1 General State Space MC Theory

1.1 Introduction

Markov chain theory is usually introduced in a course on stochastic processes under the restriction that the state space is countable. In order to apply Markov chain theory to most Bayesian problems (to sample from the posterior) we require a generalization of this theory to general state spaces. Usually we are interested in Euclidean space (e.g. $\mathbb{R}$ or $\mathbb{R}^d$ or a relevant subspace), however not much more effort is required to work in non-topological spaces.

Markov chains are constructed from a transition kernel P. A transition kernel is a conditional probability measure such that $\Phi_n \sim P(\cdot \mid \Phi_{n-1})$. The chains in MCMC settings have a strong stability property. A stationary probability distribution exists by construction of the chain. That is a distribution π exists such that for some N, if $\Phi_n \sim \pi \implies \Phi_{n+1} \sim \pi, \forall\, n > N$, if P allows moves over the entire state space. This freedom is called irreducibility. Irreducibility ensures that most MCMC chains are recurrent (i.e. the average number of visits to an arbitrary set B is infinite) or Harris recurrent (the probability of an infinite number of returns to B is 1). Harris recurrence ensures that the chain has the same limiting behavior for every starting value instead of almost every starting value. This is quite important as in practice we start chains from an arbitrary point x_0: a set of measure zero (under a continuous dominating measure).

The stationary distribution, π, is also a limiting distribution. That is, $\pi(\Phi_n) \to \pi$ under the total variation norm, independently of the initial state, x_0. Stronger forms of convergence also appear in MCMC theory, such as geometric and uniform convergence. A consequence of this convergence is that

$$\lim_{N \uparrow \infty} \left(\frac{1}{N} \sum_{n=1}^{N} h(\Phi_n) \right) = \mathbb{E}_\pi[h(\Phi)], \quad a.s.$$

If the chain is time reversible (i.e. symmetric w.r.t. time), then a Central Limit Theorem holds. We begin with some definitions in the theory of stochastic processes.

1.2 Stochastic Processes

Let $\mathcal{X}$ denote a general space and $\mathcal{B}(\mathcal{X})$ denote a σ-algebra of subsets of $\mathcal{X}$. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ be a measurable space.

Definition 1 (Random Variable) *A mapping $X : \Omega \to \mathcal{X}$ is called a random variable if*

$$X^{-1}(A) := \{\omega : X(\omega) \in A\} \in \mathcal{F}, \quad \forall A \in \mathcal{B}(\mathcal{X}).$$

Note that this mapping induces a probability measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Specifically

$$\Pr(X \in A) := P(X^{-1}(A) \in \mathcal{F}).$$

We will write $P(X \in A) \equiv \Pr(X \in A)$ although technically the probability P on $(\Omega, \mathcal{F})$ and the probability measure induced by the mapping are different.

Definition 2 (Stochastic Process) *A stochastic process with state space $\mathcal{X}$ is a collection of random variables indexed by a set T, $\mathbf{\Phi} = \{\Phi(t) : t \in T\}$. That is for $A \in \mathcal{B}(\mathcal{X})$, $P(\Phi(t) \in A) := P(\{\omega : \Phi(\omega, t) \in A\} \in \mathcal{F}) = P(\Phi^{-1}(A, t) \in \mathcal{F})$.*

Definition 3 (Discrete Time Stochastic process) *A discrete time stochastic process is a stochastic process where T is the non-negative integers $\mathbb{N}_+$.*

We will only be concerned with discrete time stochastic processes and will denote $\Phi(t) \equiv \Phi_t$.

We can also think of the whole of a discrete time stochastic process $\mathbf{\Phi}$ as an entity in its own right, called sample paths or realizations of the process, lying in the product space $\mathcal{X}^\infty = \prod_{i=0}^{\infty} \mathcal{X}_i$, where $\mathcal{X}_i \equiv \mathcal{X}$, each equipped with $\mathcal{B}(\mathcal{X})$.

Definition 4 (Countable State Space) *The state space $\mathcal{X}$ is called countable, if $\mathcal{X}$ is discrete with a finite or countable number of elements. $\mathcal{B}(\mathcal{X})$ is the σ-algebra generated by all subsets of $\mathcal{X}$.*

Definition 5 (Topological State Space) *The state space $\mathcal{X}$ is called topological if it equipped with a locally compact, separable, metrizable topology, then $\mathcal{B}(\mathcal{X})$ is the Borel σ-algebra (the smallest σ-algebra generated by the open sets).*

Definition 6 (General State Space) *The state space $\mathcal{X}$ is call general if it is equipped with a countably generated σ-algebra $\mathcal{B}(\mathcal{X})$.*

For the most part, we will be concerned mainly with general state spaces.

1.3 Probability Transition Kernels

A probability transition kernel plays the same role in general state space Markov chains as the probability transition matrix plays in countable state spaces. Let $\mathbb{R}_+$ denote the non-negative real numbers.

Definition 7 (Probability Transition Kernel) *A probability transition kernel is a function, P such that*

1. $\forall\, x \in \mathcal{X}$, $P(x, \cdot)$ *is a probability measure on* $\mathcal{B}(\mathcal{X})$*; in this case* $P(x, \cdot) : \mathcal{B}(\mathcal{X}) \to [0, 1]$
2. $\forall\, A \in \mathcal{B}(\mathcal{X})$, $P(\cdot, A)$ *is a non-negative measurable function on* $\mathcal{X}$*.*

When $\mathcal{X}$ is discrete, the transition kernel reduces to a transition matrix with elements

$$P(x, y) = \Pr(\Phi_n = y \mid \Phi_{n-1} = x), \quad x, y \in \mathcal{X}.$$

As a consequence of the Radon-Nikodym theorem, when $P(x, \cdot)$ is absolutely continuous w.r.t. some measure ν, there exists a function f, such that

$$P(x, A) = \Pr(\Phi_{i+1} \in A \mid \Phi_i = x) = \int_A f d\nu.$$

The function f is called the density of P w.r.t. ν and is commonly written as $f = dP/d\nu$.

Definition 8 (Markov Chain) *Given a probability transition kernel P, $\mathbf{\Phi} = \{\Phi_0, \Phi_1, \dots\}$ is a Markov chain if*

$$\Pr(\Phi_{i+1} \in A \mid \Phi_0, \dots, \Phi_i; \Phi_i = x) = \Pr(\Phi_{i+1} \in A \mid \Phi_i = x) = P(x, A).$$

If

$$\Pr(\Phi_{i+1} \in A_1, \dots, \Phi_{i+k} \in A_k \mid \Phi_i = x) = \Pr(\Phi_1 \in A_1, \dots, \Phi_k \in A_k \mid \Phi_0 = x)$$

for all i and k, then the Markov chain is said to be homogeneous.

1.3.1 The n-step probability transition kernel

Let $P^0(x, A) = \delta_x(A)$, the Dirac measure defined by $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ if $x \in A^c$. Then for $n \geq 1$ define

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

P^n is called then n-step probability transition kernel.

Theorem 1 (Chapman-Kolmogorov Equations) *For any integer $m \in [0, n]$,*

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

In words, the Chapman-Kolmogorov Equations state, as $\mathbf{\Phi}$ moves from x into A in n steps it must take some value $y \in \mathcal{X}$ at some intermediate time m. Since Φ is a Markov chain, it forgets its past at m and moves $n - m$ steps with the appropriate law starting at y.

We will often used the following definition:

$$P_x(\Phi_n \in A) := P^n(x, A)$$

and so the Chapman-Kolmogorov Equations can be written

$$P_x(\Phi_n \in A) = \int_{\mathcal{X}} P_x(\Phi_m \in dy) P_y(\Phi_{n-m} \in A).$$

Also, we will let μ denote the initial distribution of the chain so that

$$\begin{aligned}
&P_\mu(\Phi_0 \in A_0, \Phi_1 \in A_1, \ldots, \Phi_n \in A_n) \\
&= \int_{y_0 \in A_0} \Pr(\Phi_1 \in A_1, \ldots, \Phi_n \in A_n \mid y_0) \mu(dy_0) \\
&= \int_{y_0 \in A_0} \int_{y_1 \in A_1} \Pr(\Phi_2 \in A_2, \ldots, \Phi_n \in A_n \mid y_1) \mu(dy_0) P(y_0, dy_1) \\
&\qquad\qquad \vdots \\
&= \int_{y_0 \in A_0} \cdots \int_{y_n \in A_n} \mu(dy_0) P(y_0, dy_1) \cdots P(y_{n-1}, dy_n).
\end{aligned}$$

The m-step chain $\mathbf{\Phi}^m = \{\Phi_n^m\}$ is a sub-chain of the original chain $\mathbf{\Phi}$ with transition probabilities

$$P_x(\Phi_n^m \in A) = P^{mn}(x, A).$$

Definition 9 (Skeletons and Resolvents) *The chain $\mathbf{\Phi}^m$ with transition law $P_x(\Phi_n^m \in A) = P^{mn}(x, A)$ is called the m-skeleton of the chain Φ.*

The resolvent K_ϵ, $\epsilon \in (0, 1)$, is defined by

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

The Markov chain with probability transition kernel K_ϵ is the called the K_ϵ-chain.

Given an initial distribution μ for the chain Φ, the K_ϵ-chain is a sub-chain of Φ. The indices of the K_ϵ-chain are generated from a geometric distribution with parameter $1-\epsilon$. K_ϵ chains enjoy much stronger regularity than the original chain and can be used to establish many properties of the original chain.

Proposition 1 (The (weak) Markov property) *If $\mathbf{\Phi}$ is a Markov chain with initial measure μ and $h : \mathcal{X}^\infty \to \mathbb{R}$ is a bounded, measurable function, then*

$$\mathbb{E}_\mu[h(\Phi_{n+1}, \Phi_{n+2}, \dots) \mid \Phi_0, \dots, \Phi_n; \Phi_n = x] = \mathbb{E}_x[h(\Phi_1, \Phi_2, \dots)].$$

Note here that when h is the indicator function, this is just the definition of a Markov chain.

1.3.2 Occupation, Hitting and Stopping Times

The analysis of Markov chains concern its behavior (distributions) at certain random times in its evolution, which we define now.

Definition 10 (Occupation Times, Return Times and Hitting Times) *Let $\mathbf{\Phi}$ be a Markov chain and for any $A \in \mathcal{B}(\mathcal{X})$*

(i) The occupation time η_A is the number of visits by $\mathbf{\Phi}$ to A after time zero:

$$\eta_A := \sum_{n=1}^{\infty} \mathbb{I}_A\{\Phi_n\}.$$

(ii) The first return by $\mathbf{\Phi}$ to A after time zero is defined by

$$\tau_A := \min\{n \geq 1 : \Phi_n \in A\}.$$

(iii) The first hitting time on A by $\mathbf{\Phi}$ is

$$\sigma_A := \min\{n \geq 0 : \Phi_n \in A\}.$$

(iv) A function $\zeta : \mathcal{X}^\infty \to \mathbb{N}_+ \cup \{\infty\}$ is a stopping time for $\mathbf{\Phi}$ if for any initial distribution μ the event $\{\zeta = n\}$ is ν-measurable where ν is the σ-algebra induced by $\{\Phi_i\}_0^n$.

Note that η_A, τ_A and σ_A are all measurable functions from $\mathcal{X}^\infty$ to $\mathbb{N}_+ \cup \{\infty\}$ and that τ_A and σ_A are examples of stopping times.

Analysis of the stability properties of $\mathbf{\Phi}$ involves the kernel $U : \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to \mathbb{R}_+ \cup \{\infty\}$, defined by

$$U(x, A) := \mathbb{E}_x(\eta_A) = \mathbb{E}_x\left(\sum_{n=1}^{\infty} \mathbb{I}_A(\Phi_n)\right) = \sum_{n=1}^{\infty} P^n(x, A), \quad x \in \mathcal{X},\ A \in \mathcal{B}(\mathcal{X})$$

and the return time probabilities

$$L(x, A) := P_x(\tau_A < \infty) = P_x(\mathbf{\Phi} \text{ ever enters } A).$$

Proposition 2 *Let $\mathbf{\Phi}$ be a Markov chain with probability transition kernel $P(x, A)$ and $n \in \mathbb{N}_+$.*

(i) For all $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$

$$P_x(\tau_A = 1) = P(x, A)$$

and for $n > 1$

$$\begin{aligned} P_x(\tau_A = n) &= \int_{A^c} P(x, dy) P_y(\tau_A = n - 1) \\ &= \int_{A^c} P(x, dy_1) \int_{A^c} P(y_1, dy_2) \cdots \int_{A^c} P(y_{n-2}, dy_{n-1}) P(y_{n-1}, A). \end{aligned}$$

(ii) For all $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$

$$P_x(\sigma_A = 0) = \mathbb{I}_A(x)$$

and for $n > 0$ and $x \in A^c$

$$P_x(\sigma_A = n) = P_x(\tau_A = n).$$

Furthermore,

$$L(x, A) = P_x(\tau_A < \infty) = \sum_{n=1}^{\infty} P_x(\tau_A = n).$$

For a stopping time ζ the property that tells us that the future evolution of $\mathbf{\Phi}$ after the stopping time depends only on the value of Φ_ζ is called the strong Markov property.

Proposition 3 (The Strong Markov Property) *We say that $\mathbf{\Phi}$ has the strong Markov property if for any initial distribution μ and bounded measurable function $h : \mathcal{X}^\infty \to \mathbb{R}$ and for any stopping time ζ which is finite almost surely,*

$$\mathbb{E}_\mu[h(\Phi_{\zeta+1}, \Phi_{\zeta+2}, \dots) \mid \Phi_0, \dots \Phi_\zeta; \Phi_\zeta = x_\zeta] = \mathbb{E}_{x_\zeta}[h(\Phi_1, \Phi_2, \dots)]$$

What is the difference between the weak and strong Markov properties?

1.4 Irreducibility

The first concept of "stochastic stability" that we study is that of irreducibility. It is a first measure of the sensitivity of the Markov chain to the initial distribution μ or initial state x_0. It is crucial in Markov chain Monte Carlo as it guarantees convergence no matter what the initial conditions. Thus, we avoid the need of a detailed study of the probability transition kernel in order to determine acceptable initial conditions (to guarantee convergence).

Recall that if $\mathcal{X}$ is countable, then for any two distinct states x, $y \in \mathcal{X}$ we say that x leads to y (written $x \to y$) if $L(x, y) > 0$ and that they communicate (written $x \leftrightarrow y$) if $L(x, y) > 0$ and $L(y, x) > 0$. By convention, $x \to x$.

Proposition 4 *The communication relation "$\leftrightarrow$" is an equivalence relation. The equivalence classes $C(x) = \{y : x \leftrightarrow y\}$ form a partition of the state space $\mathcal{X}$.*

Definition 11 (Irreducible Spaces and Absorbing Sets) *Let $\mathcal{X}$ be a countable state space for a Markov chain $\mathbf{\Phi}$.*

(i) If $\exists\, x \ni C(x) = \mathcal{X}$, then we say that $\mathcal{X}$, or the chain $\mathbf{\Phi}$, is irreducible.

(ii) If $P(y, C(x)) = 1\, \forall\, y \in C(x)$, then $C(x)$ is called an absorbing set.

The problem with the communication relation "$\leftrightarrow$" is that it does not hold in general state spaces. Consider the case when $\mathcal{X} = \mathbb{R}$ and $P(x, A) = \Phi(A; 0, 1)$. Then it is obvious that $L(x, y) = 0$ for all singletons, y. We could consider a Borel measurable set A and determine whether or not $L(x, A) > 0$. If yes, then $x \to A$, but $L(A, x) = 0$, so A does not lead to x. In general state spaces, therefore, we cannot say in general whether we return to single states x.

Therefore, we define an analog to irreducibility. The analog that forms the basis for modern general state space analysis is φ-irreducibility.

