Markov Chains

- Discrete time
- Discrete (finite or countable) state space $S$
- Process $\{X_n\}$
- Homogenous transition probabilities
- matrix $P = \{P(i, j); i, j \in S\}$

$P(i, j)$, the $(i, j)^{th}$ entry of the matrix $P$, represents the probability of moving to state $j$ given that the chain is currently in state $i$.

Markov Property:

$$P(X_{n+1} = i_{n+1}|X_n = i_n, \ldots, X_0 = i_0) = P(i_n, i_{n+1})$$

This means that the states of $X_{n-1} \ldots X_0$ don’t matter. The transition probabilities only depend on the current state of the process. So,

$$P(X_{n+1} = i_{n+1}|X_n = i_n, \ldots, X_0 = i_0) = P(X_{n+1} = i_{n+1}|X_n = i_n) = P(i_n, i_{n+1})$$

To calculate the probability of a path, multiply the desired transition probabilities:

$$P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n) = P(i_0, i_1) \cdot P(i_1, i_2) \cdots \cdot P(i_{n-1}, i_n)$$

Example: iid Sequence $\{X_n\}$, $P(X_n = j) = p(j)$, and $\sum_j p(j) = 1$.

$$P(i, j) = j.$$

Example: Random Walk. $S = \mathbb{Z}$ (integers), $X_n = i_0 + D_1 + D_2 + \ldots + D_n$, where $D_i$ are iid, and $P(D_i = j) = p(j)$.

$$P(i, j) = p(j - i).$$

Example: Same random walk, but stops at 0.

$$P(i, j) = \begin{cases} 
p(j - i) & \text{if } i \neq 0; \\
1 & \text{if } i = j = 0; \\
0 & \text{if } i = 0, j \neq 0. 
\end{cases}$$
Example: $Y_0, Y_1, Y_2 \ldots$ iid. $P(Y = j) = p(j)$. $X_n = \max(Y_0, \ldots, Y_n)$.

$$P(i, j) = \begin{cases} p(j) & \text{if } i \leq j; \\ \sum_{k \leq i} p(k) & \text{if } i = j; \\ 0 & \text{if } i > j. \end{cases}$$

**Matrix Properties:**

$P(i, \cdot) = [P(i, j); \ j \in S]$, the $i^{th}$ row of the matrix $P$, is a row vector. Each row of $P$ is a probability distribution on the state-space $S$, representing the probabilities of transitions out of state $i$. Each row should sum to 1. ($\sum_{j \in S} P(i, j) = 1 \ \forall i$)

The column vectors $[P(\cdot, j); j \in S]$ represent the probability of moving into state $j$.

Use the notation $P_i =$ probability for a Markov Chain that started in state $i$ ($X_0 = i$).

In our transition matrix notation, $P_i(X_1 = j) = P(i, j)$.

**$n$-step transition probabilities:**

Want to find $P_i(X_n = j)$ for $n = 1, 2, \ldots$

This is the probability that the Markov Chain is in state $j$ after $n$ steps, given that it started in state $i$. First, let $n = 2$.

$$P_i(X_2 = k) = \sum_{j \in S} P_i(X_1 = j, X_2 = k) = \sum_{j \in S} P(i, j)P(j, k).$$

This is simply matrix multiplication.

So, $P_i(X_2 = k) = P^2(i, k)$

We can generalize this fact for $n$-step probabilities, to get:

$$P_i(X_n = k) = P^n(i, k)$$

Where $P^n(i, k)$ is the $(i, k)^{th}$ entry of $P^n$, the transition matrix multiplied by itself $n$ times. This is a handy formula, but as $n$ gets large, $P^n$ gets increasingly difficult and time-consuming to compute. This motivates theory for large values of $n$.

Suppose we have $X_0 \sim \mu$, where $\mu$ is a probability distribution on $S$, so $P(X_0 = i) =$
\(\mu(i)\) for \(i \in S\). We can find \(n^{th}\) step probabilities by conditioning on \(X_0\).

\[
P(X_n = j) = \sum_{i \in S} P(X_0 = i) \cdot P(X_n = j | X_0 = i)
= \sum_{i \in S} \mu(i) P^n(i, j)
= (\mu P^n)_j
= \text{the } j^{th} \text{ entry of } \mu P^n.
\]

Here, we are regarding the probability distribution \(\mu\) on \(S\) as a vector indexed by \(i \in S\).

