Module 4

1 Markov Processes and Markov Chains

Remember how we defined a stochastic process as a sequence of random variables $X_1, X_2, \ldots, X_t, \ldots,$ typically indexed by time. This process is called a Markov chain if

$$P\left(X_{t+1} = x_{t+1} | X_0 = x_0, X_1 = x_1, \ldots, X_t = x_t\right) = P\left(X_{t+1} = x_{t+1} | X_t = x_t\right),$$

i.e., it is called a Markov chain if the history of the chain before t does not add any information to the distribution fo X_{t+1} (note how $X_0, X_1, \ldots, X_{t-1}$ dropped out of the conditional probability above). We will illustrate the concept of a Markov chain (and their one-step transition probability matrices) through an example.

1.1 Transition Probability Matrices and Their Properties

Example 1: The Ehrenfest chain originated as a model for 2 cubicle volumes of air connected by a small hole. Mathematically, let's think of this as having two urns and a total of N balls. We pick 1 of N balls at random and move it to the other urn. Let X_n = the number of balls in the "left" urn after n^{th} draw. X_n has Markov property since it doesn't depend on distant past. In particular,

$$P\left(X_{n+1} = i+1 | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\right) = P(X_{n+1} = i+1 | X_n = i) = \frac{(N-i)}{N}, \quad \text{and} \tag{1}$$

$$P(X_{n+1} = i-1 | X_n = i) = \frac{i}{N} \tag{2}$$

These conditoinal probabilities make sense. The probability in (1), i.e., the probability that the left urn will have one more ball $(X_{n+1} = i+1)$ is simply the probability that a ball is drawn from the right urn $\frac{N-i}{N}$. And the probability that the left urn will have one less ball is the probability that a ball is drawn from the left urn $\frac{i}{N}$. With $P(i, i+1) = \frac{N-i}{N}$ and $P(i, i-1) = \frac{i}{N}$, we can construct a one-step transition matrix $\mathbf{P}$ assuming that $N = 5$.

$$\begin{aligned}
\mathbf{P} &= \begin{bmatrix} P(X_n = 0 \;\&\; X_{n+1} = 0) & P(X_n = 0 \;\&\; X_{n+1} = 1) & \cdots & \cdots & P(X_n = 0 \;\&\; X_{n+1} = 5) \\ P(X_n = 1 \;\&\; X_{n+1} = 0) & P(X_n = 1 \;\&\; X_{n+1} = 1) & \cdots & \cdots & P(X_n = 1 \;\&\; X_{n+1} = 5) \\ \vdots & & & & \vdots \\ P(X_n = 5 \;\&\; X_{n+1} = 0) & P(X_n = 5 \;\&\; X_{n+1} = 1) & \cdots & \cdots & P(X_n = 5 \;\&\; X_{n+1} = 5) \end{bmatrix} \\
&= \begin{bmatrix} 0 & \frac{5}{5} & 0 & 0 & 0 & 0 \\ \frac{1}{5} & 0 & \frac{4}{5} & 0 & 0 & 0 \\ 0 & \frac{2}{5} & 0 & \frac{3}{5} & 0 & 0 \\ 0 & 0 & 0 & \frac{4}{5} & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 0 & \frac{5}{5} & 0 \end{bmatrix}.
\end{aligned}$$

The way to read this matrix is that $P(i, j)$ = the probability that the left urn will have $j-1$ marbles in the $(n+1)^{\text{st}}$ draw given that it has $i-1$ marbles in the n^{th} draw. Once again, prove to yourself that this makes sense.

One-step transition matrices are typically used to describe discrete-spaced Markov chains, and they are typically written as

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} & p_{11} & \cdots & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots & \cdots \\ \vdots & & & \ddots & \\ \vdots & & & & \ddots \end{bmatrix}$$

where $p_{i,j} = P(X_{n+1} = j | X_n = i)$. Obviously, the properties of $p_{i,j}$ include (1) $0 \le p_{i,j} \le 1$, and (2) $\sum_j p_{i,j} = 1$ for all i. Below are two more examples.

