Introduction to Markov Chains

Witold Bednorz

Chapter 1

Definition of a Markov Chain

1.1 Stochastic Process

A Markov chain is a collection of random variables $\mathsf{X} = (X_n)_{n \in T}$, where T is a countable time-set. It is customary to write T as $\mathbb{Z}_+ := \{0, 1, 2, ...\}$, and we will do this henceforth.

We require that each X_i takes values in $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ (some measurable space which we call state space for X). There are three different types of state space:

1. The state space $\mathcal{X}$ is called *countable* if $\mathcal{X}$ is discrete with a finite or countable number of elements, and with $\mathcal{B}(\mathcal{X})$ the σ-field of all subsets of $\mathcal{X}$.

2. The state space X is called *general* if it is equipped with a countably generated σ-field $\mathcal{B}(X)$.

3. The state space $\mathcal{X}$ is called *topological* if it is equipped with a locally compact, separable, metrizable topology with $\mathcal{B}(\mathcal{X})$ as the Borel σ-field.

Whenever we consider the Markov chain as an entity we regard values of the whole chain X itself as lying in the *sequence* or *path* space formed by a countable product $\Omega = \mathcal{X}^\infty = \prod_{i=0}^{\infty} \mathcal{X}_i$ where each $\mathcal{X}_i$ is a copy of $\mathcal{X}$ equipped with a copy of $\mathcal{B}(\mathcal{X})$. To make the definition precise (so that $\mathsf{X_i}$ is a random variable) we define a σ-field $\mathcal{F}$ as the smallest that contains $\mathcal{B}(X_i)$ in projection (i.e. we require that projections on each coordinate are measurable). Finally, for each $x \in \mathcal{X}$ we there will be a measure $\mathbf{P}_x$ such that the probability of the event $\{\mathsf{X} \in A\}$ is well defined for any $A \in \mathcal{F}$; the initial condition requires of course that $\mathbf{P}_x(X_0 = x) = 1$. The triple $(\Omega, \mathcal{F}, \mathbf{P}_x)$ thus defines a stochastic process, since $\Omega = \{\omega_0, \omega_1, ... : \ \omega_i \in \mathcal{X}_i\}$ has the product structure to enable the projections ω_n at time n to be well defined realizations of the random variables X_n.

Heuristically the critical aspect of a Markov process, as opposed to any other set of random variables, is that it is forgetful of all but its most immediate past. The precise meaning of this requirement for the evolution of Markov chain in time, is that the future of the process is independent of the past given only its present value.

1.2 Markov Chain on Countable Space

Definition 1 *The process* $\mathsf{X} = (X_0, X_1, ...)$ *taking values in* $(\Omega, \mathcal{F}, \mathbf{P})$ *is a Markov chain if for every* n*, and any sequence of states* $\{x_0, x_1, x_2, ..., x_n\}$*,*

$$\mathbf{P}_\mu(X_0 = x_0, X_1 = x_1, ..., X_n = x_n) = \mu(x_0)\mathbf{P}_{x_0}(X_1 = x_1)...\mathbf{P}_{x_{n-1}}(X_1 = x_n) \tag{1.1}$$

The probability measure μ *is called the initial distribution of the chain. The process* X *is a time homogeneous Markov chain if the probabilities* $\mathbf{P}_{x_j}(X_1 = x_{j+1})$ *depend on the values* x_j, x_{j+1} *and are independent of the time points* j*.*

If X is a time homogeneous Markov chain, we write

$$P(x, y) := \mathbf{P}_x(X_1 = y);$$

then the definition (1.1) can be rewritten

$$\mathbf{P}_\mu(X_0 = x_0, X_1 = x_1, ..., X_n = x_n) = \mu(x_0)P(x_0, x_1)P(x_1, x_2)...P(x_{n-1}, x_n). \tag{1.2}$$

or equivalently in terms of the conditional probabilities of the process X,

$$\mathbf{P}_\mu(X_{n+1} = x_{n+1}|X_n = x_n, ..., X_0 = x_0) = P(x_n, x_{n+1}).$$

In applications we almost always define $\mathbf{P}_{x_0}$ for a fixed x_0 by defining the one step transition probabilities for the process and building the overall distribution using (1.2). This is done by using the Transition Probability matrix.

Definition 2 *The matrix* $P = \{P(x, y), x, y \in \mathcal{X}\}$ *is called a Markov transition matrix if*

$$P(x, y) \geqslant 0, \quad \sum_{z \in \mathcal{X}} P(x, z) = 1, \quad x, y \in \mathcal{X} \tag{1.3}$$

We define the usual matrix iterates $P^n = \{P^n(x, y), x, y \in \mathcal{X}\}$ by setting $P^0 = Id$, the identity matrix and then taking inductively

$$P^n(x, y) = \sum_{y \in \mathcal{X}} P(x, y)P^{n-1}(y, z).$$

Theorem 1 *If* $\mathcal{X}$ *is countable, and*

$$\mu = \{\mu(x), x \in \mathcal{X}\}, \quad P = \{P(x, y), x, y \in X\}$$

are an initial measure on $\mathcal{X}$ *and a Markov transition matrix satisfying (1.3) then there exists a Markov chain* X *on* $(\Omega, \mathcal{F})$ *with probability law* $\mathbf{P}_\mu$ *satisfying*

$$\mathbf{P}_\mu(X_{n+1} = y|X_n = x, ..., X_0 = x_0) = P(x, y).$$

1.3 General State space Definition

We let $\mathcal{X}$ be a general set and $\mathcal{B}(\mathcal{X})$ denote a countably generated σ-field on $\mathcal{X}$: when $\mathcal{X}$ is topological, then $\mathcal{B}(\mathcal{X})$ will be taken as the Borel σ-field, but otherwise it may be arbitrary.

Definition 3 *If* $P = \{P(x, A), x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})\}$ *is such that*

1. *for each* $A \in \mathcal{B}(X)$*,* $P(\cdot, A)$ *is a non negative measurable function on* $\mathcal{X}$

2. *for each* $x \in X$*,* $P(x, \cdot)$ *is a probability measure on* $\mathcal{B}(\mathcal{X})$

then we call P *a transition probability or Markov transition function.*

To extend the definition from the countable space case to the general one we first define a finite sequence $\{X_0, X_1, ..., X_n\}$ of random variables on the product space $\mathcal{X}^{n+1} = \prod_{i=0}^{n} \mathcal{X}_i$ equipped with the product σ-field $\mathcal{F}_n$ generated by $\mathcal{B}(\mathcal{X}_i)$. For any measurable sets $A_i \subset X_i$ we develop the set function $\mathbf{P}_x^n(\cdot)$ on $\mathcal{X}^{n+1}$ by setting, for a fixed point $x \in \mathcal{X}$ and for the 'cylinder sets' $A_1 \times ... \times A_n$

$$\begin{aligned}
&\mathbf{P}_x^1(A_1) = P(x, A_1), \\
&\mathbf{P}_x^2(A_1 \times A_2) = \int_{A_1} P(x, dy_1) P(y_1, A_2), \qquad \qquad \ldots \\
&\mathbf{P}_x^n(A_1 \times ... \times A_n) = \int_{A_1} P(x, dy_1) \int_{A_2} P(y_1, dy_2) ... P(y_{n-1}, A_n).
\end{aligned}$$

These are well-defined by the this procedure for increasing n, we find

Theorem 2 *For any initial measure* μ *on* $\mathcal{B}(\mathcal{X})$*, and any transition probability kernel* $P = \{P(x, A), x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})\}$*, there exists a stochastic process* $\mathsf{X} = \{X_0, X_1, ...\}$ *on* $\Omega = \prod_{i=0}^{\infty}$ *measurable with respect to* $\mathcal{F}$ *(generated by* $\mathcal{B}(\mathcal{X}_i)$*) and a probability measure* $\mathbf{P}_\mu$ *on* $\mathcal{F}$ *such that* $\mathbf{P}_\mu(B)$ *is the probability of the event* $\{\mathsf{X} \in B\}$ *for* $B \in \mathcal{F}$*, and for measurable* $A_i \subset \mathcal{X}_i$*,* $i = 0, ..., n$ *and any* n

$$\mathbf{P}_\mu(X_0 \in A_0, X_1 \in A_1, ..., X_n \in A_n) = \int_{A_0} ... \int_{A_n} \mu(dy_0) P(y_0, dy_1) ... P(y_{n-1}, A_n). \tag{1.4}$$

Definition 4 *The stochastic process* X *defined on* $(\Omega, \mathcal{F})$ *is called a time-homogeneous Markov chain with transition probability kernel* $P(x, A)$ *and initial distribution* μ *if the finite dimensional distributions of* X *satisfy (1.4).*

As on countable spaces the n-step transition probability kernel is defined iteratively. We set $P^0(x, A) - \delta_x(A)$, the Dirac measure defined by

$$\delta_x(A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A, \end{cases}$$

and for $n \geqslant 1$, we define inductively

$$P^n(x, A) = \int_{\mathcal{X}} P(x, dy) P^{n-1}(y, A), \quad x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X}),$$

We write P^n for the n-step transition probability kernel $\{P^n(x, A), x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})\}$, note that P^n is defined analogously to the n-step transition probability matrix for the countable space case.

