1. General Theory of Markov Chains

We have already discussed the standard random walk on the integers $\mathbb{Z}$. A Markov Chain can be viewed as a generalization of this. We shall only consider in this Chapter Markov chains on a countable (finite or infinite) state space $F$. In the finite case we can identify $F$ as the set of positive integers $F = \{1, 2, ..., m\}$ for some $m = |F|$, and in the infinite case as all positive integers $F = \{n \in \mathbb{Z} : n \geq 1\}$. The chain is determined by a set of transition probabilities $p_t(i, j)$, $i, j \in F$, $t = 0, 1, 2, ...$, where $t$ denotes discrete time and $p_t(i, j)$ is the probability of moving from state $i$ to state $j$ at time $t$. Thus

(1.1) \[ p_t(i, j) \geq 0, \; i, j \in F; \quad \sum_{j \in F} p_t(i, j) = 1, \; i \in F. \]

To define a probability space associated with these transition probabilities we need to set a distribution function $\pi : F \rightarrow \mathbb{R}$ for the initial state of the chain. Thus $\pi(\cdot)$ satisfies

(1.2) \[ \pi(j) \geq 0, \; j \in F; \quad \sum_{j \in F} \pi(j) = 1. \]

The probability space $\Omega$ is given by $\Omega = F^\infty$, which is the infinite product of the state space $F$, and the $\sigma-$field $\mathcal{F}$ is the Borel field generated by finite dimensional rectangles as usual. Random variables $X_n : \Omega \rightarrow F$, $n = 0, 1, 2, ...$, are defined by $X_n(\omega) = \omega_n$, $n = 0, 1, ...$, where $\omega = (\omega_0, \omega_1, ...) \in \Omega$. The probability measure $P$ depends on $\pi(\cdot)$, so we shall denote it by $P_\pi$. The measure $P_\pi$ is determined once we know the p.d.f. for every set of variables $(X_0, ..., X_m)$, $m \geq 0$. This is given by the formula

(1.3) \[ P(X_0 = j_0, X_1 = j_1, ..., X_m = j_m) = \pi(j_0) \prod_{t=0}^{m-1} p_t(j_t, j_{t+1}). \]

Note that the p.d.f for $X_0$ is $\pi(\cdot)$, and that for any subset $A \subset F$,

(1.4) \[ P(X_{t+1} \in A \mid X_t, X_{t-1}, .., X_0) = P(X_{t+1} \in A \mid X_t) \quad \text{with probability 1}. \]

The so called Markov property or no memory property (1.4) characterizes a Markov chain, and can be an alternative starting point for the theory of Markov chains. Note the similarity of (1.4) to the definition of a Martingale.

One big advantage of studying Markov chains is that a technique is available to compute many expectation values. This is the so called backward Kolmogorov equation. Thus let $f : F \rightarrow \mathbb{R}$ be a function and suppose we want to evaluate the expectation

(1.5) \[ E \left[ f(X_T) \mid X_0 = i \right], \; i \in F. \]
Then (1.5) is given by
\[ u(i,0) = \sum_{j \in F} p_t(i,j)u(j,t+1) , \quad i \in F, \quad t = 0, \ldots, T \]
where \( u(i,t) , \quad i \in F, \quad t = 0, \ldots, T \), is a solution to the terminal value problem,
\[ u(i,t) = \sum_{j \in F} p_t(i,j)u(j,t+1) , \quad i \in F, \quad t < T; \quad u(i,T) = f(i) , \quad i \in F. \]

Note that the equation (1.6) is to be solved backwards in time from time \( T \) to time 0.

We turn now to Markov chains where the transition probabilities are independent of time, whence \( p_t(i,j) = p(i,j) , \quad t = 0, 1, \ldots, i,j \in F \). An important property of these Markov chains is the strong Markov property. Observe from (1.4) that
\[ P(X_{t+1} \in A \mid X_t, X_{t-1}, \ldots, X_0) = P(X_{t+1} \in A \mid X_t) = P(X_1 \in A \mid X_0) \quad \text{with probability 1}. \]
The strong Markov property is a random version of (1.7). Thus let \( \tau : \Omega \to \mathbb{Z} \) be a nonnegative measurable function which has the property that
\[ \{ \omega \in \Omega : \tau(\omega) = T \} \in \mathcal{F}(X_0, X_1, \ldots, X_T), \quad T = 0, 1, \ldots. \]
Such a function is called a stopping time.