1.4.1 φ- and ψ-irreducibility

Definition 12 (φ-irreducibilty) *A Markov chain $\mathbf{\Phi}$ is φ-irreducible if there exists a measure φ on $\mathcal{B}(\mathcal{X})$ such that, for all $x \in \mathcal{X}$ and for all $A \in \mathcal{B}(\mathcal{X})$, $L(x, A) > 0$ whenever $\varphi(A) > 0$.*

There are several other equivalent definitions of φ-irreducibility that prove useful.

Proposition 5 *The following are equivalent definitions of φ-irreducibility: A Markov chain $\mathbf{\Phi}$ is φ-irreducible if there exists a measure φ on $\mathcal{B}(\mathcal{X})$ such that, for all $x \in \mathcal{X}$ and for all $A \in \mathcal{B}(\mathcal{X})$ with $\varphi(A) > 0$*

(i) $L(x, A) > 0 \quad \Longleftrightarrow$

(ii) $P^n(x, A) > 0$, *for some $n > 0$, possibly depending on x and A* $\quad \Longleftrightarrow$

(iii) $U(x, A) > 0 \quad \Longleftrightarrow$

(iv) $K_{1/2}(x, A) > 0$.

Proof:
$(i) \implies (ii)$. $0 < L(x, A) = P_x(\tau_A < \infty)$. This implies $\exists$ an $n < \infty \ni P_x(\tau_A = n) > 0$.

$$\begin{aligned} 0 < P_x(\tau_A = n) &= \int_{A^c} P(x, dy_1) \int_{A^c} P(y_1, dy_2) \cdots \int_{A^c} P(y_{n-2}, dy_{n-1}) P(y_{n-1}, A) \\ &\leq \int_{\mathcal{X}} P(x, dy_1) \int_{\mathcal{X}} P(y_1, dy_2) \cdots \int_{\mathcal{X}} P(y_{n-2}, dy_{n-1}) P(y_{n-1}, A) \\ &= \int_{\mathcal{X}} P(x, dy_1) \int_{\mathcal{X}} P(y_1, dy_2) \cdots P^2(y_{n-2}, A) \\ &= P^n(x, A). \end{aligned}$$

$(ii) \implies (i)$. Suppose not. Suppose $L(x, A) = 0$ for all $x \in \mathcal{X}$. This implies that $P_x(\tau_A = n) = 0$ for all x and $n \geq 1$. In particular $P_x(\tau_A = 1) = P(x, A) = 0$. But $P^n(x, A) = \int_{\mathcal{X}} P^{n-1}(x, dy) P(y, A) = 0 \quad \Rightarrow\Leftarrow$.
$(ii) \implies (iii)$. $U(x, A) = \sum_{n=1}^{\infty} P^n(x, A)$, so that if there exists an $n > 0$, such that $P^n(x, A) > 0$, then $U(x, A) > 0$.
$(iii) \implies (ii)$. If $U(x, A) > 0$, then there is some $n > 0$ such that $P^n(x, A) > 0$.
$(ii) \implies (iv)$. $K_{1/2}(x, A) = 1/2 \sum_{n=0}^{\infty} P^n(x, A)(1/2)^n > 0$.
$(iv) \implies (ii)$. Suppose $P^n(x, A) = 0$ for some $A \in \mathcal{B}(\mathcal{X})$ and all $n > 0$. Then $K_{1/2}(x, A) = \frac{1}{2}\mathbb{I}_A(x)$. In particular, let $x \in A^c$, then $K_{1/2}(x, A) = 0$. $\Rightarrow\Leftarrow$. $\square$

The assumption of φ-irreducibility precludes many obvious forms of reducible behavior. The definition guarantees that sets of positive φ-measure are reached with positive probability, regardless of the starting value. Therefore the chain cannot break up into separate pieces.

However, what we would also like is the reverse implication, that sets B such that $\varphi(B) = 0$ are avoided with probability one from most starting points. Instead of trying to restrict

an irreducibility measure to sets of non-zero measure, it has been more fruitful to extend φ-irreducibility to a "maximal" irreducibility measure ψ. ψ is maximal in the sense that if ϕ satisfies Definition 12, then $\phi \prec \psi$. That is ψ dominates ϕ or ϕ is absolutely continuous with respect to ψ: if $\psi(A) = 0$, then $\phi(A) = 0$ (alternatively if $\phi(A) > 0$, then $\psi(A) > 0$).

Proposition 6 *If $\mathbf{\Phi}$ is φ-irreducible for some measure φ, then there exists a probability measure ψ on $\mathcal{B}(\mathcal{X})$ such that*

(i) $\mathbf{\Phi}$ *is ψ-irreducible;*

(ii) for any other measure ξ, the chain $\mathbf{\Phi}$ is ξ-irreducible if and only if $\xi \prec \psi$.

(iii) if $\psi(A) = 0$, then $\psi\{y : L(y, A) > 0\} = 0$.

(iv) the probability measure ψ is equivalent to

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

for any finite irreducibility measure ϕ'. Equivalent means $\psi' \prec \psi$ and $\psi \prec \psi'$.

Proof:
(i). First note that any probability measure that is equivalent to the irreducibility measure φ is itself an irreducibility measure, we can assume without loss of generality that $\varphi(\mathcal{X}) = 1$. Consider the probability measure ψ on $\mathcal{B}(\mathcal{X})$ defined by

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

To show that $\mathbf{\Phi}$ is ψ-irreducible we use the sets

$$\bar{A}(k) = \left\{ y : \sum_{n=1}^{k} P^n(y, A) > k^{-1} \right\}.$$

Note that $\bar{A}(1) \subseteq \bar{A}(2) \subseteq \cdots$. Therefore $\cup_{k=1}^{\infty} \bar{A}(k) = \lim_{k\uparrow\infty} \bar{A}(k)$. Call this limit $\bar{A}_L$. Now $\mathcal{X} \setminus \bar{A}_L = \{y : \sum_{n=1}^{\infty} P^n(y, A) = 0\}$. Let $\psi(A) > 0$ for $A \in \mathcal{B}(\mathcal{X})$. Then

$$\begin{aligned} 0 < \psi(A) = \int_{\mathcal{X}} K_{1/2}(y, A)\varphi(dy) &= \int_{\bar{A}_L} K_{1/2}(y, A)\varphi(dy) + \int_{\bar{A}_L^c} K_{1/2}(y, A)\varphi(dy) \\ &= \int_{\bar{A}_L} K_{1/2}(y, A)\varphi(dy) + \int_{\bar{A}_L^c} \frac{1}{2}\mathbb{I}_A(y)\varphi(dy) \\ &= \int_{\bar{A}_L} K_{1/2}(y, A)\varphi(dy) + \frac{1}{2}\varphi(A \cap \bar{A}_L^c). \end{aligned}$$

Now we have two cases: if $\varphi(A) > 0$ then $U(x, A) > 0$ for all $x \in \mathcal{X}$. Suppose $\varphi(A) = 0$. Then

$$\frac{1}{2}\varphi(A \cap \bar{A}_L^c) < \varphi(A) = 0$$

and

$$\begin{aligned} 0 < \psi(A) &= \int_{\bar{A}_L} K_{1/2}(y, A)\varphi(dy) \\ &= \int_{\bar{A}(1)} K_{1/2}(y, A)\varphi(dy) + \sum_{k=2}^{\infty} \int_{\bar{A}(k)\setminus\bar{A}(k-1)} K_{1/2}(y, A)\varphi(dy). \end{aligned}$$

Now, there exists a k such that $0 < \varphi(\bar{A}(k)\setminus\bar{A}(k-1)) < \varphi(\bar{A}(k))$. Hence, by φ-irreducibility, for a fixed x there is an m such that $P^m(x, \bar{A}(k)) > 0$ and

$$U(x, A) > \sum_{n=1}^{k} P^{m+n}(x, A) = \int_{\mathcal{X}} P^m(x, dy) \left(\sum_{n=1}^{k} P^n(y, A) \right) > k^{-1} P^m(x, \bar{A}(k)) > 0,$$

which establishes ψ-irreducibility.
(ii) $(\Rightarrow)$. Let ξ be such that $\mathbf{\Phi}$ is ξ-irreducible. Then $\xi(A) > 0$ implies $K_{1/2}(y, A) > 0$ implies $\psi(A) > 0$ (by its definition), therefore $\xi \prec \psi$.
$(\Leftarrow)$. Now suppose $\xi \prec \psi$ and $\mathbf{\Phi}$ is ψ-irreducible. If $\xi(A) > 0$, then $\psi(A) > 0$ and so $K_{1/2}(x, A) > 0$ which establishes ξ-irreducibility.
(iii). For any m we have

$$\begin{aligned} \int_{\mathcal{X}} P^m(y, A)2^{-m}\psi(dy) &= \int_{\mathcal{X}} P^m(y, A)2^{-m} \left(\int_{\mathcal{X}} \sum_{n=0}^{\infty} P^n(z, dy)2^{-(n+1)}\varphi(dz) \right) \\ &= \int_{\mathcal{X}} \sum_{n=0}^{\infty} 2^{-(n+m+1)} \left(\int_{\mathcal{X}} P^m(y, A)P^n(z, dy) \right) \varphi(dz) \\ &= \int_{\mathcal{X}} \sum_{n=0}^{\infty} P^{m+n}(z, A)2^{-(n+m+1)}\varphi(dz) < \psi(A). \end{aligned}$$

So, when $\psi(A) = 0$, $P^m(y, A) = 0$ a.e. $[\psi]$. This then implies that $L(y, A) = 0$ a.e. $[\psi]$, which means that $\psi\{y : L(y, A) > 0\} = 0$.
(iv). Follows from the definition of ψ above and the fact that any two maximal irreducibility measures are equivalent. □

Proposition 6(iii), now guarantees that we avoid "negligible" sets B with $\psi(B) = 0$, with probability one irrespective of the initial state. And by Proposition 6(i) for sets B such that $\psi(B) > 0$, we have that $L(y, B) > 0$ so we have positive probability of reaching any non-negligible set B from any initial state.

Definition 13 (ψ-irreducible) *A Markov chain is called ψ-irreducible if it is φ-irreducible for some measure φ and the measure ψ is a maximal irreducible measure satisfying the conditions of Proposition 6. We write*

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

for the sets of positive ψ-measure. By the equivalence of maximal irreducibility measures, $\mathcal{B}^+(\mathcal{X})$ is unique defined.

Definition 14 (Full and Absorbing Sets) *A set $A \in \mathcal{B}(\mathcal{X})$ is said to be*

(i) full if $\psi(A^c) = 0$.

(ii) absorbing if $P(x, A) = 1$ for $x \in A$.

Definition 15 (Accessibility) *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 all $x \in A$.*

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

$$\inf_{x \in A} L(x, B) \geq \delta. \tag{1}$$

When (1) holds, we write $A \rightsquigarrow B$.

1.5 Pseudo-atoms and Small Sets

1.5.1 Psuedo-atoms

Most of Markov chain theory on a general state space can be develop analogously to the countable state space case when $\mathcal{X}$ contains an atom for the chain $\mathbf{\Phi}$.

Definition 16 *A set $\boldsymbol{\alpha} \in \mathcal{B}(\mathcal{X})$ is called an atom for $\mathbf{\Phi}$ if there exists a measure ν on $\mathcal{B}(\mathcal{X})$ such that*

$$P(x, A) = \nu(A), \quad \forall\, x \in \boldsymbol{\alpha},\ \forall\, A \in \mathcal{B}(\mathcal{X}).$$

If $\mathbf{\Phi}$ is ψ-irreducible and $\psi(\boldsymbol{\alpha}) > 0$, then $\boldsymbol{\alpha}$ is called an accessible atom.

Singletons are always atoms. When $\mathcal{X}$ is countable and the chain is ψ-irreducible then every point is an accessible atom. However, on general state spaces, accessible atoms may not exist.

For the random walk on $\mathbb{R}$ defined by

$$\Phi_k = \Phi_{k-1} + W_k, \quad k > 0,$$

where $W_k \sim N(0,1)$, $\forall k > 0$ and Φ_0 is chosen from an arbitrary distribution, accessible atoms do not exist.

So, if this simple example does not contain accessible atoms, then what is the point in developing a theory for general state spaces that contain atoms? It seems that the class of general state spaces that contain accessible atoms is too small to be of practical importance. However, there exists "artificial atoms" for φ-irreducible chains. These atoms are found for "strongly aperiodic" chains by construction. We will construct a "split chain" $\check{\mathbf{\Phi}}$ evolving on a split state space $\check{\mathcal{X}} = \mathcal{X}_0 \cup \mathcal{X}_1$, where $\mathcal{X}_i$ is a copy of $\mathcal{X}$ in such a way that

(i) the chain $\mathbf{\Phi}$ is the marginal chain $\check{\mathbf{\Phi}}$, in the sense that $P(\Phi_k \in A) = P(\check{\Phi}_k \in A_0 \cup A_1)$ for appropriate initial distributions, and

(ii) the "bottom level" $\mathcal{X}_1$ is an accessible atom for $\check{\mathbf{\Phi}}$.

Before we get into this, though there are two consequences of the existence of atoms that will prove useful later on.

Proposition 7 *Suppose there is an atom $\boldsymbol{\alpha}$ in $\mathcal{B}(\mathcal{X})$ such that $\sum_{n=1}^{\infty} P^n(x, \boldsymbol{\alpha}) > 0$ for all $x \in \mathcal{X}$. Then $\boldsymbol{\alpha}$ is an accessible atom and $\mathbf{\Phi}$ is ν-irreducible with $\nu(A) = P(x, A)$ for all $x \in \boldsymbol{\alpha}$ and $A \in \mathcal{B}(\mathcal{X})$.*

Proof: Let $n \geq 1$. By the Chapman-Kolmogorov equations

$$\begin{aligned} P^{n+1}(x, A) &= \int_{\mathcal{X}} P^n(x, dy) P(y, A) \\ &\geq \int_{\boldsymbol{\alpha}} P^n(x, dy) P(y, A) \\ &= \nu(A) \int_{\boldsymbol{\alpha}} P^n(x, dy) = \nu(A) P^n(x, \boldsymbol{\alpha}). \end{aligned}$$

Sum both sides over n:

$$U(x, A) \geq \sum_{n=1}^{\infty} P^{n+1}(x, A) \geq \nu(A) \sum_{n=1}^{\infty} P^n(x, \boldsymbol{\alpha}) > 0$$

when $\nu(A) > 0$. Therefore $\mathbf{\Phi}$ is ν-irreducible and so there exists a probability measure ψ that satisfies the conditions in Proposition 6. Now suppose $\boldsymbol{\alpha}$ is not accessible. Then $\psi(\boldsymbol{\alpha}) = 0$. Then, by Proposition 6(*iii*), $\psi\{y : L(y, \boldsymbol{\alpha}) > 0\} = 0$. Therefore, for almost all $x \in \mathcal{X}$, $L(x, \boldsymbol{\alpha}) = 0$. This implies that $P^n(x, \boldsymbol{\alpha}) = 0$ for all $n \geq 1$. $\Rightarrow\Leftarrow$ □

Proposition 8 *If $L(x, A) > 0$ for some $x \in \boldsymbol{\alpha}$ where $\boldsymbol{\alpha}$ is an atom, then $\boldsymbol{\alpha} \rightsquigarrow A$.*

Proof: Since $\boldsymbol{\alpha}$ is an atom we have that $P(y, B) = \nu(B)$ for some measure ν on $\mathcal{B}(\mathcal{X})$ for all $y \in \boldsymbol{\alpha}$ and all $B \in \mathcal{B}(\mathcal{X})$. Furthermore, since $L(x, A) > 0$, there exists a $\delta > 0$ such that $L(x, A) \geq \delta > 0$. We need to show that there exists a δ^* such that

$$\inf_{y \in \boldsymbol{\alpha}} L(y, A) \geq \delta^* > 0.$$

I claim that $\delta^* = \delta$ works. To prove this claim, we need to show that for any $y \in \boldsymbol{\alpha}$ with $y \neq x$, $L(y, A) = L(x, A)$. Suppose $y \in \boldsymbol{\alpha}$. Then, by Proposition 2,

$$\begin{aligned} L(y, A) &= \sum_{n=1}^{\infty} P_y(\tau_A = n) \\ &= P(y, A) + \sum_{n=2}^{\infty} \int_{A^c} P(y, dz) P_z(\tau_A = n - 1) \\ &= \nu(A) + \sum_{n=2}^{\infty} \int_{A^c} \nu(dz) P_z(\tau_A = n - 1) \\ &= \nu(A) + \int_{A^c} \nu(dz) L(z, A), \end{aligned}$$

which is independent of y and the inequality above holds with $\delta^* = \delta$. □.