Theorem 3 *Chapman-Kolmogorov equations. For any* m *with* $0 \leqslant m \leqslant n$*,*

$$P^n(x, A) = \int_{\mathcal{X}} P^m(x, dy) P^{n-m}(y, A), \quad x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X}). \tag{1.5}$$

We interpret (1.5) as saying that, as X moves from x into A is n steps, at any intermediate time m it must take some value $y \in \mathcal{X}$ and that being a Markov chain it forgets the past at that time m and moves the succeeding $(n-m)$ steps with the law appropriate to starting afresh at y. We can write equation (1.5) alternatively as

$$\mathbf{P}_x(X_n \in A) = \int_{\mathcal{X}} P_x(X_m \in dy) P_y(X_{n-m}).$$

Exactly as the one-step transition probability kernel describes a chain X, the m-step kernel satisfies the definition of a transition kernel, and thus defines a Markov chain $\mathsf{X}^{\mathsf{m}} = \{X_n^m\}$ with transition probabilities

$$\mathbf{P}_x(x_n^m \in A) = P^{nm}(x, A). \tag{1.6}$$

Definition 5 *The chain X^m with transition law (1.6) is called the m-skeleton of the chain X. The resolvent K_a is defined for $0 < \varepsilon < 1$ by*

$$K_\varepsilon(x, A) = (1-\varepsilon) \sum_{i=0}^{\infty} \varepsilon^i P^i(x, A), \quad x \in \mathsf{X}, A \in \mathcal{B}(\mathcal{X}).$$

The Markov chain with transition function K_ε is called the K_ε-chain.

1.4 Semigroup

In the general case the kernel P^n operates on quite different entities from the left and the right. As an operator P^n acts on both bounded measurable functions f on $\mathcal{X}$ and on σ-finite measures μ on $\mathcal{B}(\mathcal{X})$ via

$$P^n f(x) = \int_{\mathcal{X}} P^n(x, dy) f(y), \quad \mu P^n(A) = \int_{\mathcal{X}} \mu(dx) P^n(x, A),$$

and we shall use the notation $P^n f$, μP^n to denote these operations. We shall also write

$$P^n(x, f) := \int_{\mathcal{X}} P^n(x, dy) f(f) := \delta_x P^n f$$

if it is notationally convenient. In general, the functional notation is more compact: for example we can rewrite the Chapman-Kolmogorov equations as

$$P^{m+n} = P^n P^m, \quad m, n \in \mathbb{Z}_+.$$

On many occasions, though where we feel that the argument is more transparent when written in full form we shall revert to the more detailed presentation.

1.5 Some basic properties

Proposition 1 *If X is a Markov chain on $(\Omega, \mathcal{F})$, with initial measure μ, and $h : \Omega \to \mathbb{R}$ is bounded and measurable, then*

$$\mathbf{E}_\mu(h(X_{n+1}, X_{n+2}, ...) | X_0, ..., X_n; X_n = x) = \mathbf{E}_x h(X_1, X_2, ...).$$

The formulation of the Markov concept is made much simpler if we develop more systematic notation for the information encompassed in the past of the process, and if we introduce the shift operator on the space Ω. For a given initial distribution, define the σ-field

$$\mathcal{F}_n^{\mathsf{X}} := \sigma(X_0, ..., X_n) \subset \mathcal{B}(\mathcal{X}^{n+1}).$$

which is the smallest σ-field for which the random variable $\{X_0, ..., X_n\}$ is measurable. In many cases, $\mathcal{F}_n^{\mathsf{X}}$ will coincide with $\mathcal{B}(\mathcal{X}^n)$, although this depends in particular on the initial measure μ chosen for a particular chain. the *shift operator* θ is defined to be the mapping on Ω defined by

$$\theta(\{x_0, x_1, ...\}) = \{x_1, x_2, ...\}.$$

We write θ^k for the k-th iterate of the mapping θ, defined inductively by

$$\theta^1 = \theta, \ \ \theta^{k+1} = \theta \cdot \theta^k, \ \ k \geqslant 1.$$

The shifts θ^k define operators on random variables H on $(\Omega, \mathcal{F}, \mathbf{P}_\mu)$ by

$$(\theta^k H)(\omega) = H \cdot \theta^k(\omega).$$

It is obvious that $X_n \cdot \theta^k(\omega) = X_{n+k}$. Hence if the random variable H is of the form $H = h(X_0, X_1, ...)$ for a measurable function h on the sequence Ω then

$$\theta^k H = h(X_k, X_{k+1}, ...).$$

Since the expectation $\mathbf{E}_x H$ is a measurable function on X, it follows that $\mathbf{E}_{X_n}$ is a random variable on $(\Omega, \mathcal{F}, \mathbf{P})$ for any initial distribution. With this notation the equation

$$\mathbf{E}_\mu(\theta^n H | \mathcal{F}_n^{\mathsf{X}}) = \mathbf{E}_{X_n} H, \ \ \text{a.s. } \mathbf{P}_\mu$$

valid for any bounded measurable h and fixed $n \in \mathbb{Z}_+$, describes the *the time homogeneous Markov property* in a succinct way.

It is not always the case that $\mathcal{F}_n^{\mathsf{X}}$ is complete (contains every set of $\mathbf{P}_\mu$-measure zero). For any initial measure μ we say that an event A occurs $\mathbf{P}_\mu$-a.s. to indicate that A^c is a set contained in an element of $\mathcal{F}_n^{\mathsf{X}}$ which is of P_μ-measure zero. If A occurs $\mathbf{P}_x$-a.s. for all $x \in \mathcal{X}$ then we write that A occurs $\mathbf{P}_*$-a.s.

Definition 6 *For any set $A \in \mathcal{B}(\mathcal{X})$ the occupation time η_A is the number of visits by X to A after times zero and is given by*

$$\eta_A := \sum_{n=1}^{\infty} 1_{\{X_n \in A\}}.$$

For any set $A \in \mathcal{B}(\mathcal{X})$, the variables

$$\tau_A := \min\{n \geqslant 1 : \ X_n \in A\},$$
$$\sigma_A := \min\{n \geqslant 0 : \ X_n \in A\}$$

are called first return and first hitting times on A respectively.

We also refer to the kernel U defined as

$$U(x, A) := \sum_{n=1}^{\infty} P^n(x, A) = \mathbf{E}_x \eta_A,$$

which maps $\mathcal{X} \times \mathcal{B}(X)$ to $\mathbb{R} \cup \{\infty\}$, and the return time probabilities

$$L(x, A) := \mathbf{P}_x(\tau_A < \infty) = \mathbf{P}_x(\mathsf{X} \text{ ever enters } A).$$

A function $\zeta : \Omega \in \mathbb{Z}_+ \cup \{\infty\}$ is a stopping time for X if for any initial distribution μ the event $\{\zeta = n\} \in \mathcal{F}_n^{\mathsf{X}}$ for all $n \in \mathbb{Z}_+$.

Proposition 2 *For any $A \in \mathcal{B}(X)$, the variables τ_A and σ_A are stopping times for* X.

