1) - u(y) = \frac{q_y}{p_y}(u(y) - u(y-1))$$

Since $a < y < b$, substituting $y = a + 1$,

$$u(a+2) - u(a+1) = \frac{q_{a+1}}{p_{a+1}}(u(a+1) - u(a))$$

Similarly,

$$u(a+3) - u(a+2) = \frac{q_{a+2}}{p_{a+2}}(u(a+2) - u(a+1))$$

Therefore,

$$u(y+1) - u(y) = \frac{q_{a+1}q_{a+2}\cdots q_y}{p_{a+1}p_{a+2}\cdots p_y}(u(a+1) - u(a)) \tag{6}$$

We sum (6) in y from $a + 1$ to $b - 1$,

$$u(b) - u(a) = \left(1 + \frac{q_{a+1}}{p_{a+1}} + \frac{q_{a+1}q_{a+2}}{p_{a+1}p_{a+2}} + \ldots \frac{q_{a+1}q_{a+2}\cdots q_{b-1}}{p_{a+1}p_{a+2}\cdots p_{b-1}}\right)(u(a+1) - u(a))$$

Therefore,

$$u(a) - u(a+1) = 1/\left(1 + \frac{q_{a+1}}{p_{a+1}} + \frac{q_{a+1}q_{a+2}}{p_{a+1}p_{a+2}} + \ldots \frac{q_{a+1}q_{a+2}\cdots q_{b-1}}{p_{a+1}p_{a+2}\cdots p_{b-1}}\right)$$

Substituting above in (6),

$$u(y) - u(y+1) = \left(\frac{q_{a+1}q_{a+2}\cdots q_y}{p_{a+1}p_{a+2}\cdots p_y}\right) / \left(1 + \frac{q_{a+1}}{p_{a+1}} + \frac{q_{a+1}q_{a+2}}{p_{a+1}p_{a+2}} + \ldots \frac{q_{a+1}q_{a+2}\cdots q_{b-1}}{p_{a+1}p_{a+2}\cdots p_{b-1}}\right) \tag{7}$$

Summing (7) in y from $x, \dots, b-1$, $a < x < b$,

$$u(x) = \frac{\sum_{y=x}^{b-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y}$$

where $\gamma_y = \frac{q_1 \dots q_y}{p_1 \dots p_y}$ □

Exercises

1. Consider the gambler's ruin problem where $p_x = \frac{9}{19}$ and $q_x = \frac{10}{19}$, where she stops playing only if she is ruined. Given that the gambler's current wealth is 10 units, then evaluate the probability that the gambler is ruined before she reaches 15 units of wealth.

Proposition 2.11. *The Birth and Death chain is recurrent, i.e. all states are recurrent, iff,*

$$\sum_{x=0}^{\infty} \gamma_x = \infty$$

Proof. Note from Proposition 2.10,

$$\mathbb{P}_1(T_0 < T_n) = 1 - \frac{1}{\sum_{x=0}^{n-1} \gamma_x}$$

Then,

$$\mathbb{P}_1(T_0 < \infty) = \lim_{n \to \infty} \mathbb{P}_1(T_0 < T_n) = 1 - \frac{1}{\sum_{x=0}^{\infty} \gamma_x}$$

Say the Birth and Death chain is recurrent, then from Theorem 2.4, $\rho_{10} = 1$, i.e. $\mathbb{P}_1(T_0 < \infty) = 1$ and therefore, $\sum_{x=0}^{\infty} \gamma_x = \infty$.

Now suppose $\sum_{x=0}^{\infty} \gamma_x = \infty$.

$$\mathbb{P}_0(T_0 < \infty) = \mathbf{P}(0,0) + \mathbf{P}(0,1)\mathbb{P}_1(T_0 < \infty) + \mathbf{P}(0,2)\mathbb{P}_2(T_0 < \infty) + \mathbf{P}(0,3)\mathbb{P}_3(T_0 < \infty) + \dots$$

Now note that $P(0,x) = 0, \forall x > 1$. Therefore,

$$\mathbb{P}_0(T_0 < \infty) = \mathbf{P}(0,0) + \mathbf{P}(0,1)\mathbb{P}_1(T_0 < \infty)$$