**Proposition 1.1** (Strong Markov Property). Consider the time independent Markov chain \( (X_0, X_1, \ldots) \) with initial distribution \( \pi(\cdot) \) and let \( \tau(\cdot) \) be a stopping time for this chain. Then the sequence of variables \( (X_\tau, X_{\tau+1}, X_{\tau+2}, \ldots) \) has the same distribution as the sequence \( (X_0, X_1, \ldots) \) with initial distribution \( \pi_{\tau}(\cdot) \), where \( \pi_{\tau}(\cdot) \) is the pdf of \( X_\tau \) under \( \pi(\cdot) \).

**Proof.** We illustrate the proof with the simplest case, which is then easy to generalize. Thus
\[ P(X_\tau = j_0, X_{\tau+1} = j_1, X_{\tau+2} = j_2) = \sum_{m=0}^{\infty} P(\tau = m, X_m = j_0, X_{m+1} = j_1, X_{m+2} = j_2). \]

Now from (1.8) it follows that
\[ P(\tau = m, X_m = j_0, X_{m+1} = j_1, X_{m+2} = j_2) = P(\tau = m, X_m = j_0)p(j_0, j_1)p(j_1, j_2). \]
Since we have that
\[ \sum_{m=0}^{\infty} P(\tau = m, X_m = j_0) = \pi_{\tau}(j_0), \]
we conclude that
\[ P(X_\tau = j_0, X_{\tau+1} = j_1, X_{\tau+2} = j_2) = \pi_{\tau}(j_0)p(j_0, j_1)p(j_1, j_2), \]
whence the result holds. \( \Box \)

**Definition 1.** An initial distribution \( \pi(\cdot) \) for the time independent Markov chain is said to be stationary if it satisfies the equation
\[ \sum_{i \in F} \pi(i)p(i,j) = \pi(j) , \quad j \in F. \]
Evidently (1.13) implies that if \( X_0 \) has distribution \( \pi(\cdot) \) then \( X_1 \) also has distribution \( \pi(\cdot) \). One can easily see further that the entire sequence \((X_0, X_1, \ldots)\) is stationary in the sense of Definition 4 of Chapter II page 18. An obvious question to ask then is does a Markov chain have a stationary distribution, and if so is it unique? We can give a satisfactory answer to the uniqueness question by introducing the notion of indecomposability of the chain. We say the chain is decomposable if there exists disjoint subsets \( A_0, A_1 \) of \( F \) such that
\[
\sum_{j \in A_k} p(i, j) = 1, \quad i \in A_k, \ k = 0, 1.
\]
Thus (1.14) states that the chain allows no communication between the subsets \( A_0 \) and \( A_1 \) of \( F \). Hence we may reduce the original chain to independent chains on the reduced state spaces \( A_0, A_1 \). A chain which is not decomposable is called indecomposable.

**Proposition 1.2.** Suppose the Markov chain is indecomposable and a stationary distribution \( \pi(\cdot) \) for it exists. Then \( \pi(\cdot) \) is unique and the associated stationary sequence \((X_0, X_1, \ldots)\) is ergodic.

**Proof.** We first show that for stationary \( \pi(\cdot) \) the corresponding stationary sequence \((X_0, X_1, \ldots)\) is ergodic. Thus let \( T \) be the shift operator on \( \Omega \), \( P = P_T \) be the probability measure on \( \Omega \), and \( C \in \mathcal{F} \) be an invariant set. We define a function \( \phi : F \rightarrow \mathbb{R} \) by \( \phi(j) = P(C \mid X_0 = j), \ j \in F \). Evidently \( \phi(j) \) is only uniquely defined if \( \pi(j) > 0 \). Otherwise we can take \( \phi(j) \) to be anything we want. Note now that
\[
\phi(X_0) = P(C \mid X_0) \quad \text{with probability 1},
\]
and
\[
P(C \mid X_0) = E[P(C \mid X_1, X_0) \mid X_0] \quad \text{with probability 1}.
\]
By the invariance of \( C \) we have that \( C \in \mathcal{F}(X_1, X_2, \ldots) \), whence it follows from the Markov property (1.4) that
\[
P(C \mid X_1, X_0) = P(C \mid X_1) = \phi(X_1) \quad \text{with probability 1}.
\]
Evidently (1.16), (1.17) imply that
\[
\sum_{j \in F} p(i, j) \phi(j) = \phi(i), \ i \in F \quad \text{with } \pi(i) > 0.
\]
Note from (1.13) that if \( \pi(i) > 0 \) and \( \pi(j) = 0 \) then \( p(i, j) = 0 \), whence the values of \( \phi(j) \) for \( \pi(j) = 0 \) do not enter on the LHS of (1.18).