Proposition 3 *For all $x \in A$, $A \in \mathcal{B}(\mathcal{X})$*

$$\mathbf{P}_x(\tau_A = 1) = P(x, A),$$

and inductively for $n > 1$

$$\mathbf{P}_x(\tau_A = n) = \int_{A^c} P(x, dy)\mathbf{P}_y(\tau_A = n-1) = \int_{A^c} P(x, dy_1) \int_{A^c} P(y_1, dy_2) ... \int_{A^c} P(y_{n-2}, dy_{n-1}) P(y_{n-1}, A).$$

For all $x \in \mathcal{X}$, $a \in \mathcal{B}(\mathcal{X})$

$$\mathbf{P}_x(\sigma_A = 0) = 1_A(x)$$

and for $n \geqslant 1$, $x \in A^c$

$$\mathbf{P}_x(\sigma_A = n) = \mathbf{P}_x(\tau_A = n).$$

The simple Markov property (1.5) holds for any bounded measurable h and fixed $n \in \mathbb{Z}_+$. We now extend (1.5) to stopping times. If ζ is an arbitrary stopping time, then the fact that our time-set is $\mathbb{Z}_+$ enables us to define the random variable $X_\zeta = X_n$ on the event $\{\zeta = n\}$. For a stopping time ζ the property which tells us that the future evolution of X after the stopping time depends only on the value X_ζ, rather then on any other past values, is called the Strong Markov Property.

To describe this formally, we need to define the σ-field

$$\mathcal{F}_\zeta^{\mathsf{X}} := \{A \in \mathcal{F} : \; \{\zeta = n\} \cap A \in \mathcal{F}_n^{\mathsf{X}}, n \in \mathbb{Z}_+\},$$

which describes events which happen up to time ζ. For a stopping time ζ and a random variable $H = h(X_0, X_1, ...)$ the shift θ^ζ is defined as

$$\theta^\zeta = h(X_\zeta, X_{\zeta+1}, ...),$$

on the set $\{\zeta < \infty\}$. The required extension of (1.5) is then

Definition 7 *We say that X has the Strong Markov Property if for any initial distribution μ, any real-valued bounded measurable function h on Ω, and any stopping time ζ,*

$$\mathbf{E}_\mu(\theta^\zeta H | \mathcal{F}_\zeta^{\mathsf{X}}) = \mathbf{E}_{X_\zeta} H \;\; a.s. \; \mathbf{P}_\mu$$

on the set $\{\zeta < \infty\}$.

Proposition 4 *For a Markov chain X with discrete time parameter, the Strong Markov Property always holds.*

Chapter 2

Irreducibility

2.1 Communicating classes and irreducibility

The idea of a Markov chain X reaching sets or points is much simplified when X is countable and the behavior of the chain is governed by a transition probability matrix $P = P(x, y)$, $x, y \in \mathsf{X}$. There are then number of essential equivalent ways of defining the operation of communication between states.

The simplest is to say that state x *leads to state* y which we write as $x \to y$, if $L(x, y) > 0$, and that two distinct states x and y in X when $L(x, y) > 0$ and $L(y, x) > 0$. By convention we also define $x \to x$. The relation $x \leftrightarrow y$ is often defined equivalently by requiring that there exists $n(x, y) \geqslant 0$ and $m(y, x) \geqslant 0$ such that $P^n(x, y) > 0$, and $P^m(y, x) > 0$ i.e. $\sum_{n=0}^{\infty} P^n(x, y) > 0$ and $\sum_{n=0}^{\infty} P^n(y, x) > 0$.

Proposition 5 *The relation $\leftrightarrow$ is an equivalence relation, and so the equivalence classes $C(x) = \{y : x \leftrightarrow y\}$ cover X, with $x \in C(x)$.*

Definition 8 *If $C(x) = \mathsf{X}$ for some x, then we say that X is irreducible. We say $C(x)$ is absorbing if $P(y, C(x)) = 1$ for all $y \in C(x)$.*

When states do not all communicate, then although each state in $C(x)$ communicates with every other state in $C(x)$, it is possible that there are states $y \in C(x)^c$ such that $x \to y$. This happens if and only if $C(x)$ is not absorbing.

Suppose that X is not irreducible for X. If we reorder that states according to the equivalence classes defined by the communication operation, an if we further order the classes with absorbing classes coming first, then we have a decomposition of P such as

$$\mathcal{X} = \left(\sum_{x \in I} C(x)\right) \cup D,$$

where I is the set of indicators of each different absorbing class.

Proposition 6 *Suppose that $C := C(x)$ is an absorbing communicating class for some $x \in \mathcal{X}$. Let P_C denote the matrix P restricted to the states in C. Then there exists an irreducible Markov chain X_C whose state space is restricted to C and whose transition matrix is given by P_C.*

2.2 ψ-Irreducibility

Definition 9 *We call* X *φ-irreducible if there exists a measure φ on $\mathcal{B}(\mathcal{X})$ such that, whenever $\varphi(A) > 0$, we have $L(x, A) > 0$ for all $x \in \mathcal{X}$.*

Proposition 7 *The following are equivalent formulation of φ-irreducibility*

1. *for all $x \in \mathsf{X}$, whenever $\varphi(A) > 0$, $U(x, A) > 0$;*
2. *for all $x \in \mathsf{X}$, whenever $\varphi(A) > 0$, there exists some $n > 0$, possibly depending on both A and x such that $P^n(x, A) > 0$;*
3. *for all $x \in \mathsf{X}$, whenever $\varphi(A) > 0$ then $K_{1/2}(x, A) > 0$.*

Proposition 8 *If X is φ-irreducible for some measure φ, then there exists a probability measure ψ on $\mathcal{B}(\mathcal{X})$ such that*

1. *X is ψ-irreducible*
2. *for any other measure ψ', the chain X is φ'-irreducible if and only if $\psi \succ \varphi'$;*
3. *if $\psi(A) = 0$, then $\psi(\{y : \; L(y, A) > 0\}) = 0$;*
4. *the probability measure ψ is equivalent to*

$$\psi'(A) := \int_{\mathcal{X}} \varphi'(dy) K_{1/2}(y, A),$$

for any finite irreducible measure φ'.

Definition 10 1. *The Markov chain is called ψ-irreducible if it is φ-irreducible for some φ and the measure ψ is a maximal irreducibility measure satisfying conditions of Proposition 8.*

2. *We write*

$$\mathcal{B}_+(\mathcal{X}) := \{A \in \mathcal{B}(\mathcal{X}) : \psi(A) > 0\}$$

for the sets of positive ψ-measure; the equivalence of maximal irreducibility measures means that $\mathcal{B}_+$ is uniquely defined.

3. *We call a set $A \in \mathcal{B}(\mathcal{X})$, full if $\psi(A^c) = 0$.*
4. *We call a set $A \in \mathcal{B}(\mathcal{X})$ absorbing if $P(x, A) = 1$, for $x \in A$.*

Proposition 9 *Suppose that X is ψ=irreducible. Then*

1. *every absorbing set is full,*
2. *every full contains a non-empty, absorbing set.*

If a set C is absorbing and with there is a measure ψ for which

$$\psi(B) > 0 \Rightarrow L(x, B) > 0, \;\; x \in C$$

then we will call C an absorbing ψ-irreducible set.

Proposition 10 *Suppose that A is an absorbing set. Let P_A denote the kernel P restricted to the states in A. Then there exists a Markov chain X_A whose state space is A and whose transition matrix is given by P_A. Moreover, if X is ψ-irreducible then X_A is ψ-irreducible.*

2.3 Accessible Sets

Definition 11 *We say that a set $B \in \mathcal{B}(\mathcal{X})$ is accessible from another set $A \in \mathcal{B}(\mathcal{X})$ if $L(x,B) > 0$ for every $x \in A$;*
We say that a set $B \in \mathcal{B}(\mathcal{X})$ is uniformly accessible from another set $A \in \mathcal{B}(\mathcal{X})$ if there exists a $\delta > 0$ such that

$$\inf_{x \in A} L(x,B) \geqslant \delta; \tag{2.1}$$

and when (2.1) holds we write $A \mapsto B$

Let us define taboo probabilities

$$P^n_A(x,B) = \mathbf{P}_x(X_n \in B, \tau_A \geqslant n)$$

and

$$U_A(x,B) = \sum_{n=1}^{\infty} P^n(x,B).$$

Lemma 1 *If $A \mapsto B$ and $B \mapsto C$ then $A \mapsto C$.*

Definition 12 *The set $\bar{A} := \{x \in \mathcal{X} : \; L(x,A) > 0\}$ is the set of points from which A is accessible.*
The set $\bar{A} := \{x \in \mathcal{X} : \sum_{n=1}^{m} P^n(x,A) \geqslant m^{-1}\}$.
The set $A^0 := \{x \in \mathcal{X} : \; L(x,A) = 0\} = \bar{A}^c$ is the set of points from which A is not accessible.