In order to construct the split chain, we have to consider sets that satisfy the following

Definition 17 (Minorization Condition) *For some $\delta > 0$, some $C \in \mathcal{B}(\mathcal{X})$ and some probability measure ν such that $\nu(C) = 1$ and $\nu(C^c) = 0$*

$$P(x, A) \geq \delta \nu(A) \mathbb{I}_C(x), \quad \forall\, x \in \mathcal{X}, \forall A \in \mathcal{B}(\mathcal{X}).$$

This definition ensures that the chain has probabilities uniformly bounded below by multiples of ν for every $x \in C$.

Now we will construct the split chain. We first split the space $\mathcal{X}$ by writing $\check{\mathcal{X}} = \mathcal{X} \times \{0, 1\}$, where $\mathcal{X}_i := \mathcal{X} \times \{i\}$, $i = 0, 1$. These are thought of as copies of $\mathcal{X}$ equipped with copies $\mathcal{B}(\mathcal{X}_i)$ of the σ-algebra $\mathcal{B}(\mathcal{X})$.

Let $\mathcal{B}(\check{\mathcal{X}})$ denote the σ-algebra of subsets of $\check{\mathcal{X}}$ generated by $\mathcal{B}(\mathcal{X}_0)$ and $\mathcal{B}(\mathcal{X}_1)$: that is $\mathcal{B}(\check{\mathcal{X}})$ is the smallest σ-algebra containing sets of the form $A_0 := A\times\{0\}$, $A_1 := A\times\{1\}$, $A \in \mathcal{B}(\mathcal{X})$. We denote the elements of $\check{\mathcal{X}}$ by x_0 and x_1 where x_0 denotes members of the upper level $\mathcal{X}_0$ and x_1 denotes members of the lower level $\mathcal{X}_1$.

Suppose λ is any measure on $\mathcal{B}(\mathcal{X})$. We split λ into two measures, one on $\mathcal{X}_0$ and one on $\mathcal{X}_1$ by defining the measure λ^* on $\mathcal{B}(\check{\mathcal{X}})$:

$$\left.\begin{array}{rcl} \lambda^*(A_0) &=& \lambda(A\cap C)(1-\delta) + \lambda(A\cap C^c) \\ \lambda^*(A_1) &=& \lambda(A\cap C)\delta \end{array}\right\} \tag{2}$$

where δ and C are the constant and the set in the minorization condition. Note that λ is the marginal measure induced by λ^*: for any $A \in \mathcal{B}(\mathcal{X})$ we have

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

When $A \subset C^c$, $\lambda^*(A_0) = \lambda(A)$—only subsets of C are effectively split by this construction.

Now we need to split the chain $\mathbf{\Phi}$ to form a chain $\check{\mathbf{\Phi}}$ which lives on $(\check{\mathcal{X}}, \mathcal{B}(\check{\mathcal{X}}))$. For $x_i \in \check{\mathcal{X}}$ and $\check{A} \in \mathcal{B}(\check{\mathcal{X}})$, define the split kernel $\check{P}(x_i, \check{A})$ by

$$\check{P}(x_0, \check{A}) = P^*(x, \check{A}), \qquad x_0 \in \mathcal{X}_0 \setminus C_0; \tag{3}$$

$$\check{P}(x_0, \check{A}) = (1-\delta)^{-1}(P^*(x, \check{A}) - \delta\nu^*(\check{A})), \qquad x_0 \in C_0; \tag{4}$$

$$\check{P}(x_1, \check{A}) = \nu^*(\check{A}), \qquad x_1 \in \mathcal{X}_1, \tag{5}$$

where ν, C and δ are the measure, set and constant in the minorization condition.

Outside C_0, $\check{\mathbf{\Phi}}$ behaves as $\mathbf{\Phi}$ (Equation 3) moving on the top half $\mathcal{X}_0$ of the split space. When $\mathbf{\Phi}$ enters C, we split $\check{\mathbf{\Phi}}$. $\check{\mathbf{\Phi}}$ moves to C_1 with probability δ and stays in C_0 with probability $1-\delta$. If $\check{\mathbf{\Phi}}$ stays in C_0, we have the modified law (4). The bottom half $\mathcal{X}_1$ is an atom (5), by construction. By Equation 2, $\check{P}(x_i, \mathcal{X}_1 \setminus C_1) = 0$ for all $x_i \in \check{\mathcal{X}}$. Hence, by the Chapman-Kolmogorov equations $\check{P}^n(x_i, \mathcal{X}_1 \setminus C_1) = 0$ for all $x_i \in \check{\mathcal{X}}$ and all $n > 0$. Therefore, the atom $C_1 \subset \mathcal{X}_1$ is the only part of the bottom level that is reached with positive probability. Since C_1 is an atom, we will use the notation $\check{\boldsymbol{\alpha}} := C_1$, to emphasize this fact.

Note that the minorization condition is only used in (4). However, without it, $\check{P}$ would not be a probability law on $\mathcal{B}(\check{\mathcal{X}})$.

Theorem 2 *The following hold for the split and original chain:*

(i) The chain $\mathbf{\Phi}$ is the marginal chain of $\check{\mathbf{\Phi}}$: i.e., for any initial distribution λ on $\mathcal{B}(\mathcal{X})$ and any $A \in \mathcal{B}(\mathcal{X})$,

$$\int_{\mathcal{X}} P^n(x, A)\lambda(dx) = \int_{\check{\mathcal{X}}} \check{P}^n(x_i, A_0 \cup A_1)\lambda^*(dx_i).$$

(ii) $\check{\mathbf{\Phi}}$ *is* φ^**-irreducible if and only if* $\mathbf{\Phi}$ *is* φ*-irreducible.*

(iii) If $\mathbf{\Phi}$ *is* φ*-irreducible and* $\varphi(C) > 0$*, then* $\check{\mathbf{\Phi}}$ *is* ν^**-irreducible and* $\check{\boldsymbol{\alpha}}$ *is an accessible atom for* $\check{\mathbf{\Phi}}$*.*

Proof:
(i) Suppose $n = 1$.

$$\begin{aligned}
\int_{\check{\mathcal{X}}} \check{P}(x_i, A_0 \cup A_1)\lambda^*(dx_i) &= \int_{\check{C}^c} \check{P}(x_i, A_0 \cup A_1)\lambda^*(dx_i) + \int_{\check{C}} \check{P}(x_i, A_0 \cup A_1)\lambda^*(dx_i) \\
&= \int_{C^c} P^*(x, A_0 \cup A_1)\lambda(dx) + \int_{\check{C}} \left[\check{P}(x_0, A_0 \cup A_1)\lambda^*(dx_0) + \check{P}(x_1, A_0 \cup A_1)\lambda^*(dx_1)\right] \\
&= \int_{C^c} P^*(x, A_0 \cup A_1)\lambda(dx) + \int_{C} \left[P^*(x, A_0 \cup A_1) - \delta\nu^*(A_0 \cup A_1) + \delta\nu^*(A_0 \cup A_1)\right]\lambda(dx) \\
&= \int_{\mathcal{X}} P^*(x, A_0 \cup A_1)\lambda(dx) = \int_{\mathcal{X}} P(x, A)\lambda(dx).
\end{aligned}$$

The result follows by induction and the Markov property.

(ii) By defininiton of φ^*, $\varphi^*(A_0 \cup A_1) > 0$ iff $\varphi(A) > 0$. By (i),

$$\int_{\mathcal{X}} U(x, A)\lambda(dx) = \int_{\check{\mathcal{X}}} \check{U}(x_i, A_0 \cup A_1)\lambda^*(dx_i),$$

for any initial distribution λ. In particular, for $\lambda = \delta_y$. Whence,

$$U(y, A) = \int_{\mathcal{X}} U(x, A)\delta_y(dx) = \int_{\check{\mathcal{X}}} \check{U}(x_i, A_0 \cup A_1)\delta^*_{y_i}(dx_i) = \check{U}(y_i, A_0 \cup A_1).$$

Since y is arbitrary, it is true for all y and $U(y, A) > 0$ iff $\check{U}(y_i, A_0 \cup A_1) > 0$.

(iii) Since $\mathbf{\Phi}$ is φ-irreducible, $\check{\mathbf{\Phi}}$ is ν^*-irreducible. Also, $\nu^*(\check{\boldsymbol{\alpha}}) \equiv \nu^*(C_1) = \delta\nu(C) > 0$. Therefore $\check{\boldsymbol{\alpha}}$ is an accessible atom. □

1.5.2 Small sets

A small set is a set for which the minorization condition holds, at least for an m-skeleton chain. For the splitting construction of the previous section, then, the existence of small sets is of considerable importance as they ensure the splitting method is not vacuous. Small sets act in many ways like atoms, and thus are sometimes referred to as "pseudo-atoms". The central results regarding small sets is that for a ψ-irreducible chain every set $A \in \mathcal{B}^+(\mathcal{X})$ contains a small set: C is small with $C \subseteq A$ and $C \in \mathcal{B}^+(\mathcal{X})$. Therefore, every ψ-irreducible chain has an m-skeleton that can be split, and for which the atomic structure of the split chain can be exploited.

Definition 18 (Small sets) *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*

$$P^m(x, B) \geq \nu_m(B), \quad \forall\, x \in C, \forall\, B \in \mathcal{B}(\mathcal{X}).$$

When this holds, we say that C is ν_m-small.

Recall that the probability transition kernel $P^n(x, \cdot)$ is a probability measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Therefore there exists a Lebesgue decomposition into its absolutely continuous and singular parts with respect to some σ-finite measure ϕ on $\mathcal{B}(\mathcal{X})$: for any fixed x and $B \in \mathcal{B}(\mathcal{X})$

$$P^n(x, B) = \int_B p^n(x, y)\phi(dy) + P_\perp(x, B), \tag{6}$$

where $p^n(x, y)$ is the density of $P^n(x, \cdot)$ with respect to ϕ and $P_\perp$ and ϕ are mutually singular. (Recall that two measures ν and μ are mutually singular, $\mu \perp \nu$, if there exists disjoint sets A and B such that $A \cup B = \mathcal{X}$ and $\nu(A) = \mu(B) = 0$. Thus, the two measures "live" on different subspaces of $\mathcal{X}$.)

Theorem 3 *Suppose ϕ is a σ-finite measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Suppose $A \in \mathcal{B}(\mathcal{X})$ with $\phi(A) > 0$ such that*

$$B \subseteq A,\, \phi(B) > 0 \implies \sum_{k=1}^{\infty} P^k(x, B) > 0, \quad x \in A.$$

Then, for every n, the density p^n can be chosen to be a measurable function on $\mathcal{X}^2$ and there exists $C \subseteq A$, an $m > 1$, and a $\delta > 0$ such that $\phi(C) > 0$ and

$$p^m(x, y) > \delta, \quad x, y \in C. \tag{7}$$

Proof: (Meyn & Tweedie, pp. 103–105).

The key point in this theorem is that we can define a version of the densities of the probability transition kernel such that (7) holds uniformly over $x \in C$.

Now the main theorem of this section:

Theorem 4 *If $\mathbf{\Phi}$ is ψ-irreducible, then for every $A \in \mathcal{B}^+(\mathcal{X})$ there exists $m \geq 1$ and a ν_m-small set $C \subseteq A$ such that $C \in \mathcal{B}^+(\mathcal{X})$ and $\nu_m(C) > 0$.*

Proof: Let $\phi = \psi$. Let $\mathbf{\Phi}$ be ψ-irreducible and $A \in \mathcal{B}^+(\mathcal{X}) \subset \mathcal{B}(\mathcal{X})$. Then $\psi(A) > 0$. (That ψ is a σ-finite measure will be shown shortly). Also for all $B \subseteq A$ with $\psi(B) > 0 \implies U(x, B) > 0$ and so it certainly holds for all $x \in A$. Therefore, all conditions of Theorem 3 hold. Therefore, there exists a $C \subseteq A$, an $m > 1$ and a $\delta > 0$ such that $\psi(C) > 0$ and $p^m(x, y) > \delta$, $x, y \in C$.

I claim that C is ν_m-small, with $\nu_m(\cdot) = \delta\psi(\cdot \cap C)$.

First note that $C \in \mathcal{B}^+(\mathcal{X}) \subseteq \mathcal{B}(\mathcal{X})$. Next, $\nu_m(C) = \delta\psi(C) > 0$, so that ν_m is non-trivial. Third, for any $B \in \mathcal{B}(\mathcal{X})$ and all $x \in C$,

$$P^m(x, B) = \int_B p^m(x, y)\psi(dy) + P_\perp(x, B) \geq \int_{B\cap C} p^m(x, y)\psi(dy) \geq \delta\psi(B \cap C) = \nu_m(B),$$

where m and δ are from the results of Theorem 3. Therefore, C is ν_m-small. □

Theorem 5 *If $\mathbf{\Phi}$ is ψ-irreducible, then the minorization condition holds for some m-skeleton chain and for every K_ϵ-chain, $0 < \epsilon < 1$.*

Proof: Take the m and δ and C from Theorem 4. Then the m-skeleton chain $\mathbf{\Phi}^m$ has transition probability law $P^m(x, \cdot)$. We need to show that the minorization condition holds. That is, we need to find a probability measure ν such that $\nu(C) = 1$ and $P^m(x, A) \geq \delta^*\nu(A)\mathbb{I}_C(x)$ for all $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$. I claim $\nu(\cdot) = \psi(\cdot \cap C)/\psi(C)$ is the required probability measure with $\delta^* = \delta\psi(C)$, which is easy to check.

Now, let $0 < \epsilon < 1$. The K_ϵ-chain has transition probability kernel

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

The minorization condition holds with $\nu(\cdot) = \psi(\cdot \cap C)/\psi(C)$ and $\delta^* = (1 - \epsilon)\epsilon^m\delta\psi(C)$. □

Any $\mathbf{\Phi}$ that is ψ-irreducible is well endowed with small sets from Theorems 3 and 4. Also a small set can be rather large, in fact, it can be all of $\mathcal{X}$. Given the existence of just one small set, we now show that it is possible to cover all of $\mathcal{X}$ with small sets in the ψ-irreducible case.

Proposition 9

(i) If $C \in \mathcal{B}(\mathcal{X})$ is ν_n-small and if for any $x \in D$, $P^m(x, C) \geq \delta > 0$, then D is ν_{m+n}-small where ν_{m+n} is a multiple of ν_n.

(ii) If $\mathbf{\Phi}$ *is* ψ*-irreducible, then there exists a countable collection* C_i *of small sets in* $\mathcal{B}(\mathcal{X})$ *such that*

$$\mathcal{X} = \bigcup_{i=1}^{\infty} C_i.$$

(iii) If $\mathbf{\Phi}$ *is* ψ*-irreducible and* $C \in \mathcal{B}^+(\mathcal{X})$ *is* ν_n*-small, then there exists an* $M \geq 1$ *and a measure* ν_M *such that* C *is* ν_M*-small and* $\nu_M(C) > 0$.

Proof:
(i) Let $x \in D$ and $B \in \mathcal{B}(\mathcal{X})$. Then,

$$\begin{aligned} P^{n+m}(x, B) &= \int_X P^m(x, dy) P^n(y, B) \\ &\geq \int_C P^m(x, dy) P^n(y, B) \\ &\geq \nu_n(B) \int_C P^m(x, dy) \\ &\geq \delta \nu_n(B). \end{aligned}$$

(ii) Since $\mathbf{\Phi}$ is ψ-irreducible, there exists an ν_m-small set $C \in \mathcal{B}^+(\mathcal{X})$ (Theorem 4). By definition of ψ-irreducibility

$$\bar{C}(n, m) = \{y : P^n(y, C) \geq m^{-1}\}$$

is a countable covering of $\mathcal{X}$. From (i), each $\bar{C}(n, m)$ is small.