Example 2: An electronics store sells a video game system. If at the end of the day, the number of units they have on hand is 1 or 0, they order enough new units so the total number is 5. Assume merchandise arrives before the store opens the next day, and let X_n = the number of units on hand at end of n^{th} day. Assume that 0,1,2, or 3 video games are purchased with probabilities .3, .4, .2, and .1, respectively. In this case,

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & .1 & .2 & .4 & .3 \\ 0 & 0 & .1 & .2 & .4 & .3 \\ .3 & .4 & .3 & 0 & 0 & 0 \\ .1 & .2 & .4 & .3 & 0 & 0 \\ 0 & .1 & .2 & .4 & .3 & 0 \\ 0 & 0 & .1 & .2 & .4 & .3 \end{bmatrix}.$$

Example 3: Let X_n be the weather on day n in Laurel, MD. Let's assume there are three possible states. 1 = snowy, 2 = rainy, and 3 = sunny. Let's also assume that weather in Laurel is a Markov chain (X_{n+1} only depends on X_n). Assume the Markov chain has the following transition probability matrix.

$$\mathbf{P} = \begin{pmatrix} .4 & .6 & 0 \\ .2 & .5 & .3 \\ .1 & .7 & .2 \end{pmatrix}.$$

We know $P(X_{n+1} = 2|X_n = 1)$, but what if we want to know $P(X_{n+5} = 2|X_n = 1)$? In other words, what can $\mathbf{P}$ tell us about a forecast five days in the future? What about three or two days in the future? If we wanted to calculate the probability that Thrusday is snowy given Tuesday is rainy, then a way to do it would be to calculate

$$P(X_2 = 1|X_0 = 2) = \sum_{k=1}^{3} P(X_2 = 1, X_1 = k|X_0 = 2) = \sum_{k=1}^{3} P(2,k)P(k,1).$$

This sum ultimately ends up being the $(2,1)^{\text{st}}$ element in the matrix $\mathbf{P}^2$. We generalize this calculation to the following Theorem.

Theorem 1 *The* $m-$*step transition probability* $P(X_{n+m} = j|X_n = i)$ *is the* $(i,j)^{\text{th}}$ *element of the* m^{th} *power of* $\mathbf{P}$.

So $P(X_2 = 1|X_0 = 2)$ can be found by taking the second power of $\mathbf{P}$, and $P(X_3 = 2|X_0 = 1)$ can be found b taking the third power of $\mathbf{P}$.

$$\mathbf{P}^3 = \begin{pmatrix} .238 & .564 & .198 \\ .221 & .563 & .216 \\ .215 & .570 & .215 \end{pmatrix}.$$

This can be generalized even further by calculating the probability of going from state i to state j in $n+m$ steps. This generalization is captured in the Chapman-Kolmogorov equation, which states

$$P\left(X_{n+m} = j|X_0 = i\right) = \sum_k P\left(X_{n+m} = j|X_n = k\right) P\left(X_n = k|X_0 = i\right).$$

Below is a proof of the Chapman-Kolmogorov Equation

Proof: The Law of Total Probability states that

$$P(X_{n+m} = j|X_o = i) = \sum_k P\left(X_{n+m} = j \ \& \ X_n = k|X_0 = i\right) \tag{3}$$

Using the laws of conditional probability, the right side of (3) becomes

$$\begin{aligned} &= \sum_k P\left(X_{n+m} = j \ \& \ X_n = k \ \& \ X_0 = i\right)/P(X_0 = i) \\ &= \sum_k \left[\frac{P\left(X_{n+m} = j \ \& \ X_n = k \ \& \ X_0 = i\right)}{P\left(X_n = k \ \& \ X_0 = i\right)} \times \frac{P\left(X_n = k \ \& \ X_0 = i\right)}{P\left(X_0 = i\right)}\right] \\ &= \sum_k P\left(X_{n+m} = j|X_n = k \ \& X_0 = i\right) \times P\left(X_n = k|X_o = i\right), \end{aligned}$$

and because this is a Markov chain, $P\left(X_{n+m}=j|X_n=k \;\&X_0=i\right)=P\left(X_{n+m}=j|X_n=k\right)$, making the above equation

$$=\sum_k P\left(X_{n+m}=j|X_n=k\right)\times P\left(X_n=k|X_o=i\right). \quad \blacksquare$$

The marginal probabilities of the states at time t can also be found with these transition probability matrices. Let $\boldsymbol{\pi}^(t)$ = the probability of the states at time t, i.e., assuming there are n states, let

$$\boldsymbol{\pi}^{(t)}=\left(P(X_t=1),P(X_t=2),\ldots,P(X_t=n)\right)^T.$$

These marginal probabilities can be calculated with the equation given below

$$\boldsymbol{\pi}^{(t)}=\boldsymbol{\pi}^{(0)}\mathbf{P}^t,$$

where $\boldsymbol{\pi}^{(0)}$ is the marginal probability of the states at time 0.