Lemma 2 *The set $\bar{A} = \bigcup_m \bar{A}(m)$, and for each m we have $\bar{A}(m) \mapsto A$.*

Chapter 3

Pseudo Atoms

3.1 Minorization Condition

Definition 13 *A set $\alpha \in \mathcal{B}(\mathcal{X})$ is called an atom for* X *if there exists measure ν on $\mathcal{B}(\mathcal{X})$ such that*

$$P(x, A) = \nu(A), \quad x \in \alpha.$$

If X *is ψ-irreducible and $\psi(\alpha) > 0$ then α is called an accessible atom.*

Proposition 11 *Suppose there is an atom α in $\mathcal{X}$ such that $\sum_n P^n(x, \alpha) > 0$ for all $x \in \mathcal{X}$. Then α is an accessible atom and* X *is ν-irreducible with $\nu = P(\alpha, \cdot)$.*

Proposition 12 *If $L(x, A) > 0$ for some state $x \in \alpha$, where α is an atom, then $\alpha \mapsto A$.*

Definition 14 (Minorization Condition) *For some $\delta > 0$, some $C \in \mathcal{B}(\mathcal{X})$ and some probability measure ν with $\nu(C^c) = 0$ and $\nu(C) = 1$*

$$\mathbf{P}(x, A) \geqslant \delta 1_C(x)\nu(A), \quad A \in \mathcal{B}(\mathcal{X}), x \in \mathcal{X}. \tag{3.1}$$

The concept of splitting can be described as follows. Instead of $\mathcal{X}$ we write $\bar{\mathcal{X}} = \mathcal{X} \times \{0, 1\}$, where $\mathcal{X}_0 = \mathcal{X} \times \{0\}$ and $\mathcal{X}_1 = \mathcal{X} \times \{1\}$ are copies of $\mathcal{X}$ equipped with copies $\mathcal{B}(\mathcal{X}_0)$, $\mathcal{B}(\mathcal{X}_1)$ of $\mathcal{B}(\mathcal{X})$

We will write x_i, $i = 0, 1$ for elements of $\bar{\mathcal{X}}$, with x_0 denoting members of the upper level $\mathcal{X}_0$ and x_1 denoting members of the lower level $\mathcal{X}_1$. In order to describe more easily the calculation associated with moving between the original an the split chain, we will also sometimes call $\mathcal{X}_0$ the *copy* of $\mathcal{X}$ and will say that $A \in \mathcal{B}(\mathcal{X})$ is a *copy* of the corresponding set $A_0 \subset \mathcal{X}_0$.

If λ is any measure on $\mathcal{B}(\mathcal{X})$, then the next step in the construction is to split the measure λ^* on $\mathcal{B}(\mathcal{X})$ trough

$$\begin{aligned} \lambda^*(A_0) &= (1 - \delta)\lambda(A \cap C) + \lambda(A \cap C^c), \qquad (3.2)\\ \lambda^*(A_1) &= \delta\lambda(A \cap C), \end{aligned}$$

where δ and C are the constant and the set in (3.1). Note that in this sense the splitting is dependent on the choice of the set C, and although in general the set chosen is not relevant, we will on occasion need to make explicit the set in (3.1) when we use the split chain.

Note that λ is the marginal measure induced by λ^*, in the sense that for any $A \in \mathcal{B}(\mathcal{X})$ we have

$$\lambda^*(A_0 \cup A_1) = \lambda(A).$$

In the case when $A \subset C^c$, we have $\lambda^*(A_0) = \lambda(A)$; only subsets of C are really effectively split by this construction.

Now the third, and the most subtle step in the construction is to split the chain X to form a chain $\bar{\mathsf{X}}$ which lives on $(\bar{\mathcal{X}}, \mathcal{B}(\bar{\mathcal{X}}))$. Define the split kernel $\bar{P}(x_i, A)$ for $x_i \in \bar{\mathsf{X}}$ and $A \in \mathcal{B}(\bar{\mathcal{X}})$ by

$$\bar{P}(x_0, \cdot) = P(x, \cdot)^*, \quad x_0 \in \mathcal{X}_0 \backslash C_0; \tag{3.3}$$

$$\bar{P}(x_0, \cdot) = (1-\delta)^{-1}(P(x, \cdot)^* - \delta\nu^*(\cdot)), \quad x_0 \in C_0 \tag{3.4}$$

$$\bar{P}(x_1, \cdot) = \nu^*(\cdot), \quad x_1 \in \mathcal{X}_1. \tag{3.5}$$

where C, δ and ν are the set, the constant and the measure in the Minorization Condition.

Outside C the chain $\{\bar{X}_n\}$ behaves just like $\{X_n\}$, moving on the top half $\mathcal{X}_0$ of the split chain space. Each time it arrives in C, it is split: with probability $1-\delta$ it remains in C_0, with probability δ it drops to C_1. We can think of this splitting of the chain as tossing a δ-weighted coin to decide which level to choose on each arrival in the set C where the split takes place.

When the chain remains on the top level its next step has the modified law (3.4). That (3.4) is always nonnegative follows from (3.1). This is sole use of the Minorization Condition, although without it the chain cannot be defined.

Note here the whole point of the construction the bottom level of $\mathcal{X}_1$ is an atom, with $\varphi^*(\mathcal{X}_1) = \delta\varphi(C) > 0$ whenever the chain X is φ-irreducible. By (3.2) we have $\bar{P}^n(x_1, \mathcal{X}_1 \backslash X_1) = 0$ for all $n \geqslant 1$ and all $x_1 \in \bar{\mathcal{X}}$, so that the atom $C_1 \subset \mathcal{X}_1$ is the only part of the bottom level which is reached with positive probability. We will use the notation $\bar{\alpha} := C_1$ when we wish to emphasize the fact that all transition out of C_1 are identical, so that C_1 is an atom in $\bar{\mathcal{X}}$.

3.2 Connecting the Split and Original Chains

Theorem 4 *The chain X is the marginal chain of $\{\bar{X}_n\}$: that is for any initial distribution λ on $\mathcal{B}(\mathcal{X})$ and any $A \in \mathcal{B}(\mathcal{X})$,*

$$\int_{\mathcal{X}} P^k(x, A)\lambda(dx) = \int_{\bar{\mathcal{X}}} \bar{P}^k(y_i, A_0 \cup A_1)\lambda^*(dy_i). \tag{3.6}$$

The chain X is φ-irreducible if $\bar{\mathsf{X}}$ is φ^*-irreducible and if X is φ-irreducible with $\varphi(C) > 0$ then $\bar{\mathsf{X}}$ is ν^*-irreducible, and $\bar{\alpha}$ is an accessible atom for the split chain.

Note that for any measure μ on $\mathcal{B}(\mathcal{X})$ we have $\mu^*\bar{P} = (\mu P)^*$ that is

$$\int_{\bar{X}} \mu(dx_i)P(x_i, \cdot) = (\int_{\mathcal{X}} \mu(dx)P(x_i, \cdot))^*.$$

Since it is only the marginal chain X which is really of interest, we will usually consider only sets of the form $\bar{A} = A_0 \cup A_1$, where $A \in \mathcal{B}(\mathcal{X})$, and we will largely restrict ourselves to functions on $\bar{X}$ of the form $\bar{f}(x_i) = f(x_i)$, where f is some function on $\mathcal{X}$; that is $\bar{f}$ is identical on the two copies of $\mathcal{X}$. By (3.6) we have for any k, any initial distribution λ and any function $\bar{f}$ identical on $\mathcal{X}_0$ and $\mathcal{X}_1$

$$\mathbf{E}_\lambda f(X_k) = \bar{\mathbf{E}}\bar{f}(\bar{X}_k).$$

To emphasize the identity we will henceforth denote $\bar{f}$ by f and $\bar{A}$ by A in these special instances. The context should make clear whether A is a subset of $\mathcal{X}$ or $\bar{\mathcal{X}}$ and whether the domain of f is $\mathcal{X}$ or $\bar{\mathcal{X}}$.