(iii) Since $\mathbf{\Phi}$ is ψ-irreducible and $C \in \mathcal{B}^+(\mathcal{X})$, it follows that $K_{1/2}(x, C) > 0$, $\forall x \in \mathcal{X}$. Therefore, there exists an $m \geq 1$ such that

$$\nu_n K_{1/2}(C) := \int_{\mathcal{X}} \nu_n(dx) K_{1/2}(x, C) > 0 \implies \nu_n P^m(C) := \int_{\mathcal{X}} \nu_n(dx) P^m(x, C) > 0,$$

since ν_n is non-trivial. Set

$$\nu_M(C) := \nu_n P^m(C).$$

Now, for all $x \in C$ and $B \in \mathcal{B}(\mathcal{X})$

$$\begin{aligned} P^{n+m}(x, B) &= \int_{\mathcal{X}} P^n(x, dy) P^m(y, B) \\ &\geq \int_{\mathcal{X}} \nu_n(dy) P^m(y, B) \\ &= \nu_n P^m(B) = \nu_M(B). \end{aligned}$$

Let $M = n + m$. Then C is ν_M-small, where $M = m + n$. □

The covering in Proposition 9 will be used in the following section.

1.5.3 Cyclic Behavior

An artificial example of cyclic behavior on a continuous state space is the following. Let $\mathcal{X} = [0, d)$ and U_i denote the uniform distribution on $[i, i+1)$. Let $\mathbf{\Phi}$ be a chain with transition probability kernel

$$P(x, \cdot) := \mathbb{I}_{[i-1,i)}(x) U_i(\cdot), \quad i = 0, 1, \ldots, d-1 \pmod{d}.$$

Then it is easy to see that $\mathbf{\Phi}$ cycles through the subsets $[i, i+1)$: if $\Phi_n = x \in [i-1, i)$, then $P(\Phi_{n+1} \in [i, i+1) \mid \Phi_n = x) \equiv 1$.

We will now show that this finite cyclic behavior is the worst behavior to be found for a ψ-irreducible chain. This is due to existence of small sets.

Suppose C is any ν_M-small set. Without loss of generality, we can assume $\nu_M(C) > 0$ by Proposition 9(*iii*). This small set C and the corresponding measure ν_M will be used to define a cycle for a general irreducible Markov chain. Since C is ν_M-small, $P^M(x, C) \geq \nu_M(C) > 0$, for $x \in C$, so when the chain starts in C it returns with positive probability at time M. We will define the set E_C to be the set of time points for which C is a small set with minorizing measure proportional to ν_M:

$$E_C = \{n \geq 1 : C \text{ is } \nu_n\text{-small, with } \nu_n = \delta_n \nu_M \text{ for some } \delta_n > 0\}.$$

E_C is closed under addition: Let $B \in \mathcal{B}(\mathcal{X})$, n, $m \in E_C$. Then, for $x \in C$,

$$P^{n+m}(x, B) \geq \int_C P^n(x, dy) P^m(y, B) \geq [\delta_m \delta_n \nu_M(C)] \nu_M(B).$$

Hence, there is a natural period for the set C: the greatest common divisor of E_C. Furthermore, by a number theory result (see, Billingsley, *Probability and Measure*, Theorem A21, 1995), C is ν_{nd}-small for all n larger than some n_0.

The following theorem states that this value is a property of the chain $\mathbf{\Phi}$ and does not depend on the C chosen.

Theorem 6 *Let $\mathbf{\Phi}$ be ψ-irreducible Markov chain on $\mathcal{X}$. Let $C \in \mathcal{B}^+(\mathcal{X})$ be ν_M-small and let d be the greatest common divisor of the set E_C. Then there exist disjoint sets $D_0, \ldots, D_{d-1} \in \mathcal{B}(\mathcal{X})$ (called a "d-cycle") such that*

(i) if $x \in D_i$, then $P(x, D_{i+1}) = 1$, $\quad i = 0, \ldots, d-1 \pmod{d}$;

(ii) $\psi(N) = 0$ where $N = \left(\cup_{i=0}^{d-1} D_i\right)^c$.

The d-cycle $\{D_i\}$ is maximal in the sense that for any other collection $\{d', D'_k, k = 0, \ldots, d' - 1\}$ that satisfy the above conditions, d' divides d. If $d = d'$, then, $D'_i = D_i$ a.e. $[\psi]$, by a reordering if necessary.

Proof: (Meyn & Tweedie, pp. 111-112).

Definition 19 (Periodic and Aperiodic Chains) *Let $\mathbf{\Phi}$ be a φ-irreducible Markov chain.*

- *The largest d for which a d-cycle occurs for $\mathbf{\Phi}$ is called the period of $\mathbf{\Phi}$.*
- *If $d = 1$, then $\mathbf{\Phi}$ is called aperiodic.*
- *If there exists a ν_1-small set A with $\nu_1(A) > 0$, then $\mathbf{\Phi}$ is called strongly aperiodic.*

Proposition 10 *Let $\mathbf{\Phi}$ be ψ-irreducible.*

(i) *If $\mathbf{\Phi}$ is strongly aperiodic, the minorization condition holds.*

(ii) *For all $\epsilon \in (0, 1)$, the K_ϵ-chain is strongly aperiodic.*

(iii) *If $\mathbf{\Phi}$ is aperiodic, then every skeleton is ψ-irreducible and aperiodic, and there exists an m such that the m-skeleton is strongly aperiodic.*

Proof:
(i) Since $\mathbf{\Phi}$ is strongly aperiodic, there exists a ν_1-small set A with $\nu_1(A) > 0$ and $P(x, B) \geq \nu_1(B)$ for all $x \in A$, $B \in \mathcal{B}(\mathcal{X})$. Let $\nu(\cdot) = \nu_1(\cdot \cap A)/\nu_1(A)$ and $\delta = 1/\nu_1(A)$.

(ii). Since $\mathbf{\Phi}$ is ψ-irreducible, the minorization condition holds for K_ϵ-chains by Theorem 5 for some ν_1-small set $C \in \mathcal{B}^+(\mathcal{X})$.

(iii). Let $\mathbf{\Phi}$ be ψ-irreducible and aperiodic. By Theorem 5, there is an m-skeleton that satisfies the minorization condition. Also, there exist a $C \in \mathcal{B}^+(\mathcal{X})$ that is ν_m-small. But with regard to the m-skeleton, C is ν_1-small and thus the m skeleton is strongly aperiodic. Now let $m >= 1$ be given. Obviously if $m = 1$, the 1-skeleton is ψ-irreducible and aperiodic. Suppose it is true for $m - 1$. Then

$$P^m(x, D) = \int_{\mathcal{X}} P(x, dy) P^{m-1}(y, D) = \int_{\mathcal{X}} P(x, dy) = P(x, \mathcal{X}) = 1.$$

where $\psi(D^c) = 0$. So the m-skeleton is aperiodic, $m \geq 1$. Now need to show that for $m > 1$, the m-skeleton is ψ-irreducible. $\square$

This theorem shows it is desirable to work with strongly aperiodic, irreducible chains as then the minorization condition holds and we can take advantage of the split chain. However, this condition is not met in general and so we will restrict ourselves to aperiodic chains and prove results for strongly aperiodic chains and then use special methods to extend them to general chains through the m-skeleton or K_ϵ-chain.

Proposition 11 *If $\mathbf{\Phi}$ is ψ-irreducible with period d and d-cycle $\{D_i,\, i = 0, \ldots, d-1\}$, then each of the sets D_i is an absorbing ψ-irreducible set for the chain $\mathbf{\Phi}_d$ corresponding to the transition probability kernel P^d. Furthermore, $\mathbf{\Phi}_d$ is aperiodic on each D_i.*

Proof: For $\mathbf{\Phi}_d$ with kernel P^d it is obvious that D_i is absorbing: i.e. $P^d(x, D_i) = 1$ for all $x \in D_i$. $\mathbf{\Phi}_d$ is obviously ψ-irreducible on D_i. By definition of d as the largest value for which a cycle exists, it is obvious that $\mathbf{\Phi}_d$ is aperiodic on each D_i by the previous proposition. □

1.5.4 Petite sets

A convenient tool for the analysis of Markov chains is the sampled chain. This idea extends the idea of the m-skeleton and K_ϵ chain. Petite sets are defined in terms of the sampled chain and are a generalization of small sets.

Let $a = \{a(n)\}$ be a distribution (or probability measure) on $\mathbb{N}_+$. Consider the Markov chain $\mathbf{\Phi}_a$ with probability transition kernel

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

We call $\mathbf{\Phi}_a$ the K_a-chain with sampling distribution a. Probabilistically, $\mathbf{\Phi}_a$ has the interpretation of being the chain $\mathbf{\Phi}$ "sampled" at time points drawn successively according to the distribution a. Note that if $a = \delta_m$ is the Dirac measure with $\delta_m(m) = 1$, then the K_{δ_m}-chain is the m-skeleton with transition probability kernel P^m. If a is the geometric distribution with

$$a(n) = (1 - \epsilon)\epsilon^n, \quad n \in \mathbb{N}_+,$$

then we have the resolvent, K_ϵ. This concept of sampled chains allows us to determine conditions under which one set in uniformly accessible from another set.

Definition 20 *We say that a set $B \in \mathcal{B}(\mathcal{X})$ is 'uniformly accessible using a' (the distribution a) from another set $A \in \mathcal{B}(\mathcal{X})$ if there exists $\delta > 0$ such that*

$$\inf_{x \in A} K_a(x, B) \geq \delta. \tag{8}$$

We then write $A \overset{a}{\rightsquigarrow} B$.

Note: Compare this to the definition of uniformly accessible and make sure you understand the difference. The difference being that *uniformly accessible* is in relation to the entire chain $\mathbf{\Phi}$ and its probability transition kernel P and the condition that $\inf_{x \in A} L(x, A) \geq \delta > 0$ while *uniformly accessible using a* uses the sampled chain $\mathbf{\Phi}_a$ and its probability transition kernel K_a.

Lemma 1 *If* $A \overset{a}{\rightsquigarrow} B$ *for some distribution* a*, then* $A \rightsquigarrow B$.

Proof: By definition there exists $\delta > 0$ such that Equation (8) holds. Let η be distributed as a. Then

$$\delta \leq K_a(x, B) = P_x(\Phi_\eta \in B) \leq P_x(\tau_B < \infty) = L(x, B).$$

Which, by definition, makes B uniformly accessible from A. □

Lemma 2

(i) If a *and* b *are distributions on* $\mathbb{N}_+$*, then the sampled chains* K_a *and* K_b *satisfy the generalized Chapman-Kolmogorov equations*

$$K_{a*b}(x, A) = \int_{\mathcal{X}} K_a(x, dy) K_b(y, A)$$

where $*$ *is the convolution operator.*

(ii) If $A \overset{a}{\rightsquigarrow} B$ *and* $B \overset{b}{\rightsquigarrow} C$*, then* $A \overset{a*b}{\rightsquigarrow} C$.

(iii) If a *is a distribution on* $\mathbb{N}_+$*, then the sampled chain with transition probability kernel* K_a *satisfies the relation*

$$U(x, A) \geq \int_{\mathcal{X}} U(x, dy) K_a(y, A).$$

Proof:

(i) By the Chapman-Kolmogorov equations

$$\begin{aligned} K_{a*b}(x,A) &= \sum_{n=0}^{\infty} P^n(x,A)(a*b)(n) \\ &= \sum_{n=0}^{\infty} P^n(x,A) \sum_{m=0}^{n} a(m)b(n-m) \\ &= \sum_{n=0}^{\infty}\sum_{m=0}^{n} \int_{\mathcal{X}} P^m(x,dy)P^{n-m}(y,A)a(m)b(n-m) \\ &= \int_{\mathcal{X}} \sum_{m=0}^{\infty} P^m(x,dy)a(m) \sum_{n=m}^{\infty} P^{n-m}(y,A)b(n-m) \\ &= \int_{\mathcal{X}} K_a(x,dy)K_b(y,A). \end{aligned}$$

(ii) If $A \overset{a}{\rightsquigarrow} B$ then $\inf_{x\in A} K_a(x,B) \geq \delta_a$ for some $\delta_a > 0$. If $B \overset{b}{\rightsquigarrow} C$ then $\inf_{y\in B} K_b(y,C) \geq \delta_b$ for some $\delta_b > 0$. Then, by (i)

$$K_{a*b}(x,C) = \int_{\mathcal{X}} K_a(x,dy)K_b(y,C) \geq \int_B K_a(x,dy)K_b(y,C) \geq \delta_a\delta_b.$$

(iii) For fixed m,n

$$P^{m+n}(x,A)a(n) = \int_{\mathcal{X}} P^m(x,dy)P^n(y,A)a(n).$$

Summing both sider over m gives

$$U(x,A)a(n) \geq \sum_{m>n} P^m(x,A)a(n) = \int_{\mathcal{X}} U(x,dy)P^n(y,A)a(n).$$

Now sum over n,

$$U(x,A) \geq \int_{\mathcal{X}} U(x,dy)K_a(y,A).$$

□

Small sets always exist for ψ-irreducible chains. We will also show that every small set is petite and if the chain is aperiodic, then every petite set is small. For MCMC theory we will only be concerned with aperiodic chains. But for MCMC theory we also need the concepts of recurrence and Harris recurrence, which we will study on their own merit (independent of whether the chain is aperiodic). Some of the results of recurrence will rely on petite sets. Hence,

Definition 21 (Petite sets) *A set $C \in \mathcal{B}(\mathcal{X})$ is ν_a-petite if the sample chain satisfies the bound*

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

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

The following shows that every small set is petite.

Proposition 12 *If $C \in \mathcal{B}(\mathcal{X})$ is ν_m-small, then C is ν_{δ_m}-petite.*

Proof: Since C is ν_m-small, we have for all $x \in \mathcal{C}$, $B \in \mathcal{B}(\mathcal{X})$

$$P^m(x, B) \geq \nu_m(B).$$

Now take $a = \delta_m$ in the definition of the sampled chain. That is, $a(n) = 0$ for $n \neq m$ and $a(m) = 1$. Then

$$K_{\delta_m}(x, B) = \sum_{n=0}^{\infty} P^n(x, B)a(n) = P^m(x, B) \geq \nu_m(B).$$

The result holds by taking $\nu_{\delta_m}(B) = \nu_m(B)$. □

Proposition 13

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

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

Proof:
(*i*) Since $D \overset{d}{\rightsquigarrow} A$ there exists a $\delta > 0$ such that $K_d(x, A) \geq \delta$ for all $x \in D$. Also since A is ν_a-petite we have $K_a(x, B) \geq \nu_a(B)$ for all $x \in A$ and $B \in \mathcal{B}(\mathcal{X})$. Now

$$K_{d*a}(x, B) = \int_{\mathcal{X}} K_d(x, dy)K_a(y, B) \geq \int_A K_d(x, dy)K_a(y, B) \geq \int_A K_d(x, dy)\nu_a(B) \geq \delta\nu_a(B).$$

Thus, D is ($\nu_{d*a} = \delta\nu_a$)-petite as claimed.

