Lecture 11: Discrete Time Markov Chains

1 Introduction

We have seen that iid sequences are easiest discrete time processes. However, they don’t capture correlation well. Hence, we look at the discrete time stochastic processes of the form

\[ X_{n+1} = f(X_n, Z_{n+1}), \]

where \( \{Z_n : n \in \mathbb{N}\} \) is an iid sequence, independent of initial state \( X_0 \). If \( X_n \in E \) for all \( n \in \mathbb{N}_0 \), then \( E \) is called the state space of process \( X \). We consider a countable state space, and if \( X_n = i \in E \), then we say that the process \( X \) is in state \( i \) at time \( n \). For a countable set \( E \), a stochastic process \( \{X_n \in E, n \in \mathbb{N}_0\} \) is called a discrete time Markov chain (DTMC) if for all positive integers \( n \in \mathbb{N}_0 \) and all states \( i_0, i_1, \ldots, i_{n-1}, i, j \in E \), the process \( X \) satisfies the Markov property

\[ P\{X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = P\{X_{n+1} = j|X_n = i\}. \]

1.1 Homogeneous Markov chain

If for each \( n \in \mathbb{N}_0 \), we have \( p_{ij}(n) \triangleq P\{X_{n+1} = j|X_n = i\} = p_{ij} \). That is, when the transition probability does not depend on \( n \), the DTMC is homogeneous and the matrix \( P = \{p_{ij} : i, j \in E\} \) is called the transition matrix.

For all states \( i, j \in E \), if a non-negative matrix \( A \in \mathbb{R}^{E \times E} \) has the following property

\[ a_{ij} \geq 0, \quad \sum_{j \in E} a_{ij} \leq 1, \]

then it is called a sub-stochastic matrix. If the second property holds with equality, then it is called a stochastic matrix. If in addition, \( A^T \) is stochastic, then \( A \) is called doubly stochastic. Clearly, the transition matrix \( P \) is stochastic matrix.

1.2 Transition graph

A transition matrix \( P \) is sometimes represented by a directed graph \( G = (E, \{[i, j] \in E \times E : p_{ij} > 0\}) \). In addition, this graph has a weight \( p_{ij} \) on each edge \( e = [i, j] \).

2 Chapman Kolmogorov equations

We can define \( n \)-step transition probabilities for \( i, j \in E \) and \( m, n \in \mathbb{N} \)

\[ p^{(n)}_{ij} \triangleq P\{X_{n+m} = j|X_m = i\}. \]
It follows from the Markov property and law of total probability that 
\[ p^{(m+n)}_{ij} = \sum_{k \in E} p^{(m)}_{ik} p^{(n)}_{kj}. \]

We can write this result compactly in terms of transition probability matrix \( P \) as \( P^{(n)} = P^n \). Let \( v \in \mathbb{R}^E \) is a probability vector such that 
\[ v_n(i) = P\{X_n = i\}. \]
Then, we can write this vector \( v_n \) in terms of initial probability vector \( v_0 \) and the transition matrix \( P \) as 
\[ v_n = v_0 P^n. \]

2.1 Strong Markov property (SMP)

Let \( T \) be an integer valued stopping time with respect to the stochastic process \( X \) such that \( P\{T < \infty\} = 1 \). Then for all \( i_0, \ldots, i_{n-1}, \ldots, i, j \in E \), the process \( X \) satisfies the strong Markov property if 
\[ P\{X_{T+1} = j | X_T = i, \ldots, X_0 = i_0\} = P\{X_{T+1} = j | X_T = i\}. \]

Lemma 2.1. Markov chains satisfy the strong Markov property.