3.3 Small Sets

Definition 15 *A set $C \in \mathcal{B}(\mathcal{X})$ is called a small set if there exists an $m > 0$, and a non trivial measure ν_m on $\mathcal{B}(\mathcal{X})$, such that for all $x \in C$, $B \in \mathcal{B}(\mathcal{X})$,*

$$P^m(x, B) \geqslant \nu_m(B). \tag{3.7}$$

When (3.7) holds we say that C is ν_m-small.

Theorem 5 *Suppose φ is a σ-finite measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Suppose A is any set in $\mathcal{B}(\mathcal{X})$ with $\varphi(A) > 0$ such that*

$$\varphi(B) > 0, B \subset A \ \Rightarrow \ \sum_{k=1}^{\infty} P^k(x, B) > 0, \ \ x \in A.$$

Then for every n the function p^n defined in the decomposition

$$P^n(x)(x, B) = \int_B p^n(x, y)\varphi(dy) + P_{\perp}(x, B),$$

can be chosen to be a measurable function on $\mathcal{X} \times \mathcal{X}$, and there exists $C \subset A$, $m > 1$, and $\delta > 0$ such that $\varphi(C) > 0$ and

$$p^m(x, y) > \delta, \ \ x, y \in C.$$

Theorem 6 *If X is ψ-irreducible, then for every $A \in \mathcal{B}^+(\mathcal{X})$, there exists $m \geqslant 1$ and a ν_m-small set $C \subset A$ such that $C \in \mathcal{B}_+(\mathcal{X})$ and $\nu_m(C) > 0$.*

Theorem 7 *If X is ψ-irreducible, then the Minorization Condition holds for some m-skeleton, and for every K_ε-chain, $0 < \varepsilon < 1$.*

Proposition 13

1. *If $C \in \mathcal{B}(\mathcal{X})$ is ν_m-small, and for any $x \in D$ we have $P^m(x, C) \geqslant \delta$, then D is ν_{n+m}-small, where ν_{n+m} is a multiple of ν_n.*

2. *Suppose X is ψ-irreducible. Then there exists a countable collection of sets C_i of small sets in $\mathcal{B}(\mathcal{X})$ such that*

$$\mathcal{X} = \bigcup_{i=0}^{\infty} C_i$$

3. *Suppose X is ψ-irreducible. If $C \in \mathcal{B}^+(\mathcal{X})$ is ν_n-small, then we may find $M \in \mathbb{Z}_+$ and a measure ν_M such that C is ν_M-small, and $\nu_M(C) > 0$.*

Chapter 4

Cyclic Behavior

4.1 Cycles for a Countable Space Chain

We first discuss the question for a countable space $\mathcal{X}$. Let α be a specific state in $\mathcal{X}$, and write

$$d(\alpha) = \text{G.C.D}\{n \geqslant 1 : \mathbf{P}^n(\alpha, \alpha) > 0\}.$$

We call $d(\alpha)$ the period of α. We recall that $C(\alpha) = \{y : \alpha \leftrightarrow \alpha\}$.

Proposition 14 *Suppose α has period $d(\alpha)$: then for any $y \in C(\alpha)$, $d(\alpha) = d(y)$.*

Proposition 15 *Let X be an irreducible Markov chain on a countable space, and let d denote the common period of the states in $\mathcal{X}$. Then there exists disjoint sets $D_1, D_2, ..., D_d \subset \mathcal{X}$ such that*

$$\mathcal{X} = \bigcup_{i=1}^{d} D_k,$$

and

$$P(x, D_{k+1}) = 1, \;\; x \in D_k, \;\; k = 0, ..., d-1 \;\; (\mod d)$$

Definition 16 *An irreducible chain on a countable space X is called*

1. *aperiodic, if $d(x) = 1$, $x \in \mathcal{X}$;*
2. *strongly aperiodic, if $P(x, x)$, for some $x \in \mathcal{X}$.*

Proposition 16 *Suppose X is an irreducible chain on a countable space $\mathcal{X}$, with period d and cyclic classes $\{D_1, ..., D_d\}$. Then for the Markov chain $\mathsf{X}_d = \{X_d, X_{2d}, ...\}$ with transition matrix P^d, each D_i is an irreducible absorbing set of aperiodic states.*

4.2 Cycles for a General Space Chain

The existence of small sets enables us to show that, even on a general space, we still have a finite periodic breakup into cyclic sets for ψ-irreducible chains.

Suppose that C is any ν_M-small set, and assume that $\nu_M(C) > 0$, as we may without loss of generality by Proposition 13. To simplify the notation we will suppress the subscript on ν. Hence we have

$P^M(x,\cdot) > \nu(\cdot)$, $x \in C$, and $\nu(C) > 0$, so that, when the chain starts in C, there is a positive probability that the chain will return to C at time M. Let

$$E_C = \{n \geqslant 1 : \text{ the set } C \text{ is } \nu_n\text{-small, with } \nu_n = \delta_n \nu \text{ for some} delta_n > 0.\}$$

Notice that for $B \subset C$, $n, m \in E_C$ implies

$$P^{n+m}(x,B) \geqslant \int_C P^m(x,dy)P^n(y,B) \geqslant [\delta_m \delta_n \nu(C)]\nu(B), \quad x \in C,$$

so that E_C is closed under addition. Thus there is a natural period for the set C, given by the greatest common divisor of E_C.

Lemma 3 *The set C is ν_{nd}-small for large enough n.*

Theorem 8 *Suppose X is a ψ-irreducible chain on $\mathcal{X}$. Let $C \in \mathcal{B}^+(\mathcal{X})$ be a ν_M-small set and let d be the greatest common divisor of the set E_C. Then there exists disjoint sets $D_1, ..., D_d \in \mathcal{B}(\mathcal{X})$ (a d-cycle) such that*

1. *for $x \in D_i$, $P(x, D_{i+1}) = 1$, $i = 0, ..., d-1$, $(\mod d)$;*
2. *the set $N = (\bigcup_{i=1}^d D_i)^c$ is ψ-null.*

The d-cycle $\{D_i\}$ is maximal in the sense that for any other collection $\{d', D'_k, k = 1, ..., d'\}$ satisfying (8), (1), we have d' dividing d; whilst if $d = d'$, then by reordering the indices if necessary $D'_i = D_i$ ψ-a.s.

Definition 17 *Suppose that X is an φ-irreducible Markov chain.*

1. *The largest d for which a d-cycle occurs for X is called the period of X.*
2. *When $d = 1$, the chain X is called aperiodic.*
3. *When there exists a ν_1-small set A, with $\nu_1(A) > 0$, then the chain is called strongly aperiodic.*

Proposition 17
1. *If X is strongly aperiodic, then the Minorization Condition holds.*
2. *The resolvent K_ε chain is strongly aperiodic for all $0 < \varepsilon < 1$.*
3. *If X is aperiodic then every skeleton is ψ-irreducible and aperiodic, and some m-skeleton is strongly aperiodic.*

Proposition 18 *Suppose X is a ψ-irreducible chain with period d and d-cycle $\{D_i, i = 1, ..., d\}$. Then each of the sets D_i is absorbing ψ-irreducible set for the chain X_d corresponding to the transition probability kernel P^d, and X_d on each D_i is aperiodic.*

4.3 Petite sets

Let $a = \{a(n)\}$ be a distribution on $\mathbb{Z}_+$, and consider the Markov chain

$$K_a(x,A) := \sum_{n=0}^{\infty} P^n(x,A)a(n), \quad x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X}).$$

It is obvious that K_a is a transition kernel, so that X_a is a well-defined by Theorem 2. We will call X_a the K_a-*chain*, with *sampling distribution* a. Probabistically, X_a has the interpretation of being the chain X sampled at times-points drawn successively according to the distribution a, or more accurately, at time-points of an independent renewal process with increment distribution a.