Observe next that for any set \( C \in \mathcal{F}(X_0, X_1, \ldots) \) one has
\[
\lim_{n \to \infty} P(C \mid X_n, X_{n-1}, \ldots, X_0) = \chi_C \quad \text{with probability 1}.
\]
This intuitively obvious result follows from the Martingale convergence theorem. In fact define a sequence of random variables \( Z_n, \ n = 0, 1, 2, \ldots, \) by \( Z_n = P(C \mid X_n, X_{n-1}, \ldots, X_0) \). Then from the definition of conditional expectation one has that
\[
E[Z_n \mid X_{n-1}, \ldots, X_0] = Z_{n-1}, \ n = 1, 2, \ldots, \quad \text{with probability 1},
\]
which implies that
\[
E[Z_n \mid Z_{n-1}, \ldots, Z_0] = Z_{n-1}, \ n = 1, 2, \ldots, \quad \text{with probability 1}.
\]
Hence \( \lim_{n \to \infty} Z_n \) exists with probability 1, and it is easy to see that the limit must be \( \chi_C \).

Observe now that for the invariant set \( C \), one can see as in (1.17) that \( P(C \mid X_n, X_{n-1}, \ldots, X_0) = \phi(X_n) \), whence we have that

\[
\lim_{n \to \infty} \phi(X_n) = \chi_C \quad \text{with probability 1.} \tag{1.22}
\]

Since we may assume \( 0 \leq \phi(j) \leq 1 \), \( j \in E \), and we have from (1.22) that

\[
E[\phi(X_0)\{1-\phi(X_0)\}] = \lim_{n \to \infty} E[\phi(X_n)\{1-\phi(X_n)\}] = 0,
\]

we conclude that \( \phi(j) \) can only take the values 0 or 1 for \( j \in E \) with \( \pi(j) > 0 \). Now let us define sets \( A_1, A_0 \) by

\[
A_1 = \{i \in E : \pi(i) > 0, \phi(i) = 1\}, \quad A_0 = \{i \in E : \pi(i) > 0, \phi(i) = 0\}.
\]

It follows from (1.18) that

\[
\sum_{j \in A_1} p(i, j) = 1, \quad i \in A_1.
\]

Note that in (1.25) we are again using the fact that if \( \pi(i) > 0 \) and \( \pi(j) = 0 \) then \( p(i, j) = 0 \). Replacing the function \( \phi(\cdot) \) by \( 1 - \phi(\cdot) \) in (1.18), we see also from (1.25) that

\[
\sum_{j \in A_0} p(i, j) = 1, \quad i \in A_0.
\]

Evidently (1.25), (1.26) and the indecomposability assumption imply that either \( A_0 \) or \( A_1 \) is empty, so let us assume that \( A_0 \) is empty. It follows then from (1.15) that

\[
P(C) = E[\phi(X_0)] = 1.
\]

We similarly see that \( P(C) = 0 \) if \( A_1 \) is empty. We have proved the ergodicity of the stationary sequence.

We can prove the uniqueness of \( \pi(\cdot) \) by using Proposition 4.3 of Chapter II. Thus let \( \pi_1(\cdot) \) and \( \pi_2(\cdot) \) be two invariant distributions with corresponding probability measures \( P_1, P_2 \). From Proposition 4.3 of Chapter II it follows that \( P_1 \) and \( P_2 \) are mutually singular, so there exists \( E \in \mathcal{F} \) such that \( P_1(E) = 1 \), \( P_2(E) = 0 \). Now let \( C \in \mathcal{F} \) be the set

\[
C = \cup_{n=0}^{\infty} T^{-n} E,
\]

whence it follows that \( T^{-1}C \subset C \). Evidently \( P_1(C) = 1 \), and \( P_2(C) = 0 \) by the measure preserving property of \( P_2 \). We consider now the distribution \( \pi(\cdot) \) defined by \( \pi(\cdot) = [\pi_1(\cdot) + \pi_2(\cdot)]/2 \), which evidently also satisfies (1.13). The corresponding probability measure is \( P = [P_1 + P_2]/2 \). Hence \( T \) is measure preserving and ergodic under \( P \). Since \( T^{-1}C \subset C \), it follows from the measure preserving property of \( T \) that

\[
P([C - T^{-1}C] \cup [T^{-1}C - C]) = 0.
\]