Since, $\mathbb{P}_1(T_0 < \infty) = 1$, as $\sum_{x=0}^{\infty} \gamma_x = \infty$. Therefore,

$$\mathbb{P}_0(T_1 < \infty) = \mathbf{P}(0,0) + \mathbf{P}(0,1) = 1$$

Now $\mathbb{P}_1(T_0 < \infty) = \rho_{10} = 1$ and $\mathbb{P}_0(T_1 < \infty) = \rho_{01} = 1$, therefore from Theorem 2.4 $\{0\}$ is a recurrent state. Since the chain is irreducible and $\{0\}$ is recurrent, by Theorem 2.5 all states are recurrent. □

Corollary 2.1. *The Birth and Death chain is transient, i.e. all states are transient, iff,*

$$\sum_{x=0}^{\infty} \gamma_x < \infty$$

Exercises

1. Consider a Birth and Death chain where d is not finite. Define $p_x = \frac{x+2}{2x+2}$ and $q_x = \frac{x}{2x+2}$. Is this chain recurrent or transient?

Stationary Distribution

One of the goals of Probability theory is to understand the limiting behavior of a random system. The Central Limit Theorem (CLT) roughly states that an average of a large collection of random variables follows the Normal distribution, irrespective of the distribution of the random variables themselves. This is a profound result because it states that the limiting behavior of the random systems in some sense doesn't depend on the distribution of the random variables themselves. This makes it also a very useful result, since even if random system is too difficult to analyse, the CLT allows us to study the system in the limiting case. As many of you know CLT has immense applications in statistics and many other disciplines.

With similar motivations, we will seek to study the stationary distribution for a Markov chain. Consider the humidity in Kanpur example. What is the long-run proportion of days for which Kanpur is Humid? To remind, the transition matrix is the following,

$$\mathbf{P} = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} \end{array}$$

One approach could be to study $\mathbf{P}^{(n)}$, where n is large, i.e. $\lim_{n\to\infty} \mathbf{P}^{(n)}$. To actually evaluate $\mathbf{P}^{(n)}$ for a large value of n means that we are multiplying the transition matrix to itself many times. This is computationally expensive, the cost increases as the number of states in a Markov chain increase. Instead let us consider the system of equations,

$$\begin{aligned} 0.9\pi(0) + 0.2\pi(1) &= \pi(0) \\ 0.1\pi(0) + 0.8\pi(1) &= \pi(1) \end{aligned}$$

Since π is a probability distribution,

$$\pi(0) + \pi(1) = 1$$

The solution of these system of equations is, $\pi(0) = \frac{2}{3}, \pi(1) = \frac{1}{3}$. Therefore, we will claim this implies that the long-run proportion of days for which Kanpur is Humid is 0.667, i.e. $\lim_{n\to\infty} \mathbb{P}(X_n = 1) = 0.667$. To see why we can claim this, check the paragraph below. Notice that solving this system of equations is more feasible than evaluating $\lim_{n\to\infty} \mathbf{P}^{(n)}$. In this section, we will try to identify the conditions for which $\lim_{n\to\infty} \mathbf{P}^{(n)}(x, y) = \pi(y)$ takes place.

But why does evaluating π as a proxy for $\lim_{n\to\infty} \mathbf{P}^{(n)}$ even make sense?

From Proposition 2.3,

$$\mathbb{P}(X_n = y) = \sum_{x\in\mathcal{S}} \pi_0(x)\mathbf{P}^{(n)}(x, y)$$

Taking limit,

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(X_n = y) &= \lim_{n\to\infty} \sum_{x\in\mathcal{S}} \pi_0(x)\mathbf{P}^{(n)}(x, y) \\ &= \sum_{x\in\mathcal{S}} \pi_0(x)\pi(y) \text{ (Applying } \lim_{n\to\infty} \mathbf{P}^{(n)}(x, y) = \pi(y)\text{, if it exists)} \\ &= \pi(y) \text{ (Since } \sum_{x\in\mathcal{S}} \pi(x) = 1) \end{aligned}$$

Therefore, the stationary distribution (when it exists) gives us the long term probability of a Markov chain attaining a state. Formally the **stationary distribution**, $\pi(x)$, is the solution of the following set of equations,

$$\sum_{x\in\mathcal{S}} \pi(x)\mathbf{P}(x,y) = \pi(y), \forall y \in \mathcal{S} \tag{8}$$

Since $\pi(x)$ is a probability distribution, therefore, $\sum_{x\in\mathcal{S}} \pi(x) = 1$. In matrix notation this is written as, $\pi\mathbf{P} = \pi$.