There are two specific sampled chains which we have already invoked, and which will be used frequently in the sequel. If $a = \delta_m$ then the K_{δ_m}-chain is the m-skeleton with transition kernel P^m. If $a_\varepsilon(n) = (1-\varepsilon)\varepsilon^n$, $n = 1, 2, ...$ then the kernel K_{a_ε} is the resolvent K_ε. The concept of sampled chains immediately enables to develop conditions under which one set is uniformly accessible form another. We say that a set $B \in \mathcal{B}(\mathcal{X})$ is *uniformly accessible using* a from another set $A \in \mathcal{B}(\mathcal{X})$ if there exists $\delta > 0$ such that

$$\inf_{x \in A} K_a(x, B) > \delta \tag{4.1}$$

and when (4.1) holds we write $A \overset{a}{\mapsto} B$.

Lemma 4 *If* $A \overset{a}{\mapsto} B$ *for some distribution* a *then* $A \mapsto B$.

Lemma 5 *1. If* a *and* b *are distributions on* $\mathbb{Z}_+$ *then the sampled chains with transitions laws* K_a *and* K_b *satisfy the generalized Chapman-Kolmogorov equations*

$$K_{a*b}(x, A) = \int K_a(x, dy) K_b(y, A).$$

2. If $A \overset{a}{\mapsto} B$ *and* $B \overset{b}{\mapsto} C$*, then* $A \overset{a*b}{\mapsto} C$.

3. If a *is a distribution on* $\mathbb{Z}_+$ *then the sampled chain with transition law* K_a *satisfies the relation*

$$U(x, A) \geqslant \int U(x, dy) K_a(y, A).$$

Definition 18 *We will call a set* $C \in \mathcal{B}(\mathcal{X})$ ε_a*-petite if the sampled chain satisfies the bound*

$$K_a(x, B) \geqslant \nu_a(B),$$

for all $x \in C$, $B \in \mathcal{B}(\mathcal{X})$*, where* ν_a *is a non trivial measure on* $\mathcal{B}(\mathcal{X})$.

Proposition 19 *If* $C \in \mathcal{B}(\mathcal{X})$ *is* ν_m*-small then* C *is* ν_{δ_m}*-petite.*

Proposition 20 *1. If* $A \in \mathcal{B}(\mathcal{X})$ *is* ν_a*-petite, and* $D \overset{b}{\mapsto} A$ *then* D *is* ν_{b*a}*-petite, where* ν_{b*a} *can be chosen as a multiple of* ν_a.

2. If X *is* ψ*-irreducible and if* $A \in \mathcal{B}^+(\mathcal{X})$ *is* ν_a*-petite, then* ν_a *is an irreducibility measure for* X.

Proposition 21 *Suppose* X *is* ψ*-irreducible.*

1. If A *is* ν_a*-petite, then there exists a sampling measure* b *such that* A *is also* ψ_b*-petite where* ψ_b *is a maximal irreducibility measure.*

2. The union of two petite sets is petite.

3. *There exists a sampling measure c, an everywhere positive, measurable function $s : \mathcal{X} \to \mathbb{R}$, and a maximal irreducibility measure ψ_c such that*

$$K_c(x, B) \geqslant s(x)\psi_c(B), \quad x \in \mathcal{X}, B \in \mathcal{B}(\mathcal{X}).$$

   Thus there is an increasing sequence $\{C_i\}$ of ψ_c-petite sets, all with the same sampling distribution c and minorizing measure equivalent to ψ, with $\bigcup C_i = \mathcal{X}$.

Proposition 22 *Suppose that X is ψ-irreducible and C is ν_a-petite.*

1. *Without loss of generality we can take a to be either a uniform sampling distribution $a_m(i) = 1/m$, $1 \leqslant i \leqslant m$, or a to be the geometric sampling distribution a_ε. In either case, there is a finite mean sampling time*

$$m_a = \sum_i i a(i).$$

2. *If X is strongly aperiodic then the set $C_0 \cup C_1 \subset \bar{\mathcal{X}}$ corresponding to C is ν_a^*-petite for the split chain $\bar{\mathsf{X}}$.*

Theorem 9 *If X is irreducible and aperiodic then every petite set is small.*

Chapter 5

Topology

5.1 Weak Feller Chains

Here we assume that $\mathcal{X}$ is equipped with a locally compact, separable, metrizable topology with $\mathcal{B}(\mathcal{X})$ as the Borel σ-field. Recall that a function h from $\mathcal{X}$ to $\mathbb{R}$ is *lower semicontinuous* if

$$\liminf_{y \to x} h(y) \geqslant h(x), \quad x \in \mathcal{X},$$

a typical, and frequently used, lower semicontinuous function is the indicator function $1_G(x)$ of an open set $G \in \mathcal{B}(\mathcal{X})$.

Definition 19 *1. If $P(\cdot, G)$ is a lower semicontinuous function for any open set $G \in \mathcal{B}(\mathcal{X})$, then P is called a weak Feller chain.*

2. If a is a sampling distribution and there exists a substochastic transition kernel T satisfying

$$K_a(x, A) \geqslant T(x, A), \quad x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X}),$$

where $T(\cdot, A)$ is a lower semicontinuous function for any $A \in \mathcal{B}(\mathcal{X})$, then T is called a continuous component of K_a.

3. If X is a Markov chain for which there exists a sampling measure a such that K_a possesses a continuous component T with $T(x, \mathcal{X}) > 0$ for all x, then X is called T-chain.

Recall that the transition probability kernel P acts on bounded functions through the mapping

$$Ph(x) = \int P(d, dy)h(y), \quad , x \in \mathcal{X}.$$

Suppose that $\mathcal{X}$ is a topological space and let us denote the class of bounded continuous function form $\mathcal{X}$ to $\mathbb{R}$ by $\mathcal{C}(\mathcal{X})$.

Definition 20 *The weak Feller property is frequently defined by requiring that the transition probability kernel P maps $\mathcal{C}(\mathcal{X})$ to $\mathcal{C}(\mathcal{X})$. If the transition probability kernel P maps all bounded measurable functions to $\mathcal{C}(\mathcal{X})$ then P (and also X) is called strong Feller.*

Proposition 23 *1. The transition kernel $P1_G$ is lower semicontinuous for every open set $G \in \mathcal{B}(\mathcal{X})$ if and only if P maps $\mathcal{C}(\mathcal{X})$ to $\mathcal{C}(\mathcal{X})$ and P maps all bounded measurable functions to $\mathcal{C}(\mathcal{X})$ if and only if the function $P1_A$ is lower semicontinuous for every set $A \in \mathcal{B}(\mathcal{X})$.*

2. *If the chain weak Feller then for any closed set $C \subset \mathcal{X}$ and any non-decreasing function $m : \mathbb{Z}_+ \to \mathbb{Z}_+$ the function $\mathbf{E}_x(m(\tau_C))$ is lower semicontinuous in x. Hence for any closed set $C \subset \mathcal{X}$, $r > 1$ and $n \in \mathbb{Z}_+$ the functions*
$$\mathbf{P}_x(\tau_C \geqslant n), \quad \mathbf{E}_x(\tau_C), \quad \text{and } \mathbf{E}_x r^{\tau_C}$$
are lower semicontinuous.