Proof. Let \( X \) be a Markov chain and \( A = \{X_T = i, \ldots, X_0 = i_0\} \). Then, we have 
\[ P\{X_{T+1} = j | A\} = \frac{\sum_{n \in \mathbb{N}_0} P\{X_{T+1} = j, A, T = n\} P(A)}{P(A)} = \sum_{n \in \mathbb{N}_0} p_{ij} P(A, T = n) = p_{ij}. \]

This equality follows from the fact that \( \{T = n\} \) is completely determined by \( \{X_0, \ldots, X_n\} \)

As an exercise, if we try to use the Markov property on arbitrary random variable \( T \), the SMP may not hold. For example, define a non-stopping time \( T \) for \( j \in E \)
\[ T = \inf\{n \in \mathbb{N}_0 : X_{n+1} = j\}. \]

In this case, we have 
\[ P\{X_{T+1} = j | X_T = i, \ldots, X_0 = i_0\} = 1\{p_{ij} > 0\} \neq P\{X_1 = j | X_0 = i\} = p_{ij}. \]

A useful application of the strong Markov property is as follows. Let \( i_0 \in E \) be a fixed state and \( \tau_0 = 0 \) Let \( \tau_n \) denote the stopping times at which the Markov chain visits \( i_0 \) for the \( n \)th time. That is, 
\[ \tau_n = \inf\{n > \tau_{n-1} : X_n = i_0\}. \]
Then \( \{X_{\tau_n + m} : m \in \mathbb{N}_0\} \) is a stochastic replica of \( \{X_m : m \in \mathbb{N}_0\} \) with \( X_0 = i_0 \) and can be studied as a regenerative process.

3 Communicating classes

State \( j \in E \) is said to be accessible from state \( i \in E \) if \( p^{(n)}_{ij} > 0 \) for some \( n \in \mathbb{N} \), and denoted by \( i \rightarrow j \). If two states \( i, j \in E \) are accessible to each other, they are said to communicate with each other, denoted by \( i \leftrightarrow j \). A set of states that communicate are called a communicating class.
Proposition 3.1. Communication is an equivalence relation.

Proof. Reflexivity and Symmetry are obvious. For transitivity, suppose \( i \leftrightarrow j \) and \( j \leftrightarrow k \). Suppose \( p_{ij}^{(m)} > 0 \) and \( p_{jk}^{(n)} > 0 \). Then by Chapman Kolmogorov, we have

\[
p_{ik}^{(m+n)} = \sum_{l \in \mathbb{N}_0} p_{il}^{(m)} p_{lk}^{(n)} \geq p_{ij}^{(m)} p_{jk}^{(n)} > 0
\]

Hence transitivity is assured.

3.1 Irreducibility and periodicity

A consequence of the previous result is that communicating classes are disjoint or identical. A Markov chain with a single class is called an irreducible Markov chain. The period of state \( i \) is defined as

\[
d(i) = \gcd\{n \in \mathbb{N}_0 : p_{ii}^{(n)} > 0\}
\]

If the period is 1, we say the state is aperiodic.

Proposition 3.2. If \( i \leftrightarrow j \), then \( d(i) = d(j) \). Basically, periodicity is a class property.

Proof. Let \( m \) and \( n \) be such that \( p_{ij}^{(m)} p_{ji}^{(n)} > 0 \). Suppose \( p_{ii}^{(s)} > 0 \). Then

\[
p_{jj}^{(n+m)} \geq p_{jj}^{(n)} p_{ij}^{(m)} > 0
\]

\[
p_{jj}^{(n+s+m)} \geq p_{jj}^{(n)} p_{ij}^{(m)} > 0
\]

Hence \( d(j)|n + m \) and \( d(j)|n + s + m \) which implies \( d(j)|s \). Hence \( d(j)|d(i) \). By symmetrical arguments, we get \( d(i) = d(j) \).

3.2 Transient and recurrent states

Let \( f_{ij}^{(n)} \) denote the probability that starting from state \( i \), the first transition into state \( j \) happens at time \( n \). Then let

\[
f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}
\]

Here \( f_{ij} \) would therefore denote the probability of ever entering state \( j \) given that we start at state \( i \). State \( j \) is said to be transient if \( f_{jj} < 1 \) and recurrent if \( f_{jj} = 1 \).