Lemma 3.2 of Chapter II implies then that \( P(C) = 0 \) or \( P(C) = 1 \). However we easily see that \( P(C) = [P_1(C) + P_2(C)]/2 = 1/2 \), which is a contradiction. \( \square \)
Definition 2. A time independent Markov chain is said to be irreducible if all states communicate. That is for any $i, j \in F$,

\[(1.30) \quad P(X_n = j \mid X_0 = i) > 0 \quad \text{for some } n \geq 1 \text{ which can depend on } (i, j).\]

Note that irreducibility implies indecomposability.

Lemma 1.1. Suppose a Markov chain is time independent irreducible and a stationary distribution $\pi(\cdot)$ for it exists. Then $\pi(j) > 0$ for all $j \in F$.

Proof. We use equation (1.13) and choose $i \in F$ such that $\pi(i) > 0$. Evidently if (1.30) holds for $n = 1$ then $\pi(j) > 0$. Observe now that (1.13) implies that

\[(1.31) \quad \sum_{i_1, i_2 \in F} \pi(i_1)p(i_1, i_2)p(i_2, j) = \pi(j), \quad j \in F.\]

Since

\[(1.32) \quad P(X_2 = j \mid X_0 = i) = \sum_{i_2 \in F} p(i, i_2)p(i_2, j),\]

we conclude that if (1.30) holds for $n = 2$ then $\pi(j) > 0$. Evidently we can generalize this argument to any $n \geq 1$. \qed

We can apply Proposition 5.2 of Chapter II to obtain a relation between recurrence for the Markov chain and the invariant measure.

Proposition 1.3. Suppose a time independent Markov chain is irreducible and has stationary distribution $\pi(\cdot)$. Then for all $i \in F$ one has $P(X_n = i$ infinitely often $\mid X_0 = i) = 1$. Furthermore if $\tau_i$ is the first recurrence time to $i$ one has $E[\tau_i \mid X_0 = i] = 1/\pi(i)$.

Proof. The formula for the recurrence time is immediate from Proposition 5.2 since the stationary sequence is ergodic. This result evidently also implies that

\[(1.33) \quad P(X_n = i \text{ for some finite } n \mid X_0 = i) = 1.\]

Recurrence i.e. return to $i$ infinitely often follows immediately from (1.33). \qed

2. Markov Chains with Finite State Space

We consider the case when the Markov chain is time independent and the state space is finite, $|F| = m$, in which case the transition probabilities $p(i, j)$, $i, j \in F$, form an $m \times m$ transition matrix $T$ with non-negative entries which satisfy

\[(2.1) \quad T1 = 1, \quad 1 = [1, 1, ...,].\]

Equation (1.13) can be written in matrix form as

\[(2.2) \quad \pi T = \pi \quad \text{or alternatively } T^* \pi = \pi,\]

where $T^*$ is the adjoint of $T$. Now (2.1) implies that $T$ has eigenvalue 1, and hence $T^*$ has eigenvalue 1. Thus a non-trivial solution to (2.2) exists.

Proposition 2.1. There exists a non-trivial solution $\pi$ to (2.2) with all non-negative entries. If the chain is indecomposable -see (1.14)- it is unique.
Proof. Evidently (1.13) implies that

\[ (2.3) \quad \sum_{i \in F} |\pi(i)| \cdot p(i, j) \geq |\pi(j)|, \quad j \in F. \]

If strict inequality holds in (2.3) for some \( j \in F \), then on summing (2.3) over \( j \in F \) we conclude that

\[ (2.4) \quad \sum_{i \in F} |\pi(i)| > \sum_{j \in F} |\pi(j)|, \]

which is an impossibility. We have shown therefore that the vector with entries \( |\pi(i)|, \quad i \in F \), is an eigenvector of \( T^* \) with eigenvalue 1.