3. *If the chain is weak Feller then for any open set $G \subset \mathcal{X}$, the function $\mathbf{P}_x(\tau_G \leqslant n)$ and hence also the functions $K_a(x, G)$ and $L(x, G)$ are lower semicontinuous.*

5.2 Strong Feller Chains

Definition 21 1. *A point $x \in \mathcal{X}$ is called reachable if for every open set $G \in \mathcal{B}(\mathcal{X})$ containing x*
$$\sum_n P^n(y, G) > 0, \quad y \in \mathcal{X}.$$

2. *The chain X is called open set irreducible if every point is reachable.*

Lemma 6 *If X is ψ-irreducible then x^* is reachable if and only if $x^* \in \operatorname{supp}(\psi)$.*

Proposition 24 *If X is a strong Feller chain, and $\mathcal{X}$ contains one reachable point x^*, then X is ψ=irreducible, with $\psi(\cdot) = P(x^*, \cdot)$.*

Proposition 25 *If X is an open set irreducible strong Feller chain, then X is a ψ-irreducible chain.*

5.3 T-chains

Proposition 26 *If X is a T-chain, and X contains one reachable point x^*, then X is ψ-irreducible, with $\psi = T(x^*, \cdot)$.*

Proposition 27 *If X is an open set irreducible T-chain, then X is a ψ-irreducible T-chain.*

Proposition 28 *If an open ν_a-petite set A exists, then K_a possesses a continuous component non trivial on all A.*

Proposition 29 *Suppose that for each $x \in \mathcal{X}$ there exists a probability distribution a_x on $\mathbb{Z}_+$ such that K_{a_x} possesses a continuous component T_x which is non trivial at x. Then X is a T-chain.*

Theorem 10 *If every compact set is petite, then X is T-chain. Conversely, if X is a ψ-irreducible T-chain then every compact set is petite, and consequently if X is an open set irreducible T-chain then every compact set is petite.*

Proposition 30 *If X is a ψ-irreducible T-chain, then there is a sampling distribution b, an everywhere strictly positive, continuous function $s' : \mathcal{X} \to \mathbb{R}$, and a maximal irreducibility measure ψ_b such that*
$$K_b(x, a) \geqslant s'(x)\psi_b(B), \quad x \in \mathcal{X}, B \in \mathcal{B}(\mathcal{X}).$$

Lemma 7 *If X is a ψ-irreducible Feller chain, then the closure of every petite set is petite.*

Proposition 31 *Suppose that* X *is* ψ*-irreducible. Then all compact subsets of* X *are petite if either:*

1. X *has the Feller property and open* ψ*-positive petite set exists; or*

2. X *has the Feller property and* $\text{supp}(\psi)$ *has non-empty interior.*

Theorem 11 *If a* ψ*-irreducible chain* X *is weak Feller and if* $\text{supp}(\psi)$ *has nonempty interior then* X *is a* T*-chain.*

5.4 e-Chains

One possible way to think of Markov chain is that the transition function P gives rise to a deterministic map from $\mathcal{M}$ the space of probability measures on $\mathcal{B}(\mathcal{X})$, to itself, and we can construct on this basis a dynamical system $(P, \mathcal{M}, d)$, provided we specify a metric d, and hence also topology on $\mathcal{M}$

We recall that a sequence of probability measures $\{\mu_k : \ k \in \mathbb{Z}_+\} \subset \mathcal{M}$ converges *weakly* to $\mu_\infty \in \mathcal{M}$ (denoted $\mu_k \xrightarrow{w} \mu_\infty$) if

$$\lim_{k\to\infty} \int f d\mu_k = \int f d\mu_\infty$$

for every $f \in \mathcal{C}(\mathcal{X})$. Due to our restrictions on the state space $\mathcal{X}$, the topology of weak convergence is induced by a number of metrics on $\mathcal{M}$. One such metric may be expressed

$$d_m(\mu, \nu) = \sum_{k=0}^{\infty} |\int f_k d\mu - \int f_k d\nu| 2^{-k}, \ \ \mu, \nu \in \mathcal{M}$$

where $\{f_k\}$ is an appropriate set of functions in $\mathcal{C}_c(\mathcal{X})$, the set of continuous functions on $\mathcal{X}$ with compact support.

For $(P, \mathcal{M}, d_m)$ to be a dynamical system we require that P be a continuous map on $\mathcal{M}$. If P is continuous, then we must have in particular that if a sequence of point masses $\{\delta_{x_k} : \ k \in \mathbb{Z}_+\} \subset \mathcal{M}$ converge to some point mass $\delta_{x_\infty} \in \mathcal{M}$, then

$$\delta_{x_k} P \xrightarrow{w} \delta_{x_\infty} P \ \text{ as } k \to \infty$$

or equivalently $\lim_{k\to\infty} Pf(x_k) = Pf(x_\infty)$ for all $f \in \mathcal{C}(\mathcal{X})$. That is, if the Markov transition function induces a continuous map on $\mathcal{M}$ then Pf must be continuous of any bounded continuous function f. Conversely it is clear that for any weak feller Markov transition function P, the associated operator P on $\mathcal{M}$ is continuous.

Proposition 32 *The triple* $(P, \mathcal{M}, d_m)$ *is a dynamical system if and only if the Markov transition function* P *has the weak Feller property.*

Definition 22 *We say that* X *is an e-chain if for any* $f \in \mathcal{C}_c(\mathcal{X})$ *the sequence of functions* $\{P^k f : \ k \in \mathbb{Z}_+\}$ *is equicontinuous on compact sets.*

Proposition 33 *Suppose that the Markov chain* X *has the Feller property, and that there exists a unique probability measure* π *such that for every* x

$$P^n(x, \cdot) \xrightarrow{w} \pi.$$

Then X *is an e-chain.*

We say that the dynamical system $(P, \mathcal{M}, d_m)$ is called stable in the sense of Lyapunov if for each measure $\mu \in \mathcal{M}$,

$$\lim_{\nu \to \mu} \sup_{k \geqslant 0} d_m(\nu P^k, \mu P^k) = 0.$$

Proposition 34 *The Markov is an e-chain if and only if the dynamical system $(P, \mathcal{M}, d_m)$ is stable in the sense of Lyapunov.*

Stability in the sense of Lyapunov is a useful concept when a stationary point for the dynamical system exists. If x^* is stationary point and the dynamical systems is stable in the sense of Lyapunov, the trajectories which start near x^* will stay near x^*, and this turns out to be a useful notion of stability. For the dynamical system $(P, \mathcal{M}, d_m)$, a stationary point is an invariant probability: that is, a probability satisfying

$$\pi(A) = \int \pi(dx) P(x, A), \quad A \in \mathcal{B}(\mathcal{X}).$$

We say that $(P, \mathcal{M}, d_m)$ is Lagrange stable if, for every $\mu \in \mathcal{M}$, the orbit of measures μP^k is a precompact subset of $\mathcal{M}$.

Definition 23 *The Markov chain X is called bounded in probability if for each initial condition $x \in \mathcal{X}$ and each $\varepsilon > 0$, there exists a compact subset $C \subset \mathcal{X}$ such that*

$$\liminf_{k \to \infty} \mathbf{P}_x(X_k \in C) \geqslant 1 - \varepsilon.$$

Equivalently the family of measures $\{P(x, \cdot) : \ k \geqslant 1\}$ is tight.

Proposition 35 *The chain X is bounded in probability if and only if the dynamical system $(P, \mathcal{M}, d_m)$ is Lagrange stable.*

Note that the space $\mathcal{C}(\mathcal{X})$ can be viewed as a normed linear space, where we take the norm $|\cdot|_c$ to be defined for $f \in \mathcal{C}(\mathcal{X})$ as

$$|f|_c := \sum_{k=0}^{\infty} \sum_{k=0}^{\infty} 2^{-k} (\sup_{x \in C_k} |f(x)|)$$

where $\{C_k\}$ is a sequence of open precompact sets whose union is equal to $\mathcal{X}$. The associated metric d_c generates the topology of uniform convergence on compact subsets of $\mathcal{X}$. If P is a weak Feller kernel, then the mapping P on $\mathcal{C}(\mathcal{X})$ is continuous with respect to this norm and in this case the triple $(P, \mathcal{C}(\mathcal{X}), d_c)$ is a dynamical system.

Proposition 36 *Suppose that X is bounded in probability. Then X is an e-chain if and only if the dynamical system $(P, \mathcal{C}(\mathcal{X}), d_c)$ is Lagrange stable.*

Chapter 6

Transience and Recurrence

The idea is that for irreducible chains two sorts of behavior are expected: either samples eventually leave any bounded set with probability one or they will return infinitely many times to sets of positive measure.

