Say we would like to model the USD price of bitcoin. We could observe the actual price at every hour and record it by a sequence of real numbers \(x_1, x_2, \cdots\). However, it is more interesting to build a ‘model’ that could predict the price of bitcoin at time \(t\), or at least give some meaningful insight how the actual bitcoin price behaves over time. Since there are so many factors affecting the price at every time, it might be reasonable that the price at time \(t\) should be given by a certain RV, say \(X_t\). Then our sequence of predictions would be a sequence of RVs, \((X_t)_{t \geq 0}\). This is an example of stochastic processes. Here ‘process’ means that we are not interested in just a single RV, that their sequence as a whole: ‘stochastic’ means that the way the RVs evolve in time might be random.

In this note, we will be studying a very important class of stochastic processes called Markov chains. The importance of Markov chains lies two places: 1) They are applicable for a wide range of physical, biological, social, and economical phenomena, and 2) the theory is well-established and we can actually compute and make predictions using the models.

1. Definition and examples

**Roughly speaking, Markov processes** are used to model temporally changing systems where future state only depends on the current state. For instance, if the price of bitcoin tomorrow depends only on its price today, then bitcoin price can be modeled as a Markov process. (Of course, the entire history of price often affects decisions of buyers/sellers so it may not be a realistic assumption.)

Even though Markov processes can be defined in vast generality, we concentrate on the simplest setting where the state and time are both discrete. Let \(\Omega = \{1, 2, \cdots, m\}\) be a finite set, which we call the state space. Consider a sequence \((X_t)_{t \geq 0}\) of \(\Omega\)-valued RVs, which we call a chain. We call the value of \(X_t\) the state of the chain at time \(t\). In order to narrow down the way the chain \((X_t)_{t \geq 0}\) behaves, we introduce the following properties:

(i) (Markov property) The distribution of \(X_{t+1}\) given the history \(X_0, X_1, \cdots, X_t\) depends only on \(X_t\). That is,
\[
P(X_{t+1} = k | X_t = j_t, X_{t-1} = j_{t-1}, \cdots, X_1 = j_1) = P(X_{t+1} = k | X_t = j_t).
\]

(ii) (Time-homogeneity) The transition probabilities
\[
p_{ij} = P(X_{t+1} = j | X_t = i) \quad i, j \in \Omega
\]
do not depend on \(t\).

When the chain \((X_t)_{t \geq 0}\) satisfies the above two properties, we say it is a (discrete-time and time-homogeneous) Markov chain. We define the transition matrix \(P\) to be the \(m \times m\) matrix of transition probabilities:
\[
P = (p_{ij})_{1 \leq i, j \leq m} = \begin{pmatrix}
p_{11} & p_{12} & \cdots & p_{1m} \\
p_{21} & p_{22} & \cdots & p_{2m} \\
& & \ddots & \\
p_{m1} & p_{m2} & \cdots & p_{mm}
\end{pmatrix}
\]

Finally, since the state \(X_t\) of the chain is a RV, we represent its probability mass function (PMF) via a row vector
\[
r_t = [P(X_t = 1), P(X_t = 2), \cdots, P(X_t = m)].
\]

**Example 1.1 (Gambler’s ruin).** Suppose a gambler has fortune of \(k\) dolors initially and starts gambling. At each time he wins or loses 1 dolor independently with probability \(p\) and \(1 - p\), respectively. The game ends when his fortune reaches either 0 or \(N\) dolors. What is the probability that he wins \(N\) dolors and goes home happy?
We use Markov chains to model his fortune after betting $t$ times. Namely, let $\Omega = \{0, 1, 2, \ldots, N\}$ be the state space. Let $(X_t)_{t \geq 0}$ be a sequence of RVs where $X_t$ is the gambler’s fortune after betting $t$ times. We first draw the state space diagram for $N = 4$ below: Next, we can write down its transition probabilities as

$$
\begin{align*}
\mathbb{P}(X_{t+1} = k + 1 | X_t = k) &= p & \forall 1 \leq k < N \\
\mathbb{P}(X_{t+1} = k | X_t = k - 1) &= 1 - p & \forall 1 \leq k < N \\
\mathbb{P}(X_{t+1} = 0 | X_t = 0) &= 1 \\
\mathbb{P}(X_{t+1} = N | X_t = N) &= 1.
\end{align*}
$$

For example, the transition matrix $P$ for $N = 5$ is given by

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

We call the resulting Markov chain $(X_t)_{t \geq 0}$ the gambler’s chain. ▲

**Example 1.2 (Ehrenfest Chain).** This chain is originated from the physics literature as a model for two cubical volumes of air connected by a thin tunnel. Suppose there are total $N$ indistinguishable balls split into two “urns” $A$ and $B$. At each step, we pick up one of the $N$ balls uniformly at random, and move it to the other urn. Let $X_t$ denote the number of balls in urn $A$ after $t$ steps. This is a Markov chain called the Ehrenfest chain. (See the state space diagram in Figure 2.)