1.2 State Classification

In a discrete-space Markov chain, the states need to be classified to characterize the properties of the chain. Let's begin with some terminology.

1. If X is a discrete Markov chain with discrete state space and transition matrix $\mathbf{P}$, then if states i and j are such that $\mathbf{P}^t(i,j)>0$ for some $t>0$, then i "leads to" j (written $i\longrightarrow j$).
2. If $i\longrightarrow j$ and $j\longrightarrow i$, then i and j communicate with one another.
3. All states in an equivalence class communicate with one another, but not with any other state outside of the class. If there is only one equivalence class, then the Markov chain is irreducible. If there is a set A such that $\sum_{j\in A}P(i,j)=1$ for all $i\in A$, then A is a closed set, and if i is closed, it is an absorbing state.
4. A state to which the chain returns with probability 1 is called a recurrent state.
5. A state for which the expected time until recurrence is finite is called nonnull.
6. A Markov chain is irreducible if any state j can be reached by any state i in a finite number of steps for all i,j. In other words, $\forall\ i,j\ \ \exists\ m$ such that $P(X_{m+n}=i|X_n=j)>0$.
7. A Markov chain is periodic if it can visit certain portions of the state space only at certain regularly spaced intervals. Example: State j has period d if the probability of going from $j\longrightarrow j$ in n steps is 0 $\forall\ n$ not divisible by d.
8. If every state in a Markov chain has period 1, then chain is called a periodic.
9. A Markov chain is called ergodic if it is irreducible, aperiodic, and all states are nonnull and recurrent.

We illustrate how to classify these states with some examples.

Example 4: Consider the following Markov chain with transition probability matrix $\mathbf{P}$, where

$$\mathbf{P}=\begin{bmatrix} .3 & 0 & 0 & 0 & .7 & 0 & 0\\ .1 & .2 & .3 & .4 & 0 & 0 & 0\\ 0 & 0 & .5 & .5 & 0 & 0 & 0\\ 0 & 0 & 0 & .5 & 0 & .5 & 0\\ .6 & 0 & 0 & .4 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & .2 & .8\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Classify the states of the Markov chain.

- 2 is called "transient" state since the probability of returning state 2 n times is 0, i.e., after some point it never reappears in the Markov chain.
- State 3 is also transient. Once you leave it, you'll never return to it.
- The set $\{4,6,7\}$ and $\{1,5\}$ are recurrent and irreducible. They are irreduble because any state i can be reached by any state j in a finite number of steps. They are recurrent because if within any state of these sets, you're going to return to the state infinitely many times.

Example 5: Consider the following Markov chain with transition probability matrix $\mathbf{P}$, where

$$\mathbf{P} = \begin{bmatrix} .1 & 0 & 0 & .4 & .5 & 0 \\ .1 & .2 & .2 & 0 & .5 & 0 \\ 0 & .1 & .3 & 0 & 0 & .6 \\ .1 & 0 & 0 & .9 & 0 & 0 \\ 0 & 0 & 0 & .4 & 0 & .6 \\ 0 & 0 & 0 & 0 & .5 & .5 \end{bmatrix}.$$

In this case, the states $\{1,3\}$ are transient and the others are recurrent.

Recurrent states can also be identified by taking the limiting power of $\mathbf{P}$. Recall that

$$\mathbf{P}^t(i,j) = P\left(X_{n+t} = j | X_n = i\right).$$

It follows that $\lim_{t\to\infty} \mathbf{P}^t(i,j) = \pi_j$ where, recall, π_j is the probability that, in the long run, the Markov chain lands on state j. This result makes sense. $\mathbf{P}^t(i,j)$ is the probability that you go from state i to state j in t steps. If for all i, $\mathbf{P}^t(i,j) = \pi_j$, then π_j is the probability that if enough time passes, you'll land at state j. It turns out that if state j is recurrent, then $\pi_j > 0$. And π_j is called the "limiting distribution" of state j.

Let's generalize this to a Markov chain with k states. If a Markov chain has k states, $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)^T$ is called a vector of limiting distributions. And if the Markov chain it corresponds to is irreducible, aperiodic and has transition matrix $\mathbf{P}$, $\boldsymbol{\pi}$ is uniquely determined by the equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$.

