4.1 Stochastic Matrices

**Definition 4.1.** Let $I$ be a countable set.

(a) A square matrix $P = ((p_{ij}))_{i,j \in I}$ is called a **stochastic matrix** if

\[ p_{ij} \geq 0 \text{ for all } i, j \in I \text{ and } \sum_{j \in I} p_{ij} = 1 \text{ for all } i \in I. \]

(b) A square matrix $P$ is called a **sub-stochastic matrix** if it satisfies (a) and (iii).

\[ \sum_{j=1}^{n} p_{ij} \leq 1 \text{ for all } i \in I. \]

(c) A square matrix $P$ is called a **doubly-stochastic matrix** if $P$ and $P^T$ are both stochastic matrices.

The way to think of the entry $p_{ij}$ is, roughly, “the probability of going to state $j$ next, given that we are at state $i$ now”. We will make this more precise, but for now think of these probabilities as “transition probabilities”. As such, a stochastic matrix is also commonly called a transition/Markov matrix, or a generator.

**Definition 4.2.** A directed graph $G = (V, E)$ is a set of vertices $V$ and collection of directed edges $(i, j) \in E \subset V \times V$ from vertex $i$ to vertex $j$. A weighed graph $(V, E, W)$ is a graph $(V, E)$ with the weight of an edge $e = (i, j)$ given by $w_e = w_{ij}$. We will assume that $w_{ij} = 0$ iff $(i, j)$ is not an edge in the graph.

Given a Stochastic matrix $P = ((p_{ij}))_{i,j \in I}$ we can define a weighed directed graph with vertex set $I$ and edge weight of $(i, j)$ given by $p_{ij}$. Note that, $\sum_{j \in I} p_{ij} = 1$ for all $i$ implies that the total weight of all the out-edges from the vertex $i$ is 1.

**Example 4.3.** Given a stochastic matrix $P = \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1/4 & 1/4 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$ then the corresponding edge-
A path in a directed graph $G = (V, E)$ is a collection of edges $(i_0, i_1), (i_1, i_2), \ldots, (i_{n-1}, i_n)$ where $i_0, i_1, \ldots, i_n \in V$. Moreover, the length of a path is the number of edges in this path, and the weight of a path is defined as $\prod_{\text{edges in the path}}$ edge weights.

A cycle is a path with the same starting and ending vertices.

Example 4.5. (a) $\{(2,3), (3,2), (2,3), (3,2)\}$ is a path of $G$ with length (i.e., number of edges) 4.
(b) $\{(2,1), (1,1), (1,1)\}$ is a path of $G$ with length 3 and weight $= 1/2 \cdot 1 \cdot 1 = 1/2$.
(c) For example 4.3, to reach 1 from 2 in 2 steps, there are 3 paths:

$\{(2,1), (1,1)\}$
$\{(2,2), (2,1)\}$
$\{(2,3), (3,1)\}$

The total weight of these paths is

$$(P \cdot P)_{21} = \sum_{k=1}^{n} p_{2k} p_{k1} = p_{21} p_{11} + p_{22} p_{21} + p_{23} p_{31}.$$ 

Exercise 4.1. Show that $(P^k)_{ij} = \text{Total weight of all paths from } i \text{ to } j \text{ of length } k$.

Remark 4.6. $P$ is a stochastic matrix if and only if $P \cdot 1 = 1$ where $1 = (1,1,\ldots,1)^T$. Thus, if $P$ is stochastic, so are $P^k$ for all positive integer $k$.

Theorem 4.7. For a stochastic matrix $P$, the following statements hold:

i. $\text{spec}(P) \subseteq B(0,1)$,

ii. If $p_{ii} > 0$ for all $i$, then $\text{spec}(P) \subseteq B(0,1) \cup \{1\}$.

Proof. i. For all $i$, $R_i = \sum_{j \neq i} |p_{ij}| = 1 - p_{ii}$. Suppose $z$ is an eigenvalue of $P$. Then by Gershgorin circle theorem, $z \in B(p_{ii}, 1 - p_{ii})$ and thus, $|z| \leq 1$.

Follows from the fact that $B(p_{ii}, 1 - p_{ii}) \subseteq B(0,1) \cup \{1\}$ if $p_{ii} > 0$. ■
4.2 Discrete Time Markov Chains

**Definition 4.8.** Given a probability space $(\Omega, \mathcal{F}, P)$, a set $I$, and an ordered set $\mathcal{T}$, a stochastic process (with values in $I$) is a collection of random variables $(X_t)_{t \in \mathcal{T}}$ such that $X_t : \Omega \to I$ is a random variable for each $t \in \mathcal{T}$. We call $I$ the state space of the stochastic process and $\mathcal{T}$ the time domain. Moreover, if $\mathcal{T}$ is countable (e.g., $\mathbb{N} = \{0, 1, 2, \ldots\}$), then we get a discrete-time stochastic process.

