Markov Chains

A stochastic process is a collection of random variables over a probability space. In this class, we look at discrete time: \(X_0, X_1, X_2, \ldots\)

- \(X_0\): state of process at time 0
- \(X_n\): state of process at time \(n\)

The random variables \(X_0, X_1, \ldots\) represent the states of the process. We call the set of possible states the state space and denote it as \(S\).

**Def (Markov Property)** A process \(X_0, X_1, \ldots\) obeys the Markov Property if for every possible sequence of values \(i_0, i_1, \ldots, i_n, i_{n+1}\):

\[
\Pr(X_{n+1} = i_{n+1} \mid X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \Pr(X_{n+1} = i_{n+1} \mid X_n = i_n)
\]

That is, for each \(n \geq 0\), the distribution of \(X_{n+1}\) given \(X_0, X_1, \ldots, X_n\) depends only on \(X_n\).

**Ex** Consider the process of flipping a coin that flips heads with probability \(p\) until we see two consecutive heads.

This process obeys the Markov Property, so it is a Markov chain.

- \(S = \{H, T, HH\}\)
- \(\Pr(T \mid H) = 1 - p = q\)
- \(\Pr(T \mid T) = 1 - p = q\)
- \(\Pr(H \mid T) = p\)
- \(\Pr(HH \mid H) = p\)
- \(\Pr(HH \mid HH) = 1\)

**Claim** For \(X_0, X_1, \ldots\) a Markov chain on \(S\) and \(i_0, i_1, \ldots, i_n\) a sequence of states visited,

\[
\Pr(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \Pr(X_0 = i_0) \cdot \Pr(X_1 = i_1 \mid X_0 = i_0) \cdot \Pr(X_2 = i_2 \mid X_0 = i_0, X_1 = i_1) \ldots \cdot \Pr(X_n = i_n \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1})
\]

**Ex** Consider the process of flipping a coin that flips heads with probability \(p\) until we see two consecutive heads.

- \(\Pr(X_0 = H) = p\)
- \(\Pr(X_0 = T) = q\)
- \(\Pr(H \rightarrow H \rightarrow H) = p \cdot q \cdot q \cdot p \cdot p\)
**Transition Matrix**

**Def** (Transition Matrix) The one-step transition matrix of a chain is the matrix $P$ such that

$$P(i, j) = P(X_1 = j | X_0 = i) \quad \text{(probability of i → j in one step)}$$

Note:

1. $P$ is a square matrix.
2. Each row of $P$ is a distribution:
   $$\sum_j P(i, j) = 1 \quad \text{for any } i$$

**Ex** Consider the process of flipping a coin that flips heads with probability $p$ until we see two consecutive heads.

$$P = \begin{bmatrix} T & H & HH \\ T & q & p \\ H & 0 & 0 \\ HH & 0 & 1 \end{bmatrix}$$

**Claim** The n-step transition matrix $P_n$ is given by $P_n(i, j) = P(X_n = i \mid X_0 = j)$ (probability of i → j in n steps).

Then

$$P_2(i, j) = P(X_2 = j \mid X_0 = i)$$

$$= \sum_{k \in \mathbb{S}} P(X_1 = k, X_2 = j \mid X_0 = i)$$

$$= \sum_{k \in \mathbb{S}} P(X_1 = k \mid X_0 = i) \cdot P(X_2 = j \mid X_1 = i, X_0 = i) = \sum_k P(i, k) \cdot P(k, j)$$

By induction, $P_n(i, j) = P^n(i, j)$, so $P_n = P^n$

**Ex** Consider the process of flipping a coin that flips heads with probability $p$ until we see two consecutive heads.

$$P = \begin{bmatrix} T & H & HH \\ T & q & p \\ H & 0 & 0 \\ HH & 0 & 1 \end{bmatrix} \quad P_2 = \begin{bmatrix} q & p & 0 \\ q & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad P^2 = \begin{bmatrix} q^2 & pq & q \cdot 0 \\ q^2 & pq & q \cdot 0 \\ 0 & 0 & 1 \end{bmatrix}$$
Distribution over Time

Let \( \pi_0 \) be the initial distribution over the state space written as a row vector:
\[
\pi_0(i) = \Pr(X_0 = i)
\]
For example, with \( S = \{1, 2, \ldots, N\} \),
\[
\pi_0 = [\Pr(X_0 = 1) \ \ \ Pr(X_0 = 2) \ \ \ldots \ \ \ Pr(X_0 = N)]
\]
And let \( \pi_n \) be the distribution over the state space after \( n \geq 0 \) steps.
For example, with \( S = \{1, 2, \ldots, N\} \),
\[
\pi_n = [\Pr(X_n = 1) \ \ \ Pr(X_n = 2) \ \ \ldots \ \ \ Pr(X_n = N)]
\]
Then
\[
\pi_n = \pi_0 \mathbf{P}^n = \pi_0 \mathbf{P}^n
\]

For \( i \in S \),
\[
\pi_n(i) = \sum_{k \in S} \Pr(X_0 = k) \cdot \Pr(X_n = i \mid X_0 = k)
\]
\[
= \sum_{k \in S} \pi_0(k) \cdot \mathbf{P}_n(k,i)
\]
\[
= (\pi_0 \mathbf{P}_n)(i)
\]

Note: To specify a Markov chain, you need
- \( S \), the state space
- \( \mathbf{P} \), the transition probabilities
- \( \pi_0 \), the initial distribution.
Hitting Time

Suppose you repeatedly flip a coin with probability $p$ of heads until you see two consecutive heads. What is the expected number of flips it will take?