Exercises

1. Evaluate the stationary distribution of a Markov chain with the following transition matrix,

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{bmatrix} \end{array}$$

Remark 5. *Now consider $\sum_{x\in\mathcal{S}} \pi(x)\mathbf{P}^{(2)}(x,y)$, using Chapman-Kolmogorov equation this can be simplified in the following way,*

$$\begin{aligned}
&\sum_{x\in\mathcal{S}} \pi(x)\mathbf{P}^{(2)}(x,y) \\
=&\sum_{x\in\mathcal{S}} \pi(x)\sum_{z\in\mathcal{S}}\mathbf{P}(x,z)\mathbf{P}(z,y) \ \textit{(Applying Chapman-Kolmogorov Equation)} \\
=&\sum_{z\in\mathcal{S}}\left(\sum_{x\in\mathcal{S}} \pi(x)\mathbf{P}(x,z)\right)\mathbf{P}(z,y) \\
=&\sum_{z\in\mathcal{S}} \pi(z)\mathbf{P}(z,y) \ \textit{(From (8))} \\
=&\pi(y) \ \textit{(From (8))}
\end{aligned}$$

Similarly, we can show that for any n,

$$\sum_{x\in\mathcal{S}} \pi(x)\mathbf{P}^{(n)}(x,y) = \pi(y) \tag{9}$$

This motivates why we solve (8) to obtain the stationary distribution, as one can note that as n approaches ∞ the RHS is still the stationary distribution.

Stationary distribution of the Birth and Death chain

Consider the irreducible Birth and Death chain introduced earlier.

Now consider the system of equations,

$$\sum_{x\in\mathcal{S}} \pi(x)\mathbf{P}(x,y) = \pi(y), y \in \mathcal{S}$$

Therefore,

$$\begin{aligned}
\pi(0)r_0 + \pi(1)q_1 &= \pi(0) \\
\pi(y-1)p_{y-1} + \pi(y)r_y + \pi(y+1)q_{y+1} &= \pi(y), y \geq 1
\end{aligned}$$

Also, $p_y + q_y + r_y = 1$.

It follows that,

$$\pi(y+1) = \frac{p_y}{q_{y+1}}\pi(y)$$

Therefore,

$$\pi(x) = \frac{p_0 \cdots p_{x-1}}{q_1 \cdots q_x}\pi(0), x \geq 1$$

Let $\pi_x = \frac{p_0 \cdots p_{x-1}}{q_1 \cdots q_x}$. Then,

$$\pi(x) = \pi_x \pi(0)$$

And from $\sum_{d=0}^{\infty} \pi_d = 1$,

$$\pi(x) = \frac{\pi_x}{\sum_{d=0}^{\infty} \pi_d}$$

if $\sum_{d=0}^{\infty} \pi_d < \infty$. Therefore, the Birth and Death chain has a stationary distribution iff $\sum_{d=0}^{\infty} \pi_d < \infty$.

Exercises

1. Evaluate the stationary distribution of the Markov chain with transition matrix,

$$\begin{array}{c c} & \begin{array}{cccc} 0 & 1 & 2 & 4 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

Mean return time to a Transient state

If y is a Transient state, then $\mathbb{P}_y(T_y = \infty) > 0$, i.e., there is a positive probability that it takes forever to return to y. Therefore, $\mathbb{E}_y[T_y] = \infty$. In words, this means that the average time for a transient state to return to itself is infinite.

Now suppose y is a recurrent state. We define the **mean return time** as,

$$m_y = \mathbb{E}_y[T_y]$$

where m_y means the mean time taken by a chain at y to return to y.

If $m_y < \infty$, then we say that y is **positive recurrent**. If $m_y = \infty$, we say that it is **null recurrent**. It is a more precise classification of a recurrent state, where we identify how quickly does a recurrent state return to itself on average.

There is a connection between stationary distributions and recurrent states that we will explore now. In the remainder of the notes we state all results without proofs, interested may refer to Chapter 2 of Hoel, Port and Stone to understand the proofs.