A stochastic process is a function of two variables, and could be written $X(\omega, t)$. But it is more commonly written as $X_t(\omega)$, since we think of these two variables differently. For each fixed $t \in \mathcal{T}$, $X_t$ is a random variable, and once we choose a specific outcome in $\Omega$, we have the number $X_t(\omega)$.

In this sense, a stochastic process is a sequence of random variables, indexed by time. On the other hand, for each fixed $\omega \in \Omega$, $X_t(\omega)$ is a curve or path in the set $I$. So that we can think of a stochastic process as a random variable whose values are paths in $I$. Both of these points of view are valid and each will be useful in various contexts.

**Definition 4.9.** Consider a discrete-time stochastic process $(X_t)_{t \geq 0}$ defined on an underlying probability space $(\Omega, \mathcal{F}, P)$ and taking values in the set $I$. The random process is called a discrete-time Markov chain with initial distribution $\lambda$ and transition matrix $P$ if,

1. $X_0 \sim \lambda$, where $\lambda$ is the initial distribution
2. $P(X_n = j | X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = i) = P(X_n = j | X_{n-1} = i) = p_{ij}$, implies, given $X_{n-1}$, the random number $X_n$ is independent of the history, $X_0, X_1, \ldots, X_{n-2}$.

Here, $P = ((p_{ij}))$ is called the transition matrix and we will use the notation $(X_t)_{t \geq 0} \sim \text{Markov}(\lambda, P)$.

If $p_{ij}$'s are independent of time (consequently, transition matrix $P$ is also independent of time), then the Markov chain is called the time-homogeneous Markov chain.

### 4.2.1 Examples

**Random Walk on integers:** Consider the set of integers and a Markov chain $(X_i)_{i \geq 0}$. The random variables take integer values, i.e. $I = \mathbb{Z}$. Consider $X_0 = 1$ (or some other distribution $\lambda$). The value of the random variable in the next step increases by 1 with probability 1/4, decreases by 1 with probability 1/4, and stays the same with probability 1/2. This can be formally written as

$$p_{ij} = \begin{cases} 
1/4, & j = i + 1 \\
1/4, & j = i - 1 \\
1/2, & j = i \\
0, & \text{otherwise}
\end{cases}$$

The transition probability matrix, $P$, has a special structure (tridiagonal matrix) in this example and we write, $(X_t)_{t \geq 0} \sim \text{Markov}(\lambda, P)$. This process denoted as $(X_t)_{t \geq 0} \sim \text{Markov}(\lambda, P)$, is also called Lazy Random Walk. It is called lazy since the Markov chain stays the same with a probability of 1/2.
**Random Walk with absorption:** Consider an integer $n$ and a Markov chain $(X_t)_{t \geq 0}$. The random variables take integer values in $I = \{-N, -(N-1), \ldots, -1, 0, 1, \ldots, N\}$. Consider $X_0 = 1$ (or some distribution $\lambda$). The value of the random variable in the next step increases by 1 with probability $1/4$, decreases by 1 with probability $1/4$, and stays the same with probability $1/2$. Moreover, once the random variable reaches the end, $\{\pm N\}$, the random variable stays there with a probability of 1. We can write the transition probability matrix as,

$$P = \begin{bmatrix}
1 & 0 & 0 & 0 & \ldots & 0 \\
0.25 & 0.5 & 0.25 & 0 & \ldots & 0 \\
0 & 0.25 & 0.5 & 0.25 & \ldots & 0 \\
0 & 0 & 0.25 & \ddots & \ddots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \ldots & 1
\end{bmatrix}.$$ 

**Pattern Recognition:** Consider a sequence of i.i.d. Ber($p$) random variables $Y_t, t \geq 0$ where $Y_0$ takes value 1 with probability $p$ and takes value 0 with probability $1 - p$. Now let us focus on finding the pattern $S = (1, 1)$, i.e., the value of two successive random variables is 1. We consider new random variables $X_t$ defined as $X_0 = (Y_0, Y_1), X_1 = (Y_1, Y_2), \ldots, X_t = (Y_t, Y_{t+1}), \ldots$. The random variable $X_t$ takes values in set $I = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. The transition probabilities for $X$ can be written as,

$$P(X_n = (c, d) \mid X_{n-1} = (a, b)) = \begin{cases} 
0, & \text{if } b \neq c \\
p, & \text{if } b = c \text{ and } d = 1 \\
1 - p, & \text{if } b = c \text{ and } d = 0
\end{cases} \quad \rightarrow P = \begin{bmatrix}
1 - p & p & 0 & 0 \\
0 & 0 & 1 - p & p \\
1 - p & p & 0 & 0 \\
0 & 0 & 1 - p & p
\end{bmatrix}.$$ 

![Pattern Recognition as Markov Process.](image)