The process is given by the Markov chain

Let $R$ be the number of flips it takes to see two consecutive heads.

We can find the expected time until we see HH by conditioning. Let $\beta(i)$ be the expected time until HH starting from state $i$, i.e., $\beta(i) = \mathbb{E}[R | X_0 = i]$. 

$$
\begin{align*}
\beta(HH) &= 0 \\
\beta(H) &= 1 + p \cdot \beta(HH) + q \cdot \beta(T) \\
\beta(T) &= 1 + p \cdot \beta(H) + q \cdot \beta(T) \\
\beta(S) &= 1 + p \cdot \beta(H) + q \cdot \beta(T)
\end{align*}
$$

First Step Equations

1. $\beta(H) = 1 + p \cdot 0 + q \cdot \beta(T) = 1 + q \cdot \beta(T)$
2. $\beta(T) - q \cdot \beta(T) = p \cdot \beta(T) = (1 - p) \cdot \beta(H)$, so $\beta(T) = \frac{1}{p} \cdot \beta(H)$
3. $\beta(S) = \beta(T)$

So $\mathbb{E}[R] = \frac{1}{p} + \frac{1}{p^2}$

Note: Let $X_0, X_1, \ldots$ be a finite Markov chain on state space $S$ with transition matrix $P$. Let $A \subseteq S$ and $\beta(i)$ be the expected time to reach a state in $A$ from state $i$. Then

$$
\begin{align*}
\beta(i) &= 0 \text{ if } i \in A \\
\beta(i) &= 1 + \sum_{j \in A} P(i, j) \cdot \beta(j)
\end{align*}
$$
Q. We repeatedly roll a six-sided die and sum the rolls modulo 3 as we go. What is the chance our sum hits 1 before it hits 2?

\( S = \{0, 1, 2\} \) the value of the sum.

\[
\begin{align*}
3, 6 & \quad \rightarrow \quad 1, 4 \\
2, 5 & \quad \rightarrow \quad 1, 4 \\
2, 6 & \quad \rightarrow \quad 3, 1 \\
3, 0 & \quad \rightarrow \quad 2, 6
\end{align*}
\]

All transitions have probability \( \frac{1}{6} = \frac{1}{3} \)

Leverage Markov Property: Let \( \alpha(i) = P(1 \text{ before } 2 \mid \text{ in state } i) \).

\[
\begin{align*}
\alpha(0) &= \frac{1}{3} \alpha(0) + \frac{1}{3} \alpha(1) + \frac{1}{3} \alpha(2) \\
\alpha(1) &= 1 \\
\alpha(2) &= 0
\end{align*}
\]

\[
\frac{1}{3} \alpha(0) = \frac{1}{3} \\
\alpha(0) = \frac{1}{2}
\]

Q. Consider a sequence of iid trials, each of which results in \( n \) mutually exclusive categories. On each trial, let the chance of category \( i \) be \( p_i > 0 \).

What is the chance category \( i \) appears before category \( j \)?

\( S = \{i, j, k\} \) the category, where \( k \) is any category other than \( i \) or \( j \).

For \( \alpha(i) = P(1 \text{ before } j \mid \text{ in state } i) \),

\[
\begin{align*}
\alpha(i) &= 1 \\
\alpha(j) &= 0 \\
\alpha(k) &= p_i \alpha(i) + p_j \alpha(j) + (1 - p_i - p_j) \alpha(k)
\end{align*}
\]

\[
\alpha(k) = \frac{p_i}{p_i + p_j}
\]

Note. Let \( X_0, X_1, \ldots \) be a finite Markov chain on state space \( S \) with transition matrix \( P \).

Let \( A, B, C, S \) be mutually exclusive and \( \alpha(i) \) be the probability of hitting \( A \) before \( B \) from state \( i \).

Then

\[
\begin{align*}
\alpha(i) &= 1 \text{ if } i \in A \\
\alpha(i) &= 0 \text{ if } i \in B \\
\alpha(i) &= \sum_{k \in S} P(i, k) \alpha(k) \text{ otherwise}
\end{align*}
\]
A. An ant is sitting at the corner of a cube. At each timestep, she traverses an edge uniformly at random. What is the expected time until she reaches the other end of the cube?

Rather than defining a state for each corner, define a state as her distance from the start. $S = \{0, 1, 2, 3\}$.

Then

\[
\begin{align*}
\beta(0) &= 1 + \beta(1) \\
\beta(1) &= 1 + \frac{1}{3} \beta(0) + \frac{2}{3} \beta(2) \\
\beta(2) &= 1 + \frac{2}{3} \beta(1) + \frac{1}{3} \beta(3) \\
\beta(3) &= 0
\end{align*}
\]

\[
\begin{align*}
\beta(1) &= 1 + \frac{1}{3} + \frac{1}{3} \beta(1) + \frac{2}{3} \beta(2) \\
\beta(1) &= 2 + \beta(2) \\
\beta(2) &= 1 + \frac{4}{3} + \frac{2}{3} \beta(2) + 0 \\
\beta(2) &= 3 + 1 = 7 \\
\beta(1) &= 2 + 7 = 9 \\
\beta(0) &= 10
\end{align*}
\]