Lecture 4: Discrete Time Markov Chains, II: State Classification, Stationary Distributions, & Limiting Behavior

John MacLaren Walsh, Ph.D.

February 9, 2014

1 References

Markov chains are a widely taught subject, and hence there are a wide variety of texts. A few that cover this material at a level appropriate for this course are:


The exposition below follows the first reference (which the bookstore has copies of). The section numbers below are reading from this text.

2 State Classification and Decomposition, Cont’d

Recall from last week the notion of a communicating class (based on the notion of accessibility/reachability and communication between states), persistent vs. transient states, and null persistent vs. non-null persistent states.

One final classification of a state is whether or not it is periodic or aperiodic. The period $T_j$ of a state $j$ is

$$T_j = \gcd \left( \left\{ n \in \mathbb{Z}_{\geq 0} | p_{j,j}^{(n)} > 0 \right\} \right) \tag{1}$$

where $\gcd$ is the greatest common divisor and $\mathbb{Z}_{\geq 0}$ is the set of non-negative integers. A state $j$ is said to be aperiodic if its period $T_j = 1$, otherwise a state is said to be periodic with period $T_j$.

States that are persistent, aperiodic, and non-null are called ergodic states.

It follows from the fact that the relation of a state $i$ and a state $j$ communicating with one another being an equivalence relation, that the states $S$ of a Markov chain can be partitioned $S = C_1 \cup \cdots \cup C_k \cup T$ up into the following disjoint sets: 1) a collection of disjoint communicating classes $C_1, \ldots, C_k \cup T$, and, potentially, 2) a collection of states $T$ which communicate with no-one (including that they don’t communicate with themselves). This latter group of states $T$ is necessarily transient, since their lack of communication with themselves means that their recurrence probability $f_{jj} = 0$.

Finally, note that if we allowed the notion of accessibility to be instead that $p_{i,j}^{(n)} > 0$ for any $n \geq 0$ (instead of $n \geq 1$), as is done in some texts, then any state communicates with itself since $p_{jj}^{(0)} = 1$ for all $j$. Hence, in this case, the states $S$ can be partitioned up into a collection of disjoint communicating classes $S = C_1, \ldots, C_k$.

Note that a chain with only one communicating class $S = C$ in this decomposition is said to be irreducible.
One of the key reasons of being interested in communicating classes is that all states in a communicating class are necessarily of the same type. Namely, they are all either transient or persistent, and, if they are persistent, they are all either null persistent or non-null persistent. Furthermore, all the states in a communicating class are collectively either periodic with the same period or aperiodic.

3 Stationary Distributions

A stationary distribution \( \pi_j, j \in S \), for a Markov chain with states \( S \), is a distribution that satisfies the equations

\[
\pi_j = \sum_{i \in S} \pi_i p_{i,j}^{(n)}, \quad \pi_j \geq 0, j \in S, \quad \sum_{j \in S} \pi_j = 1
\]

when such a collection \( \{\pi_j | j \in S\} \) exists. There exist Markov chains which do not have any stationary distributions, for instance, a irreducible chain with all null persistent states has no stationary distribution.

When the Markov chain is a finite state Markov chain, we can stack the \( \pi_j \)s into a vector \( \pi = [\pi_1, \ldots, \pi_{|S|}] \), and using the state transition matrix \( P = [p_{i,j}^{(1)}] \), we can rewrite (2) as

\[
\pi = \pi P, \quad \pi \geq 0, \quad \pi 1^T = 1
\]

where \( 1 = [1, \ldots, 1] \). Equivalently, in the finite state case there must exists an eigenvector \( \pi^T \) of the matrix \( P^T \) with eigenvalue 1 (so that \( \pi^T = P^T \pi^T \)) with \( \pi \) having all elements of the same sign (either all \( \leq 0 \) or all \( \geq 0 \)), so that we can select the scaling \( \alpha \) of this eigenvector with \( \alpha \pi \geq 0 \) and \( \alpha \pi 1^T = 1 \).