(*ii*) Let $\bar{A}(n,m) = \{y : P^n(y,A) > m^{-1}\}$. Since $A \in \mathcal{B}^+(\mathcal{X})$ is ν_a-petite, then for $B \in \mathcal{B}(\mathcal{X})$ we have $K_a(x,B) \geq \nu_a(B)$ for all $x \in A$. Assume $\nu_a(B) > 0$. Let $x \in \mathcal{X}$. By ψ-irreducibility there exist an n such that $P^n(x,A) > 0$ and therefore an m such that $x \in \bar{A}(n,m)$. Now

$$\int_{\mathcal{X}} P^n(x,dy)K_a(y,B) \geq \int_A P^n(x,dy)K_a(y,B) \geq m^{-1}\nu_a(B) > 0.$$

Furthermore,

$$\begin{aligned}
\int_{\mathcal{X}} P^n(x,dy)K_a(y,B) &= \sum_{r=0}^{\infty} \int_{\mathcal{X}} P^n(x,dy)P^r(y,B)a(r) \\
&\leq \sum_{r=0}^{\infty} \int_{\mathcal{X}} P^n(x,dy)P^r(y,B) \\
&= \sum_{r=0}^{\infty} P^{n+r}(x,B) \\
&\leq U(x,B).
\end{aligned}$$

Hence, $\mathbf{\Phi}$ is ν_a-irreducible. □

So, the above proposition tells us if we have a petite set, then we can generate an irreducibility measure for $\mathbf{\Phi}$. There are other useful properties of petite sets that distinguish them from small sets.

Proposition 14 *Suppose* $\mathbf{\Phi}$ *is* ψ*-irreducible.*

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

(ii) The union of two petite sets is petite.

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

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

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* $\cup C_i = \mathcal{X}$*.*

Proof:
(*i*) To prove this we assume without loss of generality that ν_a is an irreducibility measure

even if $\psi(A) = 0$. To see this, by ψ-irreducibility, there exists a ν_b-petite (actually a small set) C such that $C \in \mathcal{B}^+(\mathcal{X})$ and $\nu_b(C) > 0$ (Proposition 9(iii)). We also have $K_\epsilon(y, C) > 0$ for any $y \in \mathcal{X}$ and $\epsilon \in (0, 1)$, since $\mathbf{\Phi}$ is ψ-irreducible and $C \in \mathcal{B}(\mathcal{X})$. So for $x \in A$

$$K_{a*\epsilon}(x, C) = \int_{\mathcal{X}} K_a(x, dy)K_\epsilon(y, C) \geq \int_{\mathcal{X}} \nu_a(dy)K_\epsilon(y, C) > 0.$$

So this shows that $A \overset{a*\epsilon}{\rightsquigarrow} C$. But then by Proposition 13(i), A is $\nu_{a*\epsilon*b}$-petite which is a multiple of ν_b. Now apply Proposition 13(ii) to C. Thus ν_b is an irreducibility measure for $\mathbf{\Phi}$. But then clearly any multiple of ν_b is an irreducibility measure for $\mathbf{\Phi}$. Hence, $\nu_{a*\epsilon*b}$ is an irreducibility measure. Therefore we can always find an irreducibility measure ν_a with A being ν_a-petite.

So we assume ν_a is an irreducibility measure. Now, for $0 < \epsilon < 1$

$$K_{a*\epsilon}(x, B) = \int_{\mathcal{X}} K_a(x, dy)K_\epsilon(y, B) \geq \int_{\mathcal{X}} \nu_a(dy)K_\epsilon(y, B), \quad x \in A,\ B \in \mathcal{B}(\mathcal{X}).$$

Let $b = a * \epsilon.$ and ψ_b equal the r.h.s. above. Then ψ_b is a maximal irreducibility measure by Proposition 6(iv) and A is ψ_b-petite.

(ii) Suppose A_1 is ψ_{a_1}-petite and A_2 is ψ_{a_2}-petite such that both measures are maximal irreducible. Let $A \in \mathcal{B}^+(\mathcal{X})$ be a fixed petite set. Define the sampling measure a on $\mathbb{N}_+$ by $a(i) = \frac{1}{2}[a_1(i) + a_2(i)]$. Let $x \in A_1 \cup A_2$, then

$$K_a(x, A) \geq \frac{1}{2} \min\left(\psi_{a_1}(A), \psi_{a_2}(A)\right) > 0$$

Therefore $A_1 \cup A_2 \overset{a}{\rightsquigarrow} A$. Hence, by Proposition 13(i), $A_1 \cup A_2$ is petite.

(iii) By Theorem 4 we can find a ν_n-small set $C \in \mathcal{B}^+(\mathcal{X})$. By (i) we may assume that C is ψ_b-petite where ψ_b is a maximal irreducibility measure. Hence $K_b(y, \cdot) \geq \mathbb{I}_C(y)\psi_b(\cdot)$ for all $y \in \mathcal{X}$.

By irreducibility we also have $K_\epsilon(x, C) > 0$ for all $\epsilon \in (0, 1)$ and all $x \in \mathcal{X}$. Thus, for any $x \in \mathcal{X}$ and $B \in \mathcal{B}^+(\mathcal{X})$,

$$K_{b*\epsilon}(x, B) = \int_{\mathcal{X}} K_\epsilon(y, dz)K_b(z, B) \geq K_\epsilon(x, C)\psi_b(B).$$

The result holds with $c = b * \epsilon$, $s(x) = K_\epsilon(x, C)$ and $\psi_c = \psi_b$.

Now, for $m \geq 1$, let

$$C_m := \{x \in \mathcal{X} : s(x) \geq m^{-1}\}$$

These sets are an increasing sequence, i.e. $C_m \subseteq C_{m+1}$ so that

$$\mathcal{X} = \lim_{m\uparrow\infty} C_m = \bigcup_{m=1}^{\infty} C_m.$$

□

We have already seen that every small set is petite. We now show that if $\mathbf{\Phi}$ is also aperiodic, then every petite set is small.

Theorem 7 *If $\mathbf{\Phi}$ is irreducible and aperiodic, then every petite set is small.*

Proof: Let A be petite. Then we can assume A is ψ_a-petite where ψ_a is a maximal irreducibility measure. Let C be the small set in the definition

$$E_C = \{n \geq 1 : C \text{ is } \nu_n\text{-small, with } \nu_n = \delta_n \nu_M \text{ for some } \delta_n > 0\}.$$

Since the chain is aperiodic, by the number theory result given before Theorem 6, there exists some $n_0 \in \mathbb{N}_+$ such that C is ν_k-small with $\nu_k = \delta\nu_M$ for some $\delta > 0$ (take $\delta = \min_k(\delta_k)$ for $k \in E_C$ with $n_0/2 - 1 \leq k \leq n_0$ for $C \in \mathcal{B}^+(\mathcal{X})$). Since $C \in \mathcal{B}^+(\mathcal{X})$ we take n_0 so large that

$$\sum_{k=\lceil n_0/2\rceil}^{\infty} a(k) \leq \frac{1}{2}\psi_a(C).$$

This holds because if $\sum_{k=0}^{\infty} a(k)$ is a convergent series, then $\lim_{n\to\infty} \sum_{k=n}^{\infty} a(k) = 0$ and the fact that ψ_a is a probability measure so that $0 < \psi_a(C) \leq 1$. Since each $0 \leq P^k(x, C) \leq 1$, we have

$$\sum_{k=\lceil n_0/2\rceil}^{\infty} P^k(x, C)a(k) \leq \sum_{k=\lceil n_0/2\rceil}^{\infty} a(k).$$

A is a ψ_a-petite set, therefore for all $x \in A$,

$$\begin{aligned}\psi_a(C) \leq K_a(x, C) = \sum_{k=0}^{\infty} P^k(x, C)a(k) &= \sum_{k=0}^{\lfloor n_0/2\rfloor} P^k(x, C)a(k) + \sum_{k=\lceil n_0/2\rceil}^{\infty} P^k(x, C)a(k) \\ &\leq \sum_{k=0}^{\lfloor n_0/2\rfloor} P^k(x, C)a(k) + \frac{1}{2}\psi_a(C)\end{aligned}$$

which implies

$$\frac{1}{2}\psi_a(C) \leq \sum_{k=0}^{\lfloor n_0/2\rfloor} P^k(x, C)a(k).$$

We then have for all $x \in A$ and $B \in \mathcal{B}(\mathcal{X})$

$$\begin{aligned} P^{n_0}(x,B) = \sum_{k=0}^{\infty} P^{n_0}(x,B)a(k) &= \sum_{k=0}^{\infty} \left\{ \int_{\mathcal{X}} P^k(x,dy) P^{n_0-k}(y,B) \right\} a(k) \\ &\geq \sum_{k=0}^{\lfloor n_0/2 \rfloor} \left\{ \int_{C} P^k(x,dy) P^{n_0-k}(y,B) \right\} a(k) \\ &\geq \left(\sum_{k=0}^{\lfloor n_0/2 \rfloor} P^k(x,C)a(k) \right) (\delta \nu_M(B)) \\ &\geq \left(\frac{1}{2}\psi_a(C) \right) (\delta \nu_M(B)) \,. \end{aligned}$$

Now the second inequality holds since for $0 \leq k \leq \lfloor n_0/2 \rfloor$, we have $n_0/2 - 1 \leq n_0 - k \leq n_0$ so that $P^{n_0-k}(y,B) \geq \delta \nu_M(B)$ for all $x \in C$ by our choice of n_0. Therefore A is ν_{n_0}-small with $\nu_{n_0} = (\frac{1}{2}\delta\psi_a(C))\nu_M$. □

1.6 Transience and Recurrence/Harris Recurrence

Definition 22 (Uniform Transience and Recurrence) *A set A is called uniformly transient if there exists a real valued number M such that for all $x \in \mathcal{A}$, $\mathbb{E}_x(\eta_A) \leq M$.*

A is called recurrent if for all $x \in A$, $\mathbb{E}_x(\eta_A) = \infty$.

Note that in this definition, it is not $U(x,A) := \mathbb{E}_x(\eta_A) = \sum_{n=1}^{\infty} P^n(x,A)$, $x \in \mathcal{X}$. We need another definition here to aid in the discussion.

Definition 23 (Taboo Probabilities) *The n-step taboo probability is*

$$ {}_AP^n(x,B) := P_x(\Phi_n \in B, \tau_A \geq n), \quad x \in \mathcal{X};\ A, B \in \mathcal{B}(\mathcal{X}).$$

${}_AP^n(x,B)$ denotes the probability of a transition to B in n steps of the chain, "avoiding" the set A. The taboo probabilities satisfy the iterative relation

$$ {}_AP^1(x,B) = P(x,B)$$

and for $n \geq 1$

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

Define

$$U_A(x,B) := \sum_{n=1}^{\infty} {}_AP^n(x,B), \quad x \in \mathcal{X};\ A,B \in \mathcal{B}(\mathcal{X}).$$

Note also that

$$L(x,A) = U_A(x,A), \quad x \in \mathcal{X};\ A \in \mathcal{B}(\mathcal{X}).$$

By convention ${}_AP^0(x,A) = 0$.

Now, for $B \in \mathcal{B}(\mathcal{X})$ consider the event

$$\{\Phi_n \in B\} = \cup_{j=1}^{n-1}\{\Phi_n \in B, \tau_A = j\} \cup \{\Phi_n \in B, \tau_A \geq n\}, \quad A \in \mathcal{B}(\mathcal{X}),\ A \neq B.$$

The sets on the r.h.s. are mutually exclusive. Based on this decomposition we have the *first-entrance decomposition*

$$P^n(x,B) = {}_AP^n(x,B) + \sum_{j=1}^{n-1} \int_A {}_AP^j(x,dy)P^{n-j}(y,B).$$

Similarly, there exists a decomposition of $\{\Phi_n \in B\}$ into mutually exclusive sets that results in the *last-exit decomposition*

$$P^n(x,B) = {}_AP^n(x,B) + \sum_{j=1}^{n-1} \int_A P^j(x,dy){}_AP^{n-j}(y,B).$$

In what follows, we will need the generating functions for the series $\{P^n\}$ and $\{{}_AP^n\}$. Let $|z| < 1$. Then the generating function for $\{P^n\}$ is

$$U^{(z)}(x,B) := \sum_{n=1}^{\infty} P^n(x,B)z^n$$

and the generating function for $\{{}_AP^n\}$ is

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

By their definitions, it is easy to see that

$$U(x,B) = \sum_{n=1}^{\infty} P^n(x,B) = \lim_{z\uparrow 1} U^{(z)}(x,B).$$

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

$$L(x,A) = \sum_{n=1}^{\infty} {}_AP^n(x,A) = \lim_{z\uparrow 1} U_A^{(z)}(x,A).$$

Now, multiply the first-entrance and last-exit decompositions by z^n and sum over n.

$$\begin{aligned} U^{(z)}(x,B) &= \sum_{n=1}^{\infty} P^n(x,B)z^n \\ &= U_A^{(z)}(x,B) + \sum_{n=1}^{\infty}\sum_{j=1}^{n-1}\int_A {}_AP^j(x,dy)P^{n-j}(y,B)z^n \\ &= U_A^{(z)}(x,B) + \int_A \sum_{n=1}^{\infty} {}_AP^n(x,dy)z^n \sum_{m=1}^{\infty} P^m(y,B)z^m \\ &= U_A^{(z)}(x,B) + \int_A U_A^{(z)}(x,dy)U^{(z)}(y,B) \end{aligned} \tag{9}$$

and

$$U^{(z)}(x,B) = U_A^{(z)}(x,B) + \int_A U^{(z)}(x,dy)U_A^{(z)}(y,B). \tag{10}$$

We now consider the case when the chain $\mathbf{\Phi}$ has an atom. We will show that the chain is either recurrent or transient, in the case when $\mathbf{\Phi}$ is ψ-irreducible. We then use the splitting techniques to classify general chains.

Lemma 3 *If $\mathbf{\Phi}$ is ψ-irreducible and has an atom $\boldsymbol{\alpha} \in \mathcal{B}^+(\mathcal{X})$, then $U(\boldsymbol{\alpha},\boldsymbol{\alpha}) = \infty$ if and only if $L(\boldsymbol{\alpha},\boldsymbol{\alpha}) = 1$.*

Proof: Use (10) with $A = B = x = \boldsymbol{\alpha}$:

$$U^{(z)}(\boldsymbol{\alpha},\boldsymbol{\alpha}) = U_{\boldsymbol{\alpha}}^{(z)}(\boldsymbol{\alpha},\boldsymbol{\alpha}) + U_{\boldsymbol{\alpha}}^{(z)}(\boldsymbol{\alpha},\boldsymbol{\alpha})U^{(z)}(\boldsymbol{\alpha},\boldsymbol{\alpha}).$$

Why does this hold? First note that since $\boldsymbol{\alpha}$ is an atom, we have $P(x,A) = \nu(A)$ for some measure ν, all $A \in \mathcal{B}(\mathcal{X})$ and $x \in \boldsymbol{\alpha}$. Let μ be any initial distribution of $\mathbf{\Phi}$. Then

$$\begin{aligned} P(\boldsymbol{\alpha},A) &= \Pr(\Phi_1 \in A \mid \Phi_0 \in \boldsymbol{\alpha}) \\ &= \frac{\int_{\boldsymbol{\alpha}} P(x,A)\mu(dx)}{\int_{\boldsymbol{\alpha}} \mu(dx)} \\ &= \frac{\int_{\boldsymbol{\alpha}} \nu(A)\mu(dx)}{\int_{\boldsymbol{\alpha}} \mu(dx)} \\ &= \nu(A). \end{aligned}$$

Hence, for all $x \in \boldsymbol{\alpha}$, $P(x, A) = P(\boldsymbol{\alpha}, A)$. Hence, for $x \in \boldsymbol{\alpha}$ and $A \in \mathcal{B}(\mathcal{X})$

$$\begin{aligned} U_{\boldsymbol{\alpha}}^{(z)}(x, A) &= \sum_{n=1}^{\infty} {}_{\boldsymbol{\alpha}}P^n(x, A)z^n \\ &= \sum_{n=1}^{\infty} \int_{\boldsymbol{\alpha}^c} P(x, dy)_{\boldsymbol{\alpha}}P^{n-1}(y, A)z^n \\ &= \sum_{n=1}^{\infty} \int_{\boldsymbol{\alpha}^c} P(\boldsymbol{\alpha}, dy)_{\boldsymbol{\alpha}}P^{n-1}(y, A)z^n \\ &= \sum_{n=1}^{\infty} {}_{\boldsymbol{\alpha}}P^n(\boldsymbol{\alpha}, A)z^n \\ &= U_{\boldsymbol{\alpha}}^{(z)}(\boldsymbol{\alpha}, A). \end{aligned}$$

Plug $A = B = x = \boldsymbol{\alpha}$ into (10) and use the above equality

$$\begin{aligned} U^{(z)}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) &= U_{\boldsymbol{\alpha}}^{(z)}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) + \int_{\boldsymbol{\alpha}} U^{(z)}(\boldsymbol{\alpha}, dy)U_{\boldsymbol{\alpha}}^{(z)}(y, \boldsymbol{\alpha}) \\ &= U_{\boldsymbol{\alpha}}^{(z)}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) + \int_{\boldsymbol{\alpha}} U^{(z)}(\boldsymbol{\alpha}, dy)U_{\boldsymbol{\alpha}}^{(z)}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \\ &= U_{\boldsymbol{\alpha}}^{(z)}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) + U^{(z)}(\boldsymbol{\alpha}, \boldsymbol{\alpha})U_{\boldsymbol{\alpha}}^{(z)}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \end{aligned}$$

Now, take the limit of both sides as $z \uparrow 1$.

$$U(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = L(\boldsymbol{\alpha}, \boldsymbol{\alpha}) + L(\boldsymbol{\alpha}, \boldsymbol{\alpha})U(\boldsymbol{\alpha}, \boldsymbol{\alpha}).$$

Solving for $U(\boldsymbol{\alpha}, \boldsymbol{\alpha})$,

$$U(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = \frac{L(\boldsymbol{\alpha}, \boldsymbol{\alpha})}{1 - L(\boldsymbol{\alpha}, \boldsymbol{\alpha})}.$$

Therefore,

$$U(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = \infty \iff L(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = 1.$$

□

Theorem 8 *If $\boldsymbol{\Phi}$ is ψ-irreducible and has an atom $\boldsymbol{\alpha} \in \mathcal{B}^+(\mathcal{X})$ and*

(i) if $\boldsymbol{\alpha}$ is recurrent, then every set in $\mathcal{B}^+(\mathcal{X})$ is recurrent;

(ii) if $\boldsymbol{\alpha}$ is transient, then there is a countable covering of $\mathcal{X}$ by uniformly transient sets.