Assume now that the chain is indecomposable and that \( \pi_1(\cdot) \) and \( \pi_2(\cdot) \) are two distinct invariant measures. Setting \( \pi(\cdot) = \pi_1(\cdot) - \pi_2(\cdot) \) and defining sets \( A_0, A_1 \) by

\[ (2.5) \quad A_0 = \{ i \in F : \pi(i) > 0 \}, \quad A_1 = \{ i \in F : \pi(i) < 0 \}, \]

we see from the fact that \( \sum_{i \in F} \pi(i) = 0 \), that both \( A_0 \) and \( A_1 \) are non-empty. Using the fact that \( \pi(\cdot) \) is a solution to (2.2), we conclude that equality holds in (2.3) for all \( j \in F \). It follows that if \( j \notin A_0 \cup A_1 \), then \( p(i, j) = 0 \) for all \( i \in A_0 \cup A_1 \). Further if \( j \in A_1 \) then \( p(i, j) = 0 \) for \( i \in A_0 \). We conclude that

\[ (2.6) \quad \sum_{j \in A_0} p(i, j) = 1, \quad i \in A_0. \]

Since we get a similar result for \( A_1 \) the chain is decomposable, a contradiction. \( \square \)

We have already seen in Proposition 1.2 that the stationary sequence \((X_0, X_1, \ldots)\) associated with the Markov chain is ergodic if the chain is indecomposable. Next we wish to show that under certain conditions on the transition matrix \( T \) the sequence is strong mixing. In order to do this we first prove a classical theorem in the theory of positive matrices.

**Theorem 2.1 (Perron-Frobenius Theorem).** Suppose \( T = (t_{i,j}) \) is an \( n \times n \) matrix with entries \( t_{i,j} \geq 0 \) and such that for some positive integer \( k \) the entries of \( T^k \) are all strictly positive. Then there exists an eigenvalue \( r \) of \( T \) such that:

(a) \( r \) is real and strictly positive.

(b) \( r \) is a simple root of the characteristic polynomial of \( T \).

(c) If \( \lambda \in \mathbb{C} \) is a root of the characteristic polynomial of \( T \) different from \( r \) then \( |\lambda| < r \).

(d) The eigenvector of \( T \) for \( r \) can be chosen so that all its entries are strictly positive.

**Proof.** We first construct an eigenvector for \( T \) with all positive entries. To do this let \( S \) be the set

\[ (2.7) \quad S = \{ x = (x_1, \ldots, x_n) \in \mathbb{R}^n : x_i > 0, \quad i = 1, \ldots, n, \quad |x| = 1 \}, \]

so \( S \) is the part of the unit sphere which is strictly in the positive quadrant (if \( n = 2 \)). Define now a function \( r : S \to \mathbb{R} \) by

\[ (2.8) \quad r(x) = \min_{1 \leq i \leq n} \frac{(Tx)_i}{x_i}. \]
It is clear that \( r(\cdot) \) is a non-negative continuous function on \( S \) and is bounded above as
\[
(2.9) \quad r(x) \leq \max_{1 \leq i \leq n} \sum_{j=1}^{n} t_{i,j}.
\]
It follows from (2.9) that
\[
(2.10) \quad r = \sup_{x \in S} r(x) < \infty,
\]
Hence for all \( x \in S \),
\[
(2.11) \quad (T x)_i \leq rx_i \quad \text{for some } i = i(x), \ 1 \leq i \leq n, \ \text{depending on } x \in S.
\]
Let \( x_N \in S, \ N = 1, 2, \ldots \), be a sequence such that \( \lim_{N \to \infty} r(x_N) = r \). Then some subsequence of \( x_N \) converges to a point \( x_\infty \in \overline{S} \), the closure of \( S \) in \( \mathbb{R}^n \) so \( |x_\infty| = 1 \).

Furthermore (2.8) implies that
\[
(2.12) \quad z = [T - r]x_\infty \text{ is a vector with all nonnegative entries.}
\]
Since the matrix \( T^k \) has all strictly positive entries, it follows that the vector \( w_k = T^k z \) has all strictly positive entries if \( z \neq 0 \). In that case (2.12) implies that
\[
(2.13) \quad r(T^k x_\infty/|T^k x_\infty|) > r,
\]
a contradiction to the definition of \( r \), from which we conclude that \( z = 0 \). This also implies that \( T^k x_\infty \) is an eigenvector of \( T \) with eigenvalue \( r \) which has all strictly positive entries. We have proven (a) and (d).