Also note that, when a chain has a stationary distribution, it may have more than one stationary distribution. Furthermore, if \( \{\pi_i^{(0)} | j \in S\} \) and \( \{\pi_i^{(1)} | j \in S\} \) are two stationary distributions for a discrete time Markov chain, then so is \( \{\pi_i^{(j)} | j \in S\} \) where \( \pi_i^{(j)} = \lambda \pi_i^{(0)} + (1 - \lambda)\pi_i^{(1)}, \forall j \in S \) for any \( \lambda \in [0, 1] \). In particular, for a finite state Markov chain, the set of stationary distributions is a polyhedron described by (3).

As can be seen by applying the Chapman Kolmogorov equations and the definition of stationarity, stationary distributions have the property that if the chain is started in with this distribution, with \( \mathbb{P}[X_0 = j] = \pi_j, \forall j \in S \), then it will remain in it for all time, i.e. then \( \mathbb{P}[X_n = j] = \pi_j, \forall j \in S \) for all \( n \geq 0 \).

4 Limiting Behavior and Limiting Distributions

In this section we would like to determine the behavior of \( p_{i,j}^{(n)} \) for \( n \to \infty \). It turns out these can be calculated from the following previously introduced quantities: the first passage probabilities \( f_{i,j} \), the mean recurrence times \( \mu_j \), and the period of a state \( T_j \), where

\[
f_{i,j}^{(n)} = \mathbb{P}[X[n] = j, X[n-k] \neq j \forall k \in \{1, \ldots, n-1\} | X[0] = i], \quad n \geq 1 \quad f_{i,j} = \sum_{n=1}^{\infty} f_{i,j}^{(n)}, \quad \mu_k = \sum_{n=1}^{\infty} n f_{i,j}^{(n)}
\]

and \( T_j \) is defined in (1).

If \( j \) is a non-null persistent aperiodic state, then for all \( i \in S \)

\[
\lim_{n \to \infty} p_{i,j}^{(n)} = \frac{f_{i,j}}{\mu_j} \quad \text{and, hence} \quad \lim_{n \to \infty} p_{j,i}^{(n)} = \frac{1}{\mu_j}
\]

If \( j \) is a null persistent aperiodic state, then for all \( i \in S \)

\[
\lim_{n \to \infty} p_{i,j}^{(n)} = 0
\]
If a chain is irreducible, aperiodic, and non-null persistent, we say that the chain is ergodic (for all of its states are ergodic). Applying (5) and the fact that all states communicate with one another and are non-null persistent, we observe that for all \( i \in S \)

\[
\lim_{n \to \infty} p_{i,j}^{(n)} = \frac{1}{\mu_j} \tag{7}
\]

so that the limit does not depend on the initial state \( i \). Hence, no matter what the initial distribution \( \{P[X[0] = j]|j \in S\} \) is, the state will asymptotically reach the limiting distribution \( \pi_j = \frac{1}{\mu_j}, j \in S \). Furthermore, such a irreducible, aperiodic, and non-null persistent chain will have a unique stationary distribution \( \pi_j = \frac{1}{\mu_j} \).

In fact, if an irreducible and aperiodic chain has a stationary distribution \( \pi_j \), it must be unique, and the chain must be ergodic (i.e. all of its states must be non-null persistent), so that this unique stationary distribution is also a limiting distribution.

More generally, if a discrete time homogeneous Markov chain has a limiting distribution, so that

\[
\tilde{\pi}_j = \lim_{n \to \infty} p_{i,j}^{(n)} \tag{8}
\]

exists for all \( i, j \) and is independent of \( i \) for each \( j \), and the chain also has a stationary distribution \( \{\pi_j\} \), then \( \tilde{\pi}_j = \pi_j \), and this stationary distribution must be unique.

Furthermore, if a finite-state Markov chain has a limiting distribution, then the limiting distribution must be a unique stationary distribution for the chain.

Thus, the existence of more than one stationary distribution is enough to show that the chain does not have a limiting distribution.

While we discussed these facts about limiting behavior during lecture on Feb. 5, we will cover their proofs during lecture on Feb. 12.