Proof:
(i) Since $\mathbf{\Phi}$ is ψ-irreducible we have for $\psi(\boldsymbol{\alpha}) > 0$ and $\psi(A) > 0$, that there is an r such that $P^r(x, \boldsymbol{\alpha}) > 0$ for all $x \in \mathcal{X}$ and an s such that $P^s(\boldsymbol{\alpha}, A) > 0$. Hence,

$$\sum_{n=1}^{\infty} P^{r+s+n}(x, A) \geq P^r(x, \boldsymbol{\alpha}) \left[\sum_{n=1}^{\infty} P^n(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \right] P^s(\boldsymbol{\alpha}, A) = \infty.$$

Since $\boldsymbol{\alpha}$ is recurrent, by assumption, the middle term above is infinite, by definition. Since A was arbitrary, the proof is complete.
(ii) Suppose $\boldsymbol{\alpha}$ is transient. Set $A = B = \boldsymbol{\alpha}$. From the last-exit decomposition (10)

$$U^{(z)}(x, \boldsymbol{\alpha}) = U^{(z)}_{\boldsymbol{\alpha}}(x, \boldsymbol{\alpha}) + U^{(z)}(x, \boldsymbol{\alpha}) U^{(z)}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}, \boldsymbol{\alpha}).$$

Solving for $U^{(z)}(x, \boldsymbol{\alpha})$ and taking limits on both sides

$$U(x, \boldsymbol{\alpha}) = \lim_{z \uparrow 1} U^{(z)}(x, \boldsymbol{\alpha}) = \lim_{z \uparrow 1} \frac{U^{(z)}_{\boldsymbol{\alpha}}(x, \boldsymbol{\alpha})}{1 - U^{(z)}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}, \boldsymbol{\alpha})} = \frac{L(x, \boldsymbol{\alpha})}{1 - L(\boldsymbol{\alpha}, \boldsymbol{\alpha})} < \infty$$

holds for all $z < 1$ and the last inequality holds from the previous lemma. Hence, $U(x, \boldsymbol{\alpha}) < \infty$ for all $x \in \mathcal{X}$.

Consider the sets

$$\bar{\boldsymbol{\alpha}}(j) = \left\{ y : \sum_{n=1}^{j} P^n(y, \boldsymbol{\alpha}) > j^{-1} \right\}. \tag{11}$$

These sets are a countable covering of $\mathcal{X}$. Using the Chapman-Kolmogorov equations,

$$U(x, \boldsymbol{\alpha}) \geq j^{-1} U(x, \bar{\boldsymbol{\alpha}}(j)) \inf_{y \in \bar{\boldsymbol{\alpha}}(j)} \sum_{n=1}^{j} P^n(y, \boldsymbol{\alpha}) \geq j^{-2} U(x, \bar{\boldsymbol{\alpha}}(j)). \tag{12}$$

Hence,

$$U(x, \bar{\boldsymbol{\alpha}}(j)) \leq j^2 U(x, \boldsymbol{\alpha}) < \infty$$

and so $\{\bar{\boldsymbol{\alpha}}(j)\}$ is the required cover by uniformly transient sets.

The second inequality in (12) is obvious by (11). To see the first inequality, note that

$$\sum_{m=1}^{k} \sum_{n=1}^{j} P^{m+n}(x, \boldsymbol{\alpha}) = \sum_{m=1}^{k} \sum_{a=m+1}^{m+j} P^a(x, \boldsymbol{\alpha}) \leq j \sum_{a=2}^{k+j} P^a(x, \boldsymbol{\alpha})$$

since the largest coefficient of any $P^a(x, \boldsymbol{\alpha})$ is $\min(k, j)$.

Also, note that

$$\begin{aligned}\sum_{m=1}^{k}\sum_{n=1}^{j} P^{m+n}(x,\boldsymbol{\alpha}) &= \sum_{m=1}^{k}\sum_{n=1}^{j}\int_{\mathcal{X}} P^m(x,dy)P^n(y,\boldsymbol{\alpha}) \\ &\geq \sum_{m=1}^{k}\int_{\bar{\boldsymbol{\alpha}}(j)} P^m(x,dy)\sum_{n=1}^{j} P^n(y,\boldsymbol{\alpha}) \\ &\geq \sum_{m=1}^{k} P^m(x,\bar{\boldsymbol{\alpha}}(j))\inf_{y\in\bar{\boldsymbol{\alpha}}(j)}\sum_{n=1}^{j} P^n(y,\boldsymbol{\alpha}).\end{aligned}$$

Therefore, by letting $k \to \infty$

$$\begin{aligned}U(x,\boldsymbol{\alpha}) &= \sum_{a=1}^{\infty} P^a(x,\boldsymbol{\alpha}) \\ &\geq \lim_{k\to\infty}\sum_{a=2}^{k} P^a(x,\boldsymbol{\alpha}) \\ &\geq \lim_{k\to\infty} j^{-1}\sum_{m=1}^{k} P^m(x,\bar{\boldsymbol{\alpha}}(j))\inf_{y\in\bar{\boldsymbol{\alpha}}(j)}\sum_{n=1}^{j} P^n(y,\boldsymbol{\alpha}) \\ &= j^{-1}U(x,\bar{\boldsymbol{\alpha}}(j))\inf_{y\in\bar{\boldsymbol{\alpha}}(j)}\sum_{n=1}^{j} P^n(y,\boldsymbol{\alpha}).\end{aligned}$$

□

Definition 24 *If $A \in \mathcal{B}(\mathcal{X})$ can be covered by a countable number of uniformly transients sets, then we call A transient.*

Now we will consider chains $\boldsymbol{\Phi}$ that do not have an atom and use the split chain construction.

Definition 25 *The chain $\boldsymbol{\Phi}$ is called recurrent if it is ψ-irreducible and $U(x,A) \equiv \infty$ for every $x \in \mathcal{X}$ and every $A \in \mathcal{B}^+(\mathcal{X})$. The chain $\boldsymbol{\Phi}$ is called transient if it is ψ-irreducible and $\mathcal{X}$ is transient.*

Next we will show that the split and original chains have mutually consistent recurrent/transient classifications.

Proposition 15 *If $\boldsymbol{\Phi}$ is ψ-irreducible and strongly aperiodic, then $\boldsymbol{\Phi}$ and $\check{\boldsymbol{\Phi}}$ are either both recurrent or both transient.*

Proof: Since $\mathbf{\Phi}$ is strongly aperiodic and ψ-irreducible, Proposition 10 guarantees that the minorization condition holds. Thus, we may split the chain to produce a chain $\check{\mathbf{\Phi}}$ on $\check{\mathcal{X}}$ which contains an accessible atom $\check{\boldsymbol{\alpha}}$. From Theorem 2, we have

$$\int_{\mathcal{X}} P^n(x, A)\lambda(dx) = \int_{\check{\mathcal{X}}} \check{P}^n(x_i, A_0 \cup A_1)\lambda^*(dx_i).$$

Take $\lambda(\cdot) = \delta_x(\cdot)$. Then, for all $x \in \mathcal{X}$ and $A \in \mathcal{B}^+(\mathcal{X})$,

$$\sum_{n=1}^{\infty} \int_{\check{\mathcal{X}}} \check{P}^n(x_i, A_0 \cup A_1)\delta_x^*(dx_i) = \sum_{n=1}^{\infty} P^n(x, A).$$

But $A \in \mathcal{B}^+(\mathcal{X})$ means $\psi(A) > 0$ implies $\psi^*(A_0 \cup A_1) > 0$. By the above equation, if $\check{\mathbf{\Phi}}$ is recurrent, then $\sum_{n=1}^{\infty} \check{P}^n(x_i, A_0 \cup A_1) = \infty$ which, in turn, implies $\sum_{n=1}^{\infty} P^n(x, A) = \infty$ and so $\mathbf{\Phi}$ is recurrent.

Now, if $\check{\mathbf{\Phi}}$ is transient, then by definition there is cover of $\check{\mathcal{X}}$ by uniformly transient sets. Therefore $\check{\mathcal{X}}$ is a transient set and by projecting the cover to $\mathcal{X}$ we see that $\mathbf{\Phi}$ is transient by definition.

Therefore $\check{\mathbf{\Phi}}$ is either recurrent or transient. This dichotomy carries over to $\mathbf{\Phi}$. □.

Lemma 4 *For $\epsilon \in (0, 1)$ we have*

$$\sum_{n=1}^{\infty} K_\epsilon^n(x, A) = \frac{1-\epsilon}{\epsilon} \sum_{n=0}^{\infty} P^n(x, A).$$

Proof: By repeated application of the generalized Chapman-Kolmogorov equations we have

$$\sum_{n=1}^{\infty} K_\epsilon^n(x, A) = \sum_{n=1}^{\infty} K_{\epsilon^{n*}}(x, A) = \sum_{n=0}^{\infty} b(n)P^n(x, A)$$

where $K_{\epsilon^{n*}}$ represents the n-fold convolution of K_ϵ and $b(k)$ is the coefficient of $P^k(x, A)$ in the expansion of the middle term above. That is, let a_ϵ denote the geometric distribution so that $a_\epsilon(k) = (1-\epsilon)\epsilon^k$. Then $b(k)$ is the kth term in the sequence $\sum_{n=1}^{\infty} a_\epsilon^{n*}$, or

$$b(k) = \left(\sum_{n=1}^{\infty} a_\epsilon^{n*}\right)(k) = (a_\epsilon + (a_\epsilon * a_\epsilon) + (a_\epsilon * a_\epsilon * a_\epsilon) + \cdots)(k).$$

Let

$$B(z) = \sum_{k=0}^{\infty} b(k)z^k \quad \text{and} \quad A(z) = \sum_{k=0}^{\infty} a_\epsilon(k)z^k,$$

then

$$A(z) = \frac{1-\epsilon}{1-\epsilon z} \quad \text{and} \quad B(z) = \sum_{n=1}^{\infty} (A(z))^n,$$

so that

$$B(z) = \left(\frac{1-\epsilon}{\epsilon}\right)(1-z)^{-1} = \frac{1-\epsilon}{\epsilon}\sum_{k=0}^{\infty} z^k, \quad |z| < 1.$$

Hence, $b(k) = (1-\epsilon)/\epsilon$ for all k by the uniqueness of the power series expansion. □

Now we can prove

Proposition 16 *If* $\mathbf{\Phi}$ *is* ψ*-irreducible, then*

(i) $\mathbf{\Phi}$ *is transient if and only if each* K_ϵ*-chain is transient.*

(ii) $\mathbf{\Phi}$ *is recurrent if and only if each* K_ϵ*-chain is recurrent.*

Proof: Suppose the K_ϵ-chain is transient, for $\epsilon \in (0,1)$. Then it is ψ-irreducible and $\mathcal{X}$ is a transient set by definition of a transient chain. But, by assumption $\mathbf{\Phi}$ is ψ-irreducible and hence it is transient. Now suppose that the K_ϵ-chain is recurrent. Then it is ψ-irreducible and $\sum_{n=1}^{\infty} K_\epsilon^n(x, A) = \infty$ for all $x \in \mathcal{X}$ and for every $A \in \mathcal{B}^+(\mathcal{X})$. But this imples $U(x, A) = \infty$ by the previous lemma and so $\mathbf{\Phi}$ is recurrent as well.

Next, suppose $\mathbf{\Phi}$ is transient. If the K_ϵ-chain were recurrent, then by (ii) above, $\mathbf{\Phi}$ would be recurrent. $\Rightarrow\Leftarrow$. Therefore the K_ϵ-chain is transient.

Suppose $\mathbf{\Phi}$ is recurrent. If the K_ϵ-chain were transient, then by (i) above, $\mathbf{\Phi}$ would be transient. $\Rightarrow\Leftarrow$. Therefore the K_ϵ-chain is recurrent. □

Now we can prove the main result of this section:

Theorem 9 *If* $\mathbf{\Phi}$ *is* ψ*-irreducible, then* $\mathbf{\Phi}$ *is either recurrent or transient.*

Proof: Since $\mathbf{\Phi}$ is ψ-irreducible, each K_ϵ-chain is strongly aperiodic (Proposition 10). By Proposition 15, each K_ϵ-chain is either recurrent or transient. This property is propogated to $\mathbf{\Phi}$ by Proposition 16. □

Theorem 10 *Suppose that* $\mathbf{\Phi}$ *is* ψ*-irreducible and aperiodic.*

(i) If there is an m*-skeleton chain* $\mathbf{\Phi}^m$ *that is transient, then* $\mathbf{\Phi}$ *is transient. If* $\mathbf{\Phi}$ *is transient, then every* m*-skeleton chain is transient.*

(ii) If there is an m*-skeleton chain* $\mathbf{\Phi}^m$ *that is recurrent, then* $\mathbf{\Phi}$ *is recurrent. If* $\mathbf{\Phi}$ *is recurrent, then every* m*-skeleton chain is recurrent.*

Proof:
(i) Suppose there exists an m such that $\mathbf{\Phi}^m$ is transient. Let A be a uniformly transient set. Then for all $x \in A$, there exists an $M < \infty$ such that

$$\sum_{j=0}^{\infty} P^{mj}(x, A) \leq M.$$

By the Chapman-Kolmogorov equations

$$\sum_{j=1}^{\infty} P^{j}(x, A) = \sum_{r=1}^{m} \int_{\mathcal{X}} P^{r}(x, dy) \sum_{j=0}^{\infty} P^{mj}(y, A) \leq mM.$$

Hence A is uniformly transient for $\mathbf{\Phi}$. Since $\mathbf{\Phi}^m$ is transient, by definition $\mathcal{X}$ is a transient set. Hence $\mathbf{\Phi}$ is transient.