Next we turn to (b). Let \( \lambda \in \mathbb{C} \) be an eigenvalue of \( T \) in which case there exists a non-zero vector \( x \in \mathbb{C}^n \) such that \( T x = \lambda x \). Letting \( \bar{x} \in \mathbb{R}^n \) be the vector \( \bar{x} = (|x_1|, \ldots, |x_n|) \), we see that
\[
(2.14) \quad (\bar{T} \bar{x})_i \geq |\lambda||x_i|, \quad i = 1, \ldots, n,
\]
whence it follows from (2.11) that \( |\lambda| \leq r \). Supposing now that \( |\lambda| = r \), we can argue in the same way we established that \( x_\infty \) is an eigenvector of \( T \) with eigenvalue \( r \), that \( \bar{x} \) is also an eigenvector of \( T \) with eigenvalue \( r \). Thus we have that
\[
(2.15) \quad T^k x = \lambda^k x, \quad T^k \bar{x} = r^k \bar{x}.
\]
This implies that
\[
(2.16) \quad |(T^k x)_i| = (T^k \bar{x})_i, \quad 1 \leq i \leq n.
\]
We conclude from (2.16) using the fact that all entries of \( T^k \) are strictly positive that
\[
(2.17) \quad x = e^{i\theta} \bar{x}, \quad \text{for some } \theta \in [0, 2\pi).
\]
Thus \( |\lambda| = r \) implies \( \lambda = r \), which completes the proof of (c).

Note that the argument of the previous paragraph rules out the possibility of having 2 independent eigenvectors with eigenvalue \( r \). In fact if two such existed we could construct an eigenvector \( x = (x_1, \ldots, x_n) \) all of whose entries do not have identical phase \( e^{i\theta} \) in (2.17). This yields a contradiction. To complete the proof of (b) we need to exclude the possibility that \( T \) has any generalized eigenvectors with eigenvalue \( r \). Let us suppose a generalized eigenvector exists, whence there exists \( y \in \mathbb{C}^d \) such that
\[
(2.18) \quad [T - r]y = x.
\]
where $x$ is the unique eigenvector of $T$ with eigenvalue $r$ and all positive entries. From what we have already proved, there also exists a unique eigenvector $x^*$ of $T^*$ with eigenvalue $r$ and all positive entries. Applying $x^*$ to the left side of the equation (2.18) we obtain that
\[(2.19)\quad x^* [T - r] y = x^* x > 0.\]
Since $x^* [T - r] = 0$ we have a contradiction. \hfill \Box

**Corollary 2.1.** Suppose a time independent Markov chain on a finite state space $F$ is irreducible. Then it has a unique invariant distribution $\pi(\cdot)$ and $\pi(j) > 0$ for all $j \in F$.

**Proof.** For $N = 1, 2, \ldots$, let $K_N$ be the matrix
\[(2.20)\quad K_N = \frac{1}{N} \sum_{n=0}^{N-1} T^n.\]
Irreducibility implies that for large enough $N$ the matrix $K_N$ has all strictly positive entries. Since $\pi K_N = \pi$ it follows from Theorem 2.1 that $\pi(\cdot)$ is unique and all entries of $\pi(\cdot)$ are strictly positive. \hfill \Box

**Definition 3.** An $n \times n$ matrix $T$ has a spectral gap if there is exactly one simple root $\lambda \in \mathbb{C}$ of the characteristic polynomial for $T$ which satisfies $|\lambda| = \sigma(T) = \max\{|\lambda| \in \mathbb{C} : \lambda \text{ eigenvalue of } T\}$.

Note that the Perron-Frobenius theorem implies that a matrix $T$ with non-negative entries such that $T^k$ has strictly positive entries for some $k \geq 1$ has a spectral gap.

**Proposition 2.2.** Suppose the transition matrix $T$ for a finite state Markov chain has a spectral gap. Then the associated stationary sequence $(X_0, X_1, \ldots)$ is strong mixing.