Example 6: Recall the weather chain that was discussed earlier, where $\mathbf{P} = \begin{pmatrix} .4 & .6 & 0 \\ .2 & .5 & .3 \\ .1 & .7 & .2 \end{pmatrix}$. What is the probability that is rainy? In other words, what is π_2? NOte again hat we're not conditioning on anything in this case. We're not assuming it started from anywhere. We just want to know what the probability is that it's rainy. We could take $\lim_{t\to\infty} \mathbf{P}^t(i,j)$ and look at the second column, or we could set up and solve the following system of equations.

$$(\pi_1, \pi_2, \pi_3) = (\pi_1, \pi_2, \pi_3) \cdot \mathbf{P} = \begin{pmatrix} .4 & .6 & 0 \\ .2 & .5 & .3 \\ .1 & .7 & .2 \end{pmatrix}.$$

Solving this system of equations, we get $\pi_1 = \frac{19}{85}$, $\pi_2 = \frac{48}{85}$, and $\pi_3 = \frac{18}{85}$.

With the idea of $\boldsymbol{\pi}$, we can now define and discuss the concept of reversibility. A stationary stochastic process $\{X_1, X_2, \ldots, X_n\}$ is said to be "reversible" if for any positive integer n and for all values $t_1, t_2, \ldots, t_n$,.

$$\left(X_{t_1}, X_{t_2}, \ldots, X_{t_n}\right) \text{ has the same distribution as } \left(X_{-t_1}, X_{-t_2}, \ldots, X_{-t_n}\right).$$

This states that the distribution of random variables going forward is equivalent to the distribution of random variables going backwards. SInce the distribution of the chain is the same going forwards as it is backwards, the chain is called reversible. And it turns out that a Markov process is reversible iff for every pair of states (i,j), there exist positive numbers π_i and π_j such that

$$\pi_i p_{i,j} = \pi_j p_{j,i}.$$

Consider the left part of the equation above. The left part of the equation calculates the joint probability of states i and j generated by starting at i and going to j $(\pi_i p_{i,j})$. The right part of the equation calculates the joint probability of states i and j generated by starting at j and going to i. It is important to understand, however, that just because a Markov chain has stationary distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)^T$ does **not** imply that the chain is reversible. Recall the weather chain. Note that $\pi_1 \times p_{1,3} \neq \pi_3 \times p_{3,1}$. You can generalize this result: if $p(x,y) > 0$ but $p(y,x) = 0$ (intuitively: if you can go y given you're at x but you can't go to x given you're at y), then you don't have reversibility.

2 Gaussian Processes

The random/stochastic process $X_1, X_2, \ldots, X_n, \ldots$ is "Gaussian" if all its finite-dimensional distributions (distributions of $(X_1, X_2), (X_1, X_2, \ldots, X_5)$, etc.) are Gaussian. Specifically, if

$$(X_1, X_2, \ldots, X_n)^T \sim N(\boldsymbol{\mu}, \Sigma),$$

where $\boldsymbol{\mu} = E\left[(X_1, X_2, \ldots, X_n)^T\right]$ and $\Sigma_{i,j} = \text{Cov}(X_i, X_j)$. A perfect example of this is Brownian motion (or the Weiner Process). In the Weiner process, $\boldsymbol{\mu}_t = \mathbf{0}$, $\text{Var}(X_t) = t$ and $\text{Cov}(X_s, X_t) = s$ where $s < t$. Writing this out in detail, we get

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2 & 2 & & 2 \\ 1 & 2 & 3 & 3 & & 3 \\ 1 & 2 & 3 & 4 & & 4 \\ \vdots & & & & \ddots & \vdots \\ 1 & 2 & 3 & 4 & \cdots & n \end{bmatrix}\right).$$

A critical property of the Weiner process is that $X_t - X_s \sim N(0, t-s)$ when $t > s$. Think about this result. $X_t - X_s$ is the difference in the values of X between time t and s. The larger the time difference, the more variable the difference in values. This is proven below.

Proof: Let's first show that $\text{Cov}(X_t, X_s) = s$. $\text{Cov}(X_t, X_s) = E\left[(X_t - \mu_t)(X_s - \mu_s)\right] = E[X_t X_s] = E\left[(X_t - 0)(X_s - 0)\right]$. We know that

$$E(X_t X_s) = E\left[(X_s + (X_t - X_s))X_s\right] = E(X_s^2) + E\left[X_s(X_t - X_s)\right] = s + E(X_s)E(X_t - X_s) = s.$$

Now let's calculate $\text{Var}(X_t - X_s)$..

$$\text{Var}(X_t - X_s) = \text{Var}(X_t) + \text{Var}(X_s) - 2 \cdot \text{Cov}(X_t, X_s) = t + s - 2 \cdot s = t - s.$$