Now suppose $\mathbf{\Phi}$ is transient. Let A be a uniformly transient set and $m > 0$ an integer. For all $x \in A$, there exists an $M < \infty$ such that

$$\sum_{j=1}^{\infty} P^{mj}(x, A) \leq \sum_{j=1}^{\infty} P^{j}(x, A) \leq M.$$

Therefore, A is a uniformly transient set for the m-skeleton. Since $\mathcal{X}$ is a uniformly transient set, then it is a uniformly transient set for the m-skeleton and so the m-skeleton is transient and m was arbitrary.

(ii) Suppose the m-skeleton chain is recurrent, for some m. Then

$$\sum_{j=1}^{\infty} P^{j}(x, A) = \sum_{r=1}^{m} \int_{\mathcal{X}} P^{r}(x, dy) \sum_{j=0}^{\infty} P^{mj}(y, A) = \infty.$$

Hence, $\mathbf{\Phi}$ is recurrent.

Now suppose that $\mathbf{\Phi}$ is recurrent. By Proposition 10, $\mathbf{\Phi}^m$ is ψ-irreducible for all m. But then by Theoerm 9, $\mathbf{\Phi}^m$ is either recurrent or transient. If it were transient, $\mathbf{\Phi}$ would be transient by (i). $\Rightarrow\Leftarrow$. Hence, $\mathbf{\Phi}^m$ is recurrent for all m. $\square$

1.6.1 Harris recurrence

In this section we will consider a stronger concept of recurrence, called Harris recurrence. To develop this concept, we will not only consider the first hitting time τ_A of a set $A \in \mathcal{B}(\mathcal{X})$, or the expected value $U(\cdot, A)$ of η_A, but also the event that $\mathbf{\Phi} \in A$ infinitely often (i.o.), or $\eta_A = \infty$. This event (set) is defined by

$$\{\mathbf{\Phi} \in A \,\text{i.o.}\} := \bigcap_{N=1}^{\infty} \bigcup_{k=N}^{\infty} \{\Phi_k \in A\}.$$

For $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$ let

$$Q(x, A) := P_x(\mathbf{\Phi} \in A \,\text{i.o.}).$$

Note that

$$Q(x, A) \leq L(x, A), \quad x \in \mathcal{X},\ A \in \mathcal{B}(\mathcal{X})$$

and by the strong Markov property

$$Q(x, A) = \mathbb{E}_x[P_{\Phi_{\tau_A}}(\mathbf{\Phi} \in A \,\text{i.o.})\mathbb{I}(\tau_A < \infty)] = \int_A U_A(x, dy)Q(y, A).$$

Definition 26 (Harris recurrence) *A set A is called Harris recurrent if*

$$Q(x, A) = P_x(\eta_A = \infty) = 1, \quad x \in A.$$

A chain $\mathbf{\Phi}$ is called Harris (recurrent) if it is ψ-irreducible and every set $A \in \mathcal{B}^+(\mathcal{X})$ is Harris recurrent.

A standard alternative definition of an Harris recurrence set is that $L(x, A) = 1$ for $x \in A$, however they are equivalent. Nevertheless, using $Q(x, A) = 1$ highlights the strengthening of recurrence to Harris recurrence from an expected infinite number of visits to a set to an almost surely infinite number of visits.

Proposition 17 *Suppose for some one set $A \in \mathcal{B}(\mathcal{X})$ we have $L(x, A) \equiv 1$ for $x \in A$. Then $Q(x, A) = L(x, A)$ for all $x \in \mathcal{X}$ and A is Harris recurrent.*

Proof:
Define $\tau_A(1) = \tau_A$ and for $k > 1$ define $\tau_A(k) = \min\{n > \tau_A(k-1) : \Phi_n \in A\}$. If $L(y, A) = 1$ for all $y \in A$, then

$$P_x(\tau_A(2) < \infty) = \int_A U_A(x, dy)L(y, A) = \int_A U_A(x, dy) = U_A(x, A) = L(x, A) = 1,$$

for all $x \in A$ by the strong Markov property. Inductively, then,

$$P_x(\tau_A(k+1) < \infty) = \int_A U_A(x, dy) P_y(\tau_A(k) < \infty) = 1.$$

Also, we have

$$P_x(\eta_A \geq k) = P_x(\tau_A(k) < \infty) = 1, \quad \forall\, x \in \mathcal{A}$$

and by monotone convergence

$$Q(x, A) = \lim_k P_x(\eta_A \geq k) = 1, \quad \forall\, x \in \mathcal{A}$$

hence it follows that $Q(x, A) \equiv 1$ for all $x \in A$. Now for $x \in \mathcal{X}$

$$Q(x, A) = \int_A U_A(x, dy) Q(y, A) = \int_A U_A(x, dy) = U_A(x, A) = L(x, A).$$

□

The most difficult proof of this section is the following.

Theorem 11 *(i) Suppose that $D \rightsquigarrow A$ for any D, $A \in \mathcal{B}(\mathcal{X})$. Let μ be any initial distribution of the chain $\mathbf{\Phi}$. Then*

$$\{\mathbf{\Phi} \in D \; i.o.\} \subseteq \{\mathbf{\Phi} \in A \; i.o.\} \quad a.s.\ [P_\mu]$$

and $Q(y, D) \leq Q(y, A)$ for all $y \in \mathcal{X}$.

(ii) If $\mathcal{X} \rightsquigarrow A$, then A is Harris recurrent and $Q(x, A) \equiv 1$ for all $x \in \mathcal{X}$.

Proof: See Meyn & Tweedie, p. 202. The proof involves Martingales and would take too much time to give the necessary background to show/prove the Martingale Convergence Theorem.

This leads us to the following strengthening of Harris recurrence...

Theorem 12 *If $\mathbf{\Phi}$ is Harris recurrent, then $Q(x, B) = 1$ for every $x \in \mathcal{X}$ and every $B \in \mathcal{B}^+(\mathcal{X})$.*

Proof: Let C_n, $n \in \mathbb{N}_+$ be petite sets with $\cup_n C_n = \mathcal{X}$. Now, the finite union of petite sets is petite. Further every small set is petite. Since we have already established that there are

many small sets in $\mathcal{B}^+(\mathcal{X})$, there are many petite sets in $\mathcal{B}^+(\mathcal{X})$ as well. Given these facts, without loss of generality, we can assume that $C_n \subset C_{n+1}$ and that $C_n \in \mathcal{B}^+(\mathcal{X})$ for all n.

Let $B \in \mathcal{B}^+(\mathcal{X})$ and $n \in \mathbb{N}_+$. Since C_n is petite, it follows that $C_n \overset{a}{\rightsquigarrow} B$ for some discrete distribution a on $\mathbb{N}_+$ (This is just a consequence of the definition of a petite set and the definition of $\overset{a}{\rightsquigarrow}$). Hence $C_n \rightsquigarrow B$ by Lemma 1. Now by the definition of a Harris recurrent chain, C_n is a Harris recurrent set, therefore by Theorem 11(i)

$$Q(x, B) \geq Q(x, C_n), \quad \forall x \in \mathcal{X},$$

and $Q(x, B) = Q(x, C_n) = 1$ for all $x \in C_n$. But $\cup_n C_n = \mathcal{X}$, and since both $B \in \mathcal{B}^+(\mathcal{X})$ and n are arbitrary, we have that $Q(x, B) = 1$ for all $x \in \mathcal{X}$ and $B \in \mathcal{B}^+(\mathcal{X})$. □

Let D be any Harris recurrent set. Let $D^\infty = \{y : L(y, D) = 1\}$. Then $D \subseteq D^\infty$ and D^∞ is absorbing. We call D a maximal absorbing set if $D = D^\infty$.

Definition 27 (Maximal Harris sets) *A set H is called maximal Harris if H is a maximal absorbing set such that $\mathbf{\Phi}$ restricted to H is Harris recurrent.*

Recall the following definitions

Definition 28 (Full and Absorbing Sets) *A set $A \in \mathcal{B}(\mathcal{X})$ is said to be*

(i) full if $\psi(A^c) = 0$.

(ii) absorbing if $P(x, A) = 1$ for $x \in A$.

In order to prove Theorem 13 we need the following three lemmas:

Lemma 5 *Suppose that $\mathbf{\Phi}$ is ψ-irreducible. Then*

(i) every absorbing set is full.

(ii) every full set contains a non-empty absorbing set.

Proof:
(i) Let A be absorbing. Suppose A is not full, then $\psi(A^c) > 0$. Let $x \in A$. By ψ-irreducibility, for every $y \in \mathcal{X}$ there exists an n such that $P^n(y, A^c) > 0$. So it is certainly true for this x.

But A is absorbing so $P(x, A) = 1$ for all $x \in A$. This implies $P^n(x, A) = 1$ for $x \in A$. But then it must be that $P^n(x, A^c) = 0$. $\Rightarrow\Leftarrow$. Therefore $\psi(A^c) = 0$ and A is full.

(ii) Let A be a full set. Set

$$B = \left\{ y \in \mathcal{X} : \sum_{n=0}^{\infty} P^n(y, A^c) = 0 \right\}.$$

Now, $B \subseteq A$. Why? Suppose not. Let $y \in B$ and suppose $y \notin A$. Then $y \in A^c$ and since $P^0(y, A^c) = \delta_y(A^c) = 1$ for all $y \in A^c$ it must be that $y \notin B$. $\Rightarrow\Leftarrow$. Hence $y \in A$.

Since A is full, $\psi(A^c) = 0$ and so $\psi(B) > 0$, hence B is non-empty. Now we need to show that $P(y, B) = 1$ for all $y \in B$. Suppose not, then $P(y, B^c) > 0$ for some $y \in B$. So, by the Chapman-Kolmogorov equations we have

$$\begin{aligned}
\sum_{n=0}^{\infty} P^{n+1}(y, A^c) &= \sum_{n=0}^{\infty} \int_{\mathcal{X}} P(y, dz) P^n(z, A^c) \\
&\geq \sum_{n=0}^{\infty} \int_{B^c} P(y, dz) P^n(z, A^c) \\
&= \int_{B^c} P(y, dz) \left(\sum_{n=0}^{\infty} P^n(z, A^c) \right) > 0.
\end{aligned}$$

This contradicts the definition of B so that $P(x, B^c) = 0$. Hence $B \subseteq A$, and is non-empty and absorbing. (Note also from (i) that B is full as well). □

Lemma 6 *If $\mathbf{\Phi}$ is a Markov chain and if $A \in \mathcal{B}(\mathcal{X})$ satisfies $L(x, A) \leq \epsilon < 1$ for $x \in A$, then $U(x, A) \leq (1 - \epsilon)^{-1}$ for all $x \in \mathcal{X}$.*

Proof: Suppose $A \in \mathcal{B}(\mathcal{X})$ and $L(x, A) \leq \epsilon < 1$ for $x \in A$. The last-exit decomposition gives us

$$U^{(z)}(x, A) = U_A^{(z)}(x, A) + \int_A U^{(z)}(x, dy) U_A^{(z)}(y, A) \leq 1 + \epsilon U^{(z)}(x, A).$$

Hence

$$U(x, A) = \lim_{z \uparrow 1} U^{(z)}(x, A) \leq (1 - \epsilon)^{-1}.$$

□

Lemma 7 *If $\mathbf{\Phi}$ is ψ-irreducible and $A \in \mathcal{B}(\mathcal{X})$ with $\psi(A) = 0$, then A is transient.*

Proof: See Meyn & Tweedie, p. 186.

Theorem 13 *If* $\mathbf{\Phi}$ *is recurrent, then*

$$\mathcal{X} = H \cup N$$

where H *is a non-empty maximal Harris set and* N *is transient. Furthermore,* $\psi(N) = 0$.

Proof: Let $C \in \mathcal{B}^+(\mathcal{X})$ be ψ_a-petite with ψ_a a maximal irreducibility measure. Let

$$H = \{x : Q(x, C) = 1\} \quad \text{and} \quad N = H^c$$

Note that $H = H^\infty$. Hence, H is either maximal absorbing or H is empty.

I claim H is non-empty. Suppose not, then $H = \emptyset$. We will show that this contradicts the assumption that $\mathbf{\Phi}$ is recurrent. Since H is empty we have that $Q(x, C) < 1$ for all $x \in \mathcal{X}$. This implies that

$$C_1 := \{x \in C : L(x, C) < 1\} \in \mathcal{B}^+(\mathcal{X}).$$

> *To see that this is the case, suppose not. Then* $\psi(C_1) = 0$. *Hence, by Lemma 5(ii), since* C_1^c *is full, it contains a non-empty absorbing set. Call it* F. *Since* $F \subset C_1^c$, *by definition we have* $L(x, C) = 1$ *for all* $x \in C \cap F$. *Also, since* F *is absorbing we must have* $L(x, C \cap F) = 1$ *for all* $x \in C \cap F$. *But by Proposition 17,* $Q(x, C \cap F) = L(x, C \cap F)$ *for all* $x \in \mathcal{X}$. *So*
>
> $$1 > Q(x, C) \geq Q(x, C \cap F) = 1. \quad \Rightarrow\Leftarrow .$$
>
> *Hence,* $\psi(C_1) > 0$, *or* $C_1 \in \mathcal{B}^+(\mathcal{X})$.

Now, since $C_1 \in \mathcal{B}^+(\mathcal{X})$ there exists $B \subseteq C_1$, such that $B \in \mathcal{B}^+(\mathcal{X})$ and a $\delta > 0$ such that $L(x, C_1) \leq \delta < 1$ for all $x \in B$. Therefore,

$$L(x, B) \leq L(x, C_1) \leq \delta, \quad x \in B.$$

So, by Lemma 6 we have $U(x, B) \leq (1 - \delta)^{-1}$, for $x \in B$. Thus contradicting the fact that $\mathbf{\Phi}$ is recurrent and so H is a non-empty maximal absorbing set. Therefore, by Lemma 5(i), H is full. Then, since H is full, $\psi(H^c) = \psi(N) = 0$. Hence by Lemma 7, N is transient.

The last thing we need to show is that H is Harris recurrent. For any $A \in \mathcal{B}^+(\mathcal{X})$ we have $C \rightsquigarrow A$ (by definitions). Then, from Theorem 11(i), $Q(x, C) \leq Q(x, A)$ for all $A \in \mathcal{B}^+(\mathcal{X})$ and so since $Q(x, C) = 1$ we have $Q(x, A) = 1$. But then for all $x \in H$ we have $Q(x, C) = 1$,

by construction of H, we have that $Q(x, A) = 1$ for all $x \in H$ and $A \in \mathcal{B}^+(\mathcal{X})$. Therefore, $\mathbf{\Phi}$ restricted to H is Harris recurrent. □

This theorem states that a recurrent Markov chain and a Harris recurrent Markov chain differ only by the existence of a ψ-null set N on which recurrence does not hold.

Theorem 14 *Suppose that $\mathbf{\Phi}$ is ψ-irreducible and aperiodic. Then $\mathbf{\Phi}$ is Harris if and only if each skeleton is Harris.*

Proof:
($\Leftarrow$) Suppose the m-skeleton is Harris recurrent. Then, since $m\tau_A^m \geq \tau_A$ for any $A \in \mathcal{B}(\mathcal{X})$, where τ_A^m is the first entrance time for the m-skeleton, it immediately follows that $\mathbf{\Phi}$ is Harris recurrent since

$$P_x(\tau_A < \infty) \geq P_x(\tau_A^m < \infty) \equiv 1.$$

($\Rightarrow$) Suppose $\mathbf{\Phi}$ is Harris recurrent. For $m \geq 2$ we know that $\mathbf{\Phi}^m$ is recurrent, hence a Harris set H_m exists for the m-skeleton. But since H_m is full, there is a set $H \subseteq H_m$ that is absorbing and hence full as well. Since $\mathbf{\Phi}$ is Harris recurrent, we have $P_x(\tau_H < \infty) \equiv 1$. Since H is absorbing, we have $m\tau_H^m \leq \tau_H + m$. Thus

$$P_x(\tau_H^m < \infty) \geq P_x(\tau_H < \infty) \equiv 1.$$

So $P_x(\tau_H^m < \infty) = 1$ and $\mathbf{\Phi}^m$ is Harris recurrent. □

The existence of a stationary distribution

Now we investigate the conditions necessary for a Markov chain to possess a stationary distribution. This is quite important for the development of MCMC theory—for we want to be able to sample from the posterior distribution, which is stationary. It is quite obvious that if we construct a Markov chain that does not possess a stationary distribution, we will not be sampling from the posterior. More importantly, we need to also show that the n-step transition kernels converge to this stationary distribution, no matter what the initial distribution, and will do so in the next section. Thus once $P^n(\mu, A) = \pi(A)$ we are guaranteed that $P^{n+1}(\mu, A) = \pi(A)$ as well (once we start sampling from the posterior distribution, we will always be sampling from the posterior).

