Markov Chains

A Markov chain is specified by:
(a) set of states \( S = \{1, 2, \ldots, m\} \)
(b) a matrix [Transition Probability Matrix]
\[ \mathbf{T} = \{P_{ij}\}, i, j \in S \]
[it is \( m \times m \) matrix]

A Markov chain is a sequence of random variables \( X_0, X_1, X_2, \ldots \) taking values in \( \{1, 2, \ldots, m\} \) and satisfying

\[ P(X_{n+1} = j \mid X_n = i, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1, X_0 = x_0) = P_{ij} \]

for all \( n, j, i, x_{n-1}, \ldots, x_0 \).

Basically, the conditional probability of \( X_{n+1} = j \) given the entire past \( \{x_n, \ldots, x_0\} \) depends only on the state \( X_{n+1} = j \) & \( X_n = i \), irrespective of the values of \( X_{n-1}, \ldots, X_0 \).

For example: Suppose it will rain tomorrow with probability 0.8 if it rained today [yesterday & days before that are irrelevant].

If it rains tomorrow with probability 0.3 if it did not rain today.
let state 1 correspond to rain
& state 2 correspond to no rain

the \( P_{11} = 0.8 \) \[= P(\text{rain tomorrow | rain today}) \]
\( P_{12} = 0.2 \) \[= P(\text{no rain tomorrow | rain today}) \]
\( P_{21} = 0.3 \) \[= P(\text{rain tomorrow | no rain today}) \]
\( P_{22} = 0.7 \) \[= P(\text{no rain tomorrow | no rain today}) \]

Note that \( P_{11} + P_{12} = 1 \)
\& \( P_{21} + P_{22} = 1 \)

\[ \Phi T = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix} \]

Rows of \( \Phi T \) sum to 1.

We can also represent this Markov chain as a "transition probability graph"

\[ \begin{array}{c}
0.8 \quad 1 \quad 0.2 \\
\text{rain} \quad \text{no rain} \\
\end{array} \]

If \( X_0, X_1, \ldots \) is a Markov chain
then
\[ P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n) = P(X_0 = i_0) P(X_1 = i_1 | X_0 = i_0) P(X_2 = i_2 | X_0 = i_0, X_1 = i_1) \ldots P(X_n = i_n | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \]  [Chain Rule]
Thus the entire joint probability for the sequence \( X_0 \rightarrow \cdots \rightarrow X_n \) can be expressed in terms of \( P(X_0 = i_0) \) & the transition matrix.

Similar to \( \Pi \), we can define the \( n \)-step transition probability matrix

\[
R(n) = \{ r_{ij}(n) \}_{i,j=1}^n
\]

where entry \( r_{ij}(n) = P(X_n = j | X_0 = i) \)

Note that \( r_{ij}(1) = P(X_1 = j | X_0 = i) = P_{ij} \)

\[
R(1) = \Pi
\]

Can we express \( R(n) \) in terms of \( R(n-1) \)?

Consider \( r_{ij}(n) = P(X_n = j | X_0 = i) \)

\[
= \sum_{k=1}^{m} P(X_n = j, X_{n-1} = k | X_0 = i)
\]

[Total Probability Rule]

\[
= \sum_{i=1}^r P(X_n = j | X_0 = i)
\]

\[
= \sum_{i=1}^r P(X_{n-1} = k | X_0 = i) P(X_n = j | X_{n-1} = k, X_0 = i)
\]

[Chain Rule]

\[
= \sum_{i=1}^r \gamma_{ik}(n-1) \pi_{kj}
\]
We can write in terms of matrix product

\[ R(n) = R(n-1) \times \Pi \]

Chapman-Kolmogorov Equation.

In particular, since \( R(1) = \Pi \)

\[ R(2) = \Pi^2 \]

\[ R(n) = \Pi^n \]

We are usually interested in the evolution of the state probabilities of different states as the chain progresses.

\[ \therefore \text{ we define } V(n) = [V_1(n), \ldots, V_m(n)] \]

\[ \text{ where } V_i(n) = P(X_n = i) \]

\[ \therefore V(n) \text{ is the PMF of } X_n \]

Recall that \( P(X_0 = i) \) needs to be specified in order to calculate the joint probabilities. Thus, we will attempt to express \( V(n) \) in terms of \( V(0) \) and \( \Pi \).

\[ V_i(n) = P(X_n = i) = \sum_{k=1}^{m} P(X_0 = k) P(X_n = i | X_0 = k) \]

[Total Probability & Chain Rules]

\[ = \sum_{k=1}^{m} V_i(0) \chi_{ki}(n) \]
Let \( V(n) = V(0) \cdot R(n) \)

\[
\begin{align*}
V(n) &= V(0) \cdot T^n \\
V(n) &= V(0) \cdot T^n \\
V(n) &= V(0) \cdot T^n \\
\end{align*}
\]

Note dimensions:
- \( V(n) \): \( 1 \times m \)
- \( V(0) \): \( 1 \times m \)
- \( R(n) \): \( m \times m \)

**Def:** \( V(0) \) is a "stationary distribution" if

\[ V(0) = V(0) \cdot T \]

[math in other words, \( V(0) \) is a left eigenvector of the matrix \( T \) with eigenvalue \( = 1 \) ]

Why is it called "stationary"?

If \( V(0) \) is stationary

\[
\begin{align*}
V(1) &= V(0) \cdot T = V(0) \\
V(2) &= V(0) \cdot T^2 = V(0) \cdot T \cdot T = V(0) \cdot T = V(0) \\
\therefore V(n) &= V(0) \quad \forall \ n
\end{align*}
\]

\[ \therefore V(n) = V(0) \quad \forall \ n \]

If Markov chains starts with distribution \( V(0) \), it stays in the distribution \( V(0) \) \( \forall \ n \).

**Example:** Recall our earlier example with

\[
T = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}
\]

Let's compute its stationary distribution

\[
V(0) = (V_1(0) \ V_2(0)) \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}
\]
we get two equations
\[ v_1(0) = 0.8 v_1(0) + 0.3 v_2(0) \]
\[ v_2(0) = 0.2 v_1(0) + 0.7 v_2(0) \]

Solving: \[ v_1(0) = \frac{3}{2} v_2(0) \]

Also see note: \[ v_1(0) + v_2(0) = 1 \]

\[ \Rightarrow \frac{3}{2} v_2(0) + v_2(0) = 1 \]
\[ \Rightarrow v_2(0) = \frac{2}{5} \]
\[ \Rightarrow v_1(0) = \frac{3}{5} \]

Thus the stationary distribution is \( v(0) = (\frac{3}{5}, \frac{2}{5}) \)

Another very important property of the stationary distribution:
Under certain benign conditions, \( v(n) \to^{\text{stationary distribution}} \) for any \( v(0) \).

i.e., whatever distribution you start with, as \( n \to \infty \), you will approach the stationary distribution.

Note: The precise conditions under which the convergence occurs is beyond the scope of this course. You can look it up in the book, if you wish. We'll explore some of the conditions as part of Homework 8.