So, we have \(X_1 \sim \mu P, X_2 \sim \mu P^2, \ldots, X_n \sim \mu P^n\).

Note: To compute the distribution of \(X_n\) for a particular \(\mu\), it is not necessary to find \(P^n(i, j)\) for all \(i, j, n\). In fact, there are very few examples where \(P^n(i, j)\) can be computed explicitly. Often, \(P\) has certain special initial distributions \(\mu\) so that computing \(\mu P^n\) is fairly simple.

**Example:** Death and Immigration Process

\(X_n = \) the number of individuals in the population at time (or generation) \(n\).

\(S = \{0, 1, 2, \ldots\}\)

Idea: between times \(n\) and \(n + 1\), each of the \(X_n\) individuals dies with probability \(p\), and the survivors contribute to generation \(X_{n+1}\). Also, immigrants arrive each generation, following a Poisson(\(\lambda\)) distribution.

Theory: What is the transition matrix, \(P\), for this process? To find it, we condition on the number of survivors to obtain:

\[
P(i, j) = \sum_{k=0}^{i \wedge j} \binom{i}{j} (1 - p)^k (p)^{i-k} \cdot \frac{e^{-\lambda} \lambda^{i-k}}{(i-k)!}
\]

Here, \(i \wedge j = \min\{i, j\}\). This formula cannot be simplified in any useful way, but we can analyze the behavior for large \(n\) using knowledge of the Poisson/Binomial relationship. We start by considering a special initial distribution for \(X_0\).

Let \(X_0 \sim \text{Poisson}(\lambda_0)\). Then the survivors of \(X_0\) have a Poisson(\(\lambda_0 q\)) distribution, where \(q = 1 - p\). Since the number of immigrants each generation follows a Poisson(\(\lambda\)) distribution, independent of the number of people currently in the population, we have \(X_1 \sim \text{Poisson}(\lambda_0 q + \lambda)\). We can repeat this logic to find:

\[
X_2 \sim \text{Poisson}((\lambda_0 q + \lambda)q + \lambda) = \text{Poisson}(\lambda_0 q^2 + \lambda q + \lambda), \text{ and}
\]
\[ X_3 \sim \text{Poisson}(\lambda_0 q^2 + \lambda q + \lambda) = \text{Poisson}(\lambda_0 q^3 + \lambda q^2 + \lambda + \lambda) . \]

In general,

\[ X_n \sim \text{Poisson}(\lambda_0 q^n + \lambda \sum_{k=0}^{n-1} q^k) . \]

In this formula, \( \lambda_0 q^n \) represents the survivors from the initial population, \( \lambda q^{n-1} \) represents the survivors from the first immigration, and so on, until \( \lambda q \) represents the survivors from the previous immigration, and \( \lambda \) represents the immigrants in the current generation.

Now we’ll look at what happens as \( n \) gets large.

As we let \( n \to \infty \):

\[ \lambda_0 q^n \to 0, \quad \text{and} \quad \lambda_0 q^n + \lambda \sum_{k=0}^{n-1} q^k \to \frac{\lambda}{1-q} = \frac{\lambda}{p} . \]

So, no matter what \( \lambda_0 \), if \( X_0 \) has Poisson distribution with mean \( \lambda_0 \),

\[ \lim_{n \to \infty} P(X_n = k) = \frac{e^{-\nu} \nu^k}{k!}, \quad \text{where} \quad \nu = \frac{\lambda}{1-q} = \frac{\lambda}{p} . \]

It is easy enough to show that this is true no matter what the distribution of \( X_0 \). The particular choice of initial distribution for \( X_0 \) that is Poisson with mean \( \lambda_0 = \nu \) gives an invariant (also called stationary, or equilibrium, or steady-state) distribution of \( X_0 \). With \( \lambda_0 = \nu \), we find that each \( X_n \) will follow a Poisson(\( \nu \)) distribution. This initial distribution \( \mu(i) = \frac{e^{-\nu} \nu^i}{i!} \) is special because it has the property that

\[ \sum_{i \in S} \mu(i)P(i,j) = \mu(j) \text{ for all } j \in S. \]

Or, in matrix form, \( \mu P = \mu \). It can be shown that this \( \mu \) is the unique stationary probability distribution for this Markov Chain.