Definition 24 *1. We say that a set A is uniformly transient if for there exists $M < \infty$ such that $\mathbf{E}_x(\eta_A) \leqslant M$ for all $x \in A$.*

2. The set A is called recurrent if $\mathbf{E}_x(\eta_A) = \infty$ for all $x \in A$.

We say that X is recurrent if every set in $\mathcal{B}_+(\mathcal{X})$ is recurrent. If there is a countable covering of $\mathcal{X}$ with uniformly transient sets then we say that X is transient.

6.1 Transience and recurrence on countable state space

Definition 25 *The state α is called transient if $\mathbf{E}_\alpha(\eta_\alpha) < \infty$, and recurrent if $\mathbf{E}_\alpha(\eta_\alpha) = \infty$.*

Proposition 37 *When $\mathcal{X}$ is countable and X is irreducible, either $U(x,y) = \infty$ for all $x, y \in \mathcal{X}$ or $U(x,y) < \infty$ for all $x, y \in \mathcal{X}$.*

Definition 26 *1. If every state is transient the chain is called transient.*

2. If every state is recurrent the chain is called recurrent.

Proposition 38 *For any $x \in \mathcal{X}$, $U(x,x) = \infty$, if and only if $L(x,x) = 1$.*

Proposition 39 *When X is irreducible, either $L(x,y) = 1$ for all $x, y \in \mathcal{X}$ or $L(x,x) < 1$ for all $x \in \mathcal{X}$.*

6.2 Transience and recurrence for individual sets

We recall that for general $A, B \in \mathcal{B}(\mathcal{X})$ the taboo probabilities are of the form

$$P_A^n(x,B) = \mathbf{P}_x(X_n \in B, \tau_A \geqslant n),$$

and by convention $P_A^0(x,A) = 0$.

Remark 1 *There are two basic decompositions of $P^n(x,\cdot)$ over the first return to $A \in \mathcal{B}(\mathcal{X})$:*

1. *first entrance decomposition*

$$P^n(x,B) = P^n_A(x,B) + \sum_{j=1}^{n-1} \int_A P^j_A(x,dw)P^{n-j}(w,B).$$

2. *last exit decomposition*

$$P^n(x,B) = P^n_A(x,B) + \sum_{j=1}^{n-1} \int_A P^j(x,dw)P^{n-j}_A(w,B),$$

To analyze the transience and recurrence behavior we need

$$U^{(z)}(x,B) := \sum_{n=1}^{\infty} P^n(x,B)z^n, \quad |z|<1;$$
$$U^{(z)}_A(x,B) := \sum_{n=1}^{\infty} P^n_A(x,B)z^n, \quad |z|<1.$$

Th kernel U has the property

$$U(x,A) = \sum_{n=1}^{\infty} P^n(x,A) = \lim_{z\to 1} U^{(z)}(x,A)$$

and as in the countable case, for any $x \in \mathcal{X}$, $A \in \mathcal{B}(\mathcal{X})$

$$\mathbf{E}_x(\eta_A) = U(x,A).$$

The return probabilities $L(x,A) = P_x(\tau_A < \infty)$ satisfy

$$L(x,A) = \sum_{n=1}^{\infty} P^n_A(x,A) - \lim_{z\to 1} U^{(z)}_A(x,A).$$

Exploiting the convolution forms in first entrance last exit decompositions we obtain

$$U^{(z)}(x,B) = U^{(z)}_A(x,B) + \int_A U^{(z)}_A(x,dw)U^{(z)}(w,B),$$
$$U^{(z)}(x,B) = U^{(z)}_A(x,B) + \int_A U^{(z)}(x,dw)U^{(z)}_A(w,B).$$

6.3 The recurrence and transience for chains with an atom

Theorem 12 *Suppose that X is ψ-irreducible and admits an atom α $\alpha \in \mathcal{B}_+(\mathcal{X})$. Then*

1. *if α is recurrent then every set in $\mathcal{B}_+(\mathcal{X})$ is recurrent.*

2. *if α is transient, then there is a countable covering of $\mathcal{X}$ by uniformly transient sets.*

Definition 27 *If $A \in \mathcal{B}(\mathcal{X})$ can be covered with a countable number of uniformly transient sets, then we call A transient.*

6.4 The general recurrence/transience dichotomy

Definition 28 *1. The chain X is called recurrent if it is ψ-irreducible and $U(x, A) = \infty$ for every $A \in \mathcal{B}_+(\mathcal{X})$.*

2. The chain X is called transient if it is ψ-irreducible and $\mathcal{X}$ is transient.

Proposition 40 *Suppose that X is ψ-irreducible and strongly aperiodic.Then either both X, $\bar{\mathsf{X}}$ are recurrent, or both X and $\bar{\mathsf{X}}$ are transient.*

Lemma 8 *For any $0 < \varepsilon < 1$ the following identity holds:*

$$\sum_{n=1}^{\infty} K_\varepsilon^n = \frac{1-\varepsilon}{\varepsilon} \sum_{n=0}^{\infty} P^n.$$

Theorem 13 *If X is ψ-irreducible, then X is either recurrent or transient.*

Theorem 14 *Suppose that X is ψ-irreducible and aperiodic.*

1. *The chain X is transient if and only if one and then very m-skeleton is transient.*

2. *The chain X is recurrent if and only if one and there very m-skeleton is recurrent.*

6.5 Transience and recurrence

Proposition 41 *Suppose that X is a Markov chain (not necessarily irreducible)*

1. *If any set $A \in \mathcal{B}(\mathcal{X})$ is uniformly transient with $U(x, A) \leqslant M$ for $x \in A$, then $U(x, A) \leqslant 1 + M$ for every $x \in \mathcal{X}$.*

2. *If any set $A \in \mathcal{B}(\mathcal{X})$ satisfies $L(x, A) = 1$ for all $x \in A$, then A is recurrent. If X is ψ irreducible, then $A \in \mathcal{B}^+(\mathcal{X})$ and we have $U(x, A) \equiv \infty$ for $x \in \mathcal{X}$.*

3. *If any set $A \in \mathcal{B}(\mathcal{X})$ satisfies $L(x, A) \leqslant \varepsilon < 1$, $x \in A$, then we have $U(x, A) \leqslant 1/(1-\varepsilon)$ for $x \in \mathcal{X}$, so that in particular A is uniformly transient.*

4. *Let $\tau_A(k)$ denote the k-th return time to A, and suppose that for some m*

$$\mathbf{P}_x(\tau_A(m) < \infty) \leqslant \varepsilon < 1, \quad x \in A;$$

then $U(x, A) \leqslant 1 + m/(1-\varepsilon)$ for very $x \in \mathcal{X}$.

Proposition 42 *If A is uniformly transient and $B \overset{a}{\mapsto} A$ for some a, then B is uniformly transient. Hence if A is uniformly transient, there is a countable covering of $\bar{A}$ by uniformly transient sets.*

Proposition 43 *Suppose D^c is absorbing and $L(x, D^c) > 0$ for all $x \in D$. Then D is transient.*

6.6 Transient sets

We recall that

Theorem 15 *If* X *is* ψ*-irreducible, then* X *is either recurrent or transient.*

Theorem 16 *If* X *is* ψ*-irreducible and transient then every petite set is uniformly transient.*

Theorem 17 *Suppose that* X *is* ψ*-irreducible. Then*

1. X *is recurrent if there exists some petite set* $C \in \mathcal{B}(\mathcal{X})$ *such that* $L(x, C) = 1$*, for all* $x \in C$*.*

2. X *is transient if and only if there exist two sets* $D, C \in \mathcal{B}_+(\mathcal{X})$ *with* $L(x, C) < 1$ *for all* $x \in D$*.*

Proposition 44 *If* X *is* ψ*-irreducible then every* ψ*-null set is transient.*

Proposition 45 *Suppose that* X *is a* ψ*-irreducible Markov chain on* $(\mathcal{X}, \mathcal{B}(\mathcal{X})$*. Then there exists sets* $D_1, ..., D_d \in \mathcal{B}(\mathcal{X})$ *such that*

1. *for* $x \in D_i$*,* $P(x, D_i) = 1$*,* $i = 0, ..., d-1 \pmod d$

2. *the set* $N = (\bigcup_{i=1}^d D_i)^c$ *is* ψ*-null and transient.*