Definition 29 (Invariant measure) *If a σ-finite measure π on $\mathcal{B}(\mathcal{X})$ has the property*

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

then we call π invariant.

If an invariant measure is finite, then we can renormalize it to be a probability measure. Obviously, this is the situation in which we are most interested.

Definition 30 (Positive and null chains) *If $\mathbf{\Phi}$ is ψ-irreducible and admits an invariant probability measure π, then we call $\mathbf{\Phi}$ a positive chain. If $\mathbf{\Phi}$ does not admit such a measure, we call it null.*

Recall that a process is stationary if for any k, the marginal distribution of $\{\Phi_n, \ldots, \Phi_{n+k}\}$ does not change as n varies. In general, Markov chains are not stationary (e.g. consider the a chain with initial distribution δ_x). However, with an appropriate choice for the initial distribution for Φ_0 we can produce a stationary process $\{\Phi_n, n \in \mathbb{N}_+\}$.

For Markov chains we only need to consider first step stationarity to generate an entire stationary process. Suppose π is the initial invariant measure (initial distribution of the chain) with

$$\pi(A) = \int_{\mathcal{X}} \pi(dw) P(w, A).$$

Now, by iterating and the Chapman-Kolmogorov equations we have all n and $A \in \mathcal{B}(\mathcal{X})$

$$\begin{aligned}
\pi(A) &= \int_{\mathcal{X}} \pi(dw) P(w, A) \\
&= \int_{\mathcal{X}} \left[\int_{\mathcal{X}} \pi(dx) P(x, dw) \right] P(w, A) \\
&= \int_{\mathcal{X}} \pi(dx) \int_{\mathcal{X}} P(x, dw) P(w, A) \\
&= \int_{\mathcal{X}} \pi(dx) P^2(x, A) \\
&= \vdots \\
&= \int_{\mathcal{X}} \pi(dx) P^n(x, A) = P_\pi(\Phi_n \in A).
\end{aligned}$$

From the Markov property $\mathbf{\Phi}$ is stationary if and only if the distribution of Φ_n does not depend on n (time).

Lemma 8 *Let $\mathbf{\Phi}$ be a Markov chain and if $A \in \mathcal{B}(\mathcal{X})$ is uniformly transient with $U(x, A) \leq M$ for $x \in A$, then $U(x, A) \leq 1 + M$ for all $x \in \mathcal{X}$.*

Proof: Use the first-entrance decomposition. In the limit we have

$$\begin{aligned} U(x,A) &= U_A(x,A) + \int_A U_A(x,dy)U(y,A) \\ &\leq L(x,A) + \int_A U_A(x,dy) \sup_{y\in A} U(y,A) \\ &= L(x,A) + L(x,A) \sup_{y\in A} U(y,A) \\ &\leq 1 + M. \end{aligned}$$

Proposition 18 *If the chain* $\mathbf{\Phi}$ *is positive, then it is recurrent.*

Proof: Suppose that the chain is transient. Then there is a countable cover of $\mathcal{X}$ with uniformly transient sets A_j. Hence, there exists an M_j such that $U(x, A_j) \leq M_j$ by the previous lemma. Now for any j, k we have

$$\pi(A_j) = k^{-1} \sum_{n=1}^{k} \int_{\mathcal{X}} \pi(dx) P^n(x, A_j) \leq k^{-1} M_j$$

As $k \uparrow \infty$ we have $\pi(A_j) = 0$. Therefore π cannot be a probability measure and $\mathbf{\Phi}$ is null. $\Rightarrow\Leftarrow$. □

Positive chains are often referred to as positive recurrent to reinforce the fact that they are recurrent.

Definition 31 (Positive Harris chains) *If* $\mathbf{\Phi}$ *is Harris recurrent and positive, then* $\mathbf{\Phi}$ *is called a positive Harris (recurrent) chain.*

Now we set out to show that an invariant probability measure exists and that it is unique, up to a multiplicative constant, for certain chains. We will begin by showing that chains that admit atoms are positive, and then extend to strongly aperiodic chains and then to recurrent chains.

Definition 32 (Subinvariant measures) *if* μ *is* σ*-finite and satisfies*

$$\mu(A) \geq \int_{\mathcal{X}} \mu(dw) P(w, A), \quad A \in \mathcal{B}(\mathcal{X}) \tag{13}$$

then μ *is called subinvariant.*

Iterating we get

$$\mu(A) \geq \int_{\mathcal{X}} \mu(dx) P^n(x, A). \tag{14}$$

Multiplying by $a(n)$, where a is a sampling distribution on $\mathbb{N}_+$, and then summing we get

$$\mu(A) \geq \int_{\mathcal{X}} \mu(dx) K_a(x, A). \tag{15}$$

Equations (14) and (15) tell us, respectively, that if μ is a subinvariant measure for $\mathbf{\Phi}$ is it also a subinvariant measure for any m-skeleton and for any sampled chain.

Proposition 19 *If $\mathbf{\Phi}$ is ψ-irreducible and if μ is any measure satisfying (13) with $\mu(A) < \infty$ for some $A \in \mathcal{B}^+(\mathcal{X})$, then*

(i) μ is σ-finite and thus μ is a subinvariant measure;

(ii) $\psi \prec \mu$;

(iii) if C is ν_a-petite, then $\mu(C) < \infty$;

(iv) if $\mu(\mathcal{X}) < \infty$, then μ is invariant.

Proof:
(i) Suppose $A \in \mathcal{B}^+(\mathcal{X})$ and $\mu(A) < \infty$. Consider the sets

$$A^*(j) = \{y : K_{1/2}(y, A) > j^{-1}\}.$$

Then

$$\begin{aligned} \infty > \mu(A) &\geq \int_{\mathcal{X}} \mu(dx) K_{1/2}(x, A) \\ &\geq \int_{A^*(j)} \mu(dx) K_{1/2}(x, A) \\ &> j^{-1} \mu(A^*(j)). \end{aligned}$$

So, each $A^*(j)$ has μ-finite measure. Furthermore since, $\lim_{j \uparrow \infty} A^*(j) = \cup_j A^*(j) = \mathcal{X}$, μ is σ-finite.

(ii) Let $A \in \mathcal{B}^+(\mathcal{X})$, i.e. $\psi(A) > 0$. Since $\mathbf{\Phi}$ is ψ-irreducible, $K_{1/2}(x, A) > 0$ which implies

$$\mu(A) \geq \int_{\mathcal{X}} \mu(dx) K_{1/2}(x, A) > 0.$$

Hence, $\mu(A) > 0$ whenever $\psi(A) > 0$, or $\psi \prec \mu$.

(*iii*) Suppose C is ν_a-petite. Then ν_a is a non-trivial measure and

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

for all $B \in \mathcal{B}(\mathcal{X})$ and $x \in \mathcal{C}$. Hence, there exists a set $A \in \mathcal{B}(\mathcal{X})$ with $\nu_a(A) > 0$ and, by assumption, $\mu(A) < \infty$. So, by (i),

$$\infty > \mu(A) \geq \int_{\mathcal{X}} \mu(dx) K_a(x, A) \geq \int_C \mu(dx) K_a(x, A) \geq \mu(C)\nu_a(A)$$

so that $\mu(C) < \infty$.

(*iv*) Suppose not. Suppose $\mu(\mathcal{X}) < \infty$ and μ is not invariant. Then there exists an A such that $\mu(A) > \int_{\mathcal{X}} \mu(dx)P(x, A)$.

$$\begin{aligned} \mu(\mathcal{X}) = \mu(A) + \mu(A^c) &> \int_{\mathcal{X}} \mu(dx)P(x, A) + \int_{\mathcal{X}} \mu(dx)P(x, A^c) \\ &= \int_{\mathcal{X}} \mu(dx)P(x, \mathcal{X}) \\ &= \mu(\mathcal{X}). \end{aligned}$$

This implies that $\mu(\mathcal{X}) = \infty$. $\Rightarrow\Leftarrow$. Hence, μ must be invariant. □

1.6.2 The existence of an invariant measure—chains with atoms

We are interested in Harris recurrent ψ-irreducible chains for MCMC theory. However, to show the existence of an invariant measure for recurrent ψ-irreducible chain, we will first proof the existence for chains with atoms (not necessarily recurrent) and then use Nummelin's splitting technique to extend the results to recurrent chains.

Lemma 9 *Suppose Φ is a Markov chain. Let $A \in \mathcal{B}(\mathcal{X})$. If $L(x, A) = 1$ for all $x \in A$, then A is a recurrent set.*

Proof: Suppose $L(x, A) = 1$ for all $x \in A$. Use the last-exit decomposition to get

$$U^{(z)}(x, A) = U_A^{(z)}(x, A) + \int_A U^{(z)}(x, dy) U_A^{(z)}(y, A).$$

Now take the limit as $z \uparrow 1$. Then

$$U(x, A) = L(x, A) + L(x, A)U(x, A) = 1 + U(x, A).$$

Therefore $U(x, A) = \infty$ for all $x \in A$ and A is recurrent by definition. □

Theorem 15 *Let $\mathbf{\Phi}$ be ψ-irreducible and suppose $\mathcal{X}$ contains an accessible atom $\boldsymbol{\alpha}$.*

(i) There exists a subinvariant measure $\mu^o_{\boldsymbol{\alpha}}$ for $\mathbf{\Phi}$ given by

$$\mu^o_{\boldsymbol{\alpha}}(A) = U_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}, A) = \sum_{n=1}^{\infty} {}_{\boldsymbol{\alpha}}P^n(\boldsymbol{\alpha}, A), \quad \forall\, A \in \mathcal{B}(\mathcal{X}),$$

where $\mu^o_{\boldsymbol{\alpha}}$ is invariant if and only if $\mathbf{\Phi}$ is recurrent.

(ii) The measure $\mu^o_{\boldsymbol{\alpha}}$ is minimal in the sense that if μ is subinvariant with $\mu(\boldsymbol{\alpha}) = 1$ then

$$\mu(A) \geq \mu^o_{\boldsymbol{\alpha}}(A), \quad \forall\, A \in \mathcal{B}(\mathcal{X}).$$

When $\mathbf{\Phi}$ is recurrent, $\mu^o_{\boldsymbol{\alpha}}$ is the unique (sub)invariant measure with $\mu(\boldsymbol{\alpha}) = 1$.

(iii) The subinvariant measure $\mu^o_{\boldsymbol{\alpha}}$ is a finite measure if and only if

$$\mathbb{E}_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}}) < \infty,$$

in which case $\mu^o_{\boldsymbol{\alpha}}$ is invariant.

Proof: (i) Let $A \in \mathcal{B}(\mathcal{X})$. Then

$$\begin{aligned}
\int_{\mathcal{X}} \mu^o_{\boldsymbol{\alpha}}(dy)P(y, A) &= \int_{\boldsymbol{\alpha}} \mu^o_{\boldsymbol{\alpha}}(dy)P(y, A) + \int_{\boldsymbol{\alpha}^c} \mu^o_{\boldsymbol{\alpha}}(dy)P(y, A) \\
&= \mu^o_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})P(\boldsymbol{\alpha}, A) + \int_{\boldsymbol{\alpha}^c} \sum_{n=1}^{\infty} {}_{\boldsymbol{\alpha}}P^n(\boldsymbol{\alpha}, dy)P(y, A) \\
&\leq P(\boldsymbol{\alpha}, A) + \sum_{n=1}^{\infty} \int_{\boldsymbol{\alpha}^c} {}_{\boldsymbol{\alpha}}P^n(\boldsymbol{\alpha}, dy)P(y, A) \\
&= {}_{\boldsymbol{\alpha}}P(\boldsymbol{\alpha}, A) + \sum_{n=2}^{\infty} {}_{\boldsymbol{\alpha}}P^n(\boldsymbol{\alpha}, A) \\
&= \mu^o_{\boldsymbol{\alpha}}(A).
\end{aligned}$$

Hence, $\mu^o_{\boldsymbol{\alpha}}$ is a subinvariant measure.

Now, $\mu^o_{\boldsymbol{\alpha}}$ is invariant if and only if $\mu^o_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) = 1$. But by definition, $\mu^o_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) = U_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = L(\boldsymbol{\alpha}, \boldsymbol{\alpha})$. But if $L(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = 1$, then $\mathbf{\Phi}$ is recurrent by Lemma 9.

(ii) Let μ be any subinvariant measure with $\mu(\boldsymbol{\alpha}) = 1$. We have

$$\begin{aligned}
\mu(A) &\geq \int_{\mathcal{X}} \mu(dy)P(y, A) \\
&\geq \int_{\boldsymbol{\alpha}} \mu(dy)P(y, A) \\
&= \mu(\boldsymbol{\alpha})P(\boldsymbol{\alpha}, A) = P(\boldsymbol{\alpha}, A).
\end{aligned}$$

Now assume that $\mu(A) \geq \sum_{m=1}^{n} {}_{\boldsymbol{\alpha}}P^m(\boldsymbol{\alpha}, A)$ for all A. Then by subinvariance

$$\begin{aligned}
\mu(A) &\geq \mu(\boldsymbol{\alpha})P(\boldsymbol{\alpha}, A) + \int_{\boldsymbol{\alpha}^c} \mu(dx)P(x, A) \\
&\geq P(\boldsymbol{\alpha}, A) + \int_{\boldsymbol{\alpha}^c} \left(\sum_{m=1}^{n} {}_{\boldsymbol{\alpha}}P^m(\boldsymbol{\alpha}, dx) \right) P(x, A) \\
&= \sum_{m=1}^{n+1} {}_{\boldsymbol{\alpha}}P^m(\boldsymbol{\alpha}, A).
\end{aligned}$$

Taking the limit as $n \uparrow \infty$ shows that $\mu(A) \geq \mu^o_{\boldsymbol{\alpha}}(A), \quad \forall\, A \in \mathcal{B}(\mathcal{X})$.

Next, suppose $\boldsymbol{\Phi}$ is recurrent so that $\mu^o_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) = 1$. If μ and $\mu^o_{\boldsymbol{\alpha}}$ differ, then $\mu(A) > \mu^o_{\boldsymbol{\alpha}}(A)$ for some $A \in \mathcal{B}(\mathcal{X})$. By ψ-irreducibility there exists an n such that $P^n(x, \boldsymbol{\alpha}) > 0$ for all $x \in \mathcal{X}$, since $\psi(\boldsymbol{\alpha}) > 0$ ($\boldsymbol{\alpha}$ is an accessible atom). Then

$$\begin{aligned}
1 = \mu(\boldsymbol{\alpha}) &\geq \int_{\mathcal{X}} \mu(dx)P^n(x, \boldsymbol{\alpha}) \\
&= \int_{A} \mu(dx)P^n(x, \boldsymbol{\alpha}) + \int_{A^c} \mu(dx)P^n(x, \boldsymbol{\alpha}) \\
&> \int_{A} \mu^o_{\boldsymbol{\alpha}}(dx)P^n(x, \boldsymbol{\alpha}) + \int_{A^c} \mu^o_{\boldsymbol{\alpha}}(dx)P^n(x, \boldsymbol{\alpha}) \\
&= \int_{\mathcal{X}} \mu^o_{\boldsymbol{\alpha}}(dx)P^n(x, \boldsymbol{\alpha}) = \mu^o_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) = 1.
\end{aligned}$$

This leads to a contradiction so that $\mu(A) = \mu^o_{\boldsymbol{\alpha}