Theorem 2.7 (Some results on Positive Recurrent states)**.** *1. Let x be a positive recurrent state and suppose $x \to y$. Then y is positive recurrent.*

2. Let C be finite irreducible set of states, then all states in C are positive recurrent.

3. A Markov chain with finite number of states, has no null recurrent states, i.e. all recurrent states are positive recurrent in such a chain.

Theorem 2.8. *Let π be a stationary distribution. If x is a transient state or a null recurrent state, then $\pi(x) = 0$.*

Theorem 2.9. *An irreducible positive recurrent Markov chain has a unique stationary distribution π, given by,*

$$\pi(x) = \frac{1}{m_x}$$

Theorem 2.10. *An irreducible Markov chain is positive recurrent, i.e. all states are positive recurrent, iff it has a stationary distribution.*

Example 2.12 (Example of null recurrent chain)**.** *We have established that for an irreducible Birth and Death chain to be transient,*

$$\sum_{x=1}^{\infty} \gamma_x = \frac{q_x \cdots q_1}{p_x \cdots p_1} < \infty \tag{10}$$

For a Birth and death chain to have a stationary distribution it was shown that,

$$\sum_{x=1}^{\infty} \pi_x = \sum_{x=1}^{\infty} \frac{p_0 \cdots p_{x-1}}{q_1 \cdots q_{x-1}} < \infty \tag{11}$$

thus, it is also the condition for positive recurrence (from Theorem 2.10).

Thus the chain is null recurrent iff (10) and (11) fail, i.e., they are null recurrent iff the below conditions hold simultaneously,

$$\sum_{x=1}^{\infty} \frac{q_x \cdots q_1}{p_x \cdots p_1} = \infty$$

and

$$\sum_{x=1}^{\infty} \frac{p_{x-1} \cdots p_0}{q_x \cdots q_1} = \infty$$

Convergence to stationary distribution

Consider a Markov chain where $\rho_{xx} > 0, \forall x \in \mathcal{S}$. The **period** of such a Markov chain is defined as,

$$d_x = \gcd\{n \geq 1 : \mathbf{P}^{(n)}(x, x) > 0\}$$

where gcd is the greatest common divisor. We require the $\rho_{xx} > 0$ condition, otherwise, if there is 0 probability of a chain returning to a state the notion of periodicity becomes meaningless in such a case.

Exercises

1. Consider the Transition Matrix,

$$\begin{array}{c c} & \begin{array}{cccc} 0 & 1 & 2 & 4 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

What is d_1?

Theorem 2.11. *All states in an irreducible Markov chain have a common period.*

We say that an irreducible chain is **periodic** with period d if $d > 1$ and **aperiodic** if $d = 1$.

Exercises

1.

$$\begin{array}{cc} & \begin{array}{cccc} 0 & 1 & 2 & 4 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{array}$$

What is the period of this Markov chain?

In the beginning of this section we wanted to understand when does the following limit exist?

$$\lim_{n\to\infty} \mathbf{P}^{(n)}(x,y) = \pi(y)$$

Now we answer this question,

Theorem 2.12. *Let X_n be an irreducible, positive recurrent, aperiodic Markov chain having stationary distribution π. Then,*

$$\lim_{n\to\infty} \mathbf{P}^{(n)}(x,y) = \pi(y), x, y \in \mathcal{S}$$

Remark 6. *Note that a Markov chain with transient or null recurrent states may also possess a stationary distribution, however the limit of the transition matrix may not converge to the stationary distribution. More importantly, the long-run interpretation of the stationary distribution also may not hold. In the Kanpur Humidity example, we were able to interpret the stationary distribution as the long-run probability because it satisfies the above conditions for convergence. Verify that the Transition Matrix of the Kanpur Humidity case satisfies the properties above!*

Remark 7. *There is a more nuanced notion of the limit when the chain is periodic, but we shall not pursue this here. Check Chapter 2, Hoel, Port and Stone if you are interested.*

References

1. Sheldon Ross, Introduction to Probability Models, Academic Press, 2024.
2. Hoel, Port, Stone, Introduction to Stochastic Processes, Houghton Mifflin Company, 1972.
3. Rick Durrett, Essentials of Stochastic Processes, Springer, 1999.
4. Sidney Resnick, Adventures in Stochastic Processes, Springer, 1992.