Proposition 3.3. The total number of visits to a state \( j \in E \) is denoted by \( N_j = \sum_{n \in \mathbb{N}_0} 1\{X_n = j\} \). Then, for each \( m \in \mathbb{N} \) and \( i \neq j \), we have

\[
P_j\{N_j = m\} = f_{jj}^{m-1}(1 - f_{jj}), \quad m \in \mathbb{N}
\]

\[
P_i\{N_j = m\} = \begin{cases} 1 - f_{ij} & m = 0, \\ f_{ij} f_{jj}^{m-1}(1 - f_{jj}) & m \in \mathbb{N}. \end{cases}
\]
Proof. For each \( k \in \mathbb{N} \), the time of the \( k \)th visit to the state \( j \) is a stopping time. From strong Markov property, the next return to state \( j \) is independent of the past. Hence, each return to state \( j \) is an iid Bernoulli random variable with probability \( f_{jj} \). It follows that the number of visits \( j \) is the time for first failure to return. Conditioned on \( X_0 = j \), the distribution of \( N_j \) is geometric random variable with success probability \( 1 - f_{jj} \).

Conditioned on \( X_0 = i \), the stopping time of first visit to \( j \) is a Bernoulli random variable with probability \( f_{ij} \). Hence, the second result follows.

Corollary 3.4. For a Markov chain \( X \), \( P_j \{N_j < \infty\} = 1 \{f_{jj} < 1\} \).

Proof. We can write the event \( \{N_j < \infty\} \) as disjoint union of events \( \{N_j = n\} \), to get

\[
P_j \{N_j \in \mathbb{N}\} = \sum_{n \in \mathbb{N}} P_j \{N_j = n\} = 1 \{f_{jj} < 1\}.
\]

In particular, this corollary implies the following

1. A transient state is visited a finite amount of times almost surely.
2. A recurrent state is visited infinitely often almost surely.
3. In a finite state Markov chain, not all states may be transient.

Proposition 3.5. A state \( j \) is recurrent iff \( \sum_{k \in \mathbb{N}} p_{jj}^{(k)} = \infty \).

Proof. For any state \( i \in E \), we can write

\[
p_{ii}^{(k)} = P_i \{X_k = i\} = E_i 1 \{X_k = i\}.
\]

Using monotone convergence theorem to exchange expectation and summation, we obtain

\[
\sum_{k \in \mathbb{N}} p_{ii}^{(k)} = E_i \sum_{k \in \mathbb{N}} 1 \{X_k = i\} = E_i N_i.
\]

Thus, \( \sum_{k \in \mathbb{N}} p_{ii}^{(k)} \) represents the expected number of returns \( E_i N_i \) to a state \( i \) starting from state \( i \), which we know to be finite if the state is transient and infinite if the state is recurrent.

Proposition 3.6. Transience and recurrence are class properties.

Proof. Let us start with proving recurrence is a class property. Let \( i \) be a recurrent state and let \( i \leftrightarrow j \). Hence there exist some \( m,n > 0 \), such that \( p_{ij}^{(m)} > 0 \) and \( p_{ji}^{(n)} > 0 \). As a consequence of the recurrence, \( \sum_{s \in \mathbb{N}_0} p_{ii}^{(s)} = \infty \). It follows that \( j \) is recurrent by observing

\[
\sum_{s \in \mathbb{N}_0} p_{jj}^{(m+n+s)} \geq \sum_{s \in \mathbb{N}_0} p_{ji}^{(n)} p_{ii}^{(s)} p_{ij}^{(m)} = \infty.
\]

Now, if \( i \) were transient instead, we conclude that \( j \) is also transient by the following observation

\[
\sum_{s \in \mathbb{N}_0} p_{jj}^{(s)} \leq \sum_{s \in \mathbb{N}_0} p_{ij}^{(m+n+s)} p_{ji}^{(n)} p_{ij}^{(m)} < \infty.
\]

Corollary 3.7. If \( j \) is recurrent, then for any state \( i \) such that \( i \leftrightarrow j \), \( f_{ij} = 1 \).