**Proof.** From (3.26) of Chapter II we need to show that for any bounded Borel measurable functions $f, g : \mathbb{R}^{m+1} \to \mathbb{R}$ one has
\[(2.21)\quad \lim_{n \to \infty} E[ f(X_0, \ldots, X_m) g(X_n, X_{n+1}, \ldots, X_{n+m}) ] = E[ f(X_0, X_1, \ldots, X_m) ] E[ g(X_0, \ldots, X_m) ].\]
Observe now that for $n \geq m$ the LHS of (2.21) can be written as
\[(2.22)\quad E[ f(X_0, \ldots, X_m) \sum_{z \in F} T_{X_0,z}^{n-m} E[ g(X_0, X_1, \ldots, X_m) \mid X_0 = z ] ],\]
where $T_{a,b}^{n-m}, y, z \in F$, denotes the entries of the matrix $T^{n-m}$. Hence (2.21) will follow from (2.22) if we can show that
\[(2.23)\quad \lim_{k \to \infty} T_{y,z}^k = \pi(z), \quad y, z \in F.\]
Using the notation of (2.1), (2.2), consider the matrix $A = T - 1\pi$, where we are taking $\pi(\cdot)$ to be a row vector and $1$ a column vector. By the spectral gap condition all eigenvalues of $A$ have absolute value strictly less than 1, whence $\lim_{k \to \infty} A^k = 0$. Since $A^k = T^k - 1\pi$ the limit (2.23) holds. \hfill \Box
We have used the Perron-Frobenius theorem to establish some properties of
time independent Markov chains, assuming certain conditions on the transfer ma-
trix. Another condition on the transfer matrix which can help us understand the
对应的 Markov chain is so called reversibility. This condition is in effect
stating that the transfer matrix is self-adjoint with respect to some positive defi-
nite diagonal quadratic form. In that case much more is known about the transfer
matrix than in the general case, in particular that all its eigenvalues are real and
the matrix is diagonalizable. To see what this states about the entries \((t_{i,j})\) of the
\(n \times n\) matrix \(T\), suppose the quadratic form is given by
\[
(\xi, \zeta)_\mu = \sum_{i=1}^{n} \mu(i) \xi_i \zeta_i, \quad \xi, \zeta \in \mathbb{R}^n,
\]
where \(\mu \in \mathbb{R}^n\) has all positive entries. Then the self-adjointness condition
\[
(\xi, T\zeta)_\mu = (T\xi, \zeta)_\mu,
\]
is equivalent to
\[
\mu(i)t_{i,j} = \mu(j)t_{j,i}, \quad 1 \leq i, j \leq n.
\]
From (2.26) we see that if we normalize the vector \(\mu(\cdot)\) so that the sum of its
elements is 1, then \(\mu\) is the invariant measure for the Markov chain. Note that for
a reversible indecomposable Markov chain, if \(T\) does not have eigenvalue \(-1\) there
is a spectral gap and hence the chain is strong mixing.

3. Examples of Markov Chains

We first note a trivial case of a time independent Markov chain, which is a
sequence of i.i.d. random variables \((X_0, X_1, \ldots)\). Let \(p(i), \ i \in F\), be the probability
density function of \(X_0\). Then the transition probabilities are \(p(i,j) = p(j), \ i, j \in F\).
If \(F\) is finite then all the rows of the matrix \(T\) are identical, so \(T\) has rank 1. In
particular the pdf of \(X_0\) is the invariant measure \(\pi\) and the matrix \(A = T - 1_\pi\) is
the 0 matrix. Hence the sequence is strong mixing by Proposition 2.2.

Another example of a time independent Markov chain is the sum \(S_n = X_1 +
\cdots + X_n, \ n = 0, 1, 2, \ldots\) of i.i.d. variables. In that case \(p(i,j) = p(j-i), \ where \(p(\cdot)\)
is the pdf of \(X_1\). Thus sums of i.i.d. variables give rise to chains which are spatially
translation invariant. It is therefore often useful to use the Fourier transform to
analyse them. We essentially did this in our proof of the CLT in Chapter I.

In the case of the Bernoulli variables \(X_0 = \pm 1\) with probability \(1/2\), for which
\(S_n\) is the standard random walk on \(Z\), we have that \(p(i,j) = p(j,i) = 1/2\) if
and only if \(|i-j| = 1\). This chain is therefore reversible, and it is also easy to
see that it is irreducible. We have in Chapter 1 proved that it is recurrent, but
no invariant measure exists (compare Proposition 1.3). To see this note that if
\(\pi(\cdot)\) is an invariant measure then trivially \(\tau\pi(\cdot)\) is also an invariant measure where
\(\tau\pi(i) = \pi(i+1), \ i \in Z\). The uniqueness theorem Proposition 1.2 implies then that
\(\tau\pi(\cdot) = \pi(\cdot)\), in which case \(\pi : Z \to \mathbb{R}\) is a constant function, contradicting the
normalization condition for \(\pi(\cdot)\).