It is easy to figure out the transition probabilities by considering different cases. If $X_t = k$, then urn $B$ has $N - k$ balls at time $t$. If $0 < k < N$, then with probability $k/N$ we move one ball from $A$ to $B$ and with probability $(N - k)/N$ we move one from $B$ to $A$. If $k = 0$, then we must pick up a ball from urn $B$ so $X_{t+1} = 1$ with probability 1. If $k = N$, then we must move one from $A$ to $B$ and $X_{t+1} = N - 1$ with probability 1. Hence, the transition kernel is given by

$$
\begin{align*}
\mathbb{P}(X_{t+1} = k + 1 | X_t = k) &= (N - k)/N & \forall 0 \leq k < N \\
\mathbb{P}(X_{t+1} = k - 1 | X_t = k) &= k/N & \forall 0 < k \leq N \\
\mathbb{P}(X_{t+1} = 1 | X_t = 0) &= 1 \\
\mathbb{P}(X_{t+1} = N - 1 | X_t = N) &= 1.
\end{align*}
$$
For example, the transition matrix $P$ for $N = 5$ is given by

$$P = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 \\
1/5 & 0 & 4/5 & 0 & 0 \\
0 & 2/5 & 0 & 3/5 & 0 \\
0 & 0 & 3/5 & 0 & 2/5 \\
0 & 0 & 0 & 4/5 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}. \quad (8)$$

**Example 1.3.** Let $\Omega = \{1, 2\}$ and let $(X_t)_{t \geq 0}$ be a Markov chain on $\Omega$ with the following transition matrix

$$P = \begin{bmatrix} p_{11} & p_{12} \\
p_{21} & p_{22} \end{bmatrix}. \quad (9)$$

We can also represent this Markov chain pictorially as in Figure 3, which is called the 'state space diagram' of the chain $(X_t)_{t \geq 0}$.

For some concrete example, suppose $p_{11} = 0.2, \ p_{12} = 0.8, \ p_{21} = 0.6, \ p_{22} = 0.4$. (10)

If the initial state of the chain $X_0$ is 1, then

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 1 \mid X_0 = 1)\mathbb{P}(X_0 = 1) + \mathbb{P}(X_1 = 1 \mid X_0 = 2)\mathbb{P}(X_0 = 2)$$

$$= \mathbb{P}(X_1 = 1 \mid X_0 = 1) = p_{11} = 0.2 \quad (11)$$

and similarly,

$$\mathbb{P}(X_1 = 2) = \mathbb{P}(X_1 = 2 \mid X_0 = 1)\mathbb{P}(X_0 = 1) + \mathbb{P}(X_1 = 2 \mid X_0 = 2)\mathbb{P}(X_0 = 2)$$

$$= \mathbb{P}(X_1 = 2 \mid X_0 = 1) = p_{12} = 0.8 \quad (12)$$

Also we can compute the distribution of $X_2$. For example,

$$\mathbb{P}(X_2 = 1) = \mathbb{P}(X_2 = 1 \mid X_1 = 1)\mathbb{P}(X_1 = 1) + \mathbb{P}(X_2 = 1 \mid X_1 = 2)\mathbb{P}(X_1 = 2)$$

$$= p_{11}\mathbb{P}(X_1 = 1) + p_{21}\mathbb{P}(X_1 = 2) \quad (13)$$

$$= 0.2 \cdot 0.2 + 0.6 \cdot 0.8 = 0.04 + 0.48 = 0.52. \quad (14)$$

In general, the distribution of $X_{t+1}$ can be computed from that of $X_t$ via a simple linear algebra. Note that for $i = 1, 2$,

$$\mathbb{P}(X_{t+1} = i) = \mathbb{P}(X_{t+1} = i \mid X_t = 1)\mathbb{P}(X_t = 1) + \mathbb{P}(X_{t+1} = i \mid X_t = 2)\mathbb{P}(X_t = 2)$$

$$= p_{i1}\mathbb{P}(X_t = 1) + p_{i2}\mathbb{P}(X_t = 2). \quad (15)$$

This can be written as

$$[\mathbb{P}(X_{t+1} = 2), \mathbb{P}(X_{t+1} = 2)] = [\mathbb{P}(X_{t+1} = 1), \mathbb{P}(X_{t+1} = 2)] \begin{bmatrix} p_{11} & p_{12} \\
p_{21} & p_{22} \end{bmatrix}. \quad (16)$$

That is, if we represent the distribution of $X_t$ as a row vector, then the distribution of $X_{t+1}$ is given by multiplying the transition matrix $P$ to the left.

We generalize this observation in the following exercise.
Exercise 1.4. Let \((X_t)_{t \geq 0}\) be a Markov chain on state space \(\Omega = \{1, 2, \ldots, m\}\) with transition matrix \(P = (p_{ij})_{1 \leq i, j \leq m}\). Let \(r_t = [P(X_t = 1), \ldots, P(X_t = m)]\) denote the row vector of the distribution of \(X_t\).

(i) Show that for each \(i \in \Omega\),
\[
P(X_{t+1} = i) = \sum_{j=1}^{m} p_{ji} P(X_t = j).
\]  
(21)

(ii) Show that for each \(t \geq 0\),
\[
r_{t+1} = r_t P
\]  
(22)

(iii) Show by induction that for each \(t \geq 0\),
\[
r_t = r_0 P^t.
\]  
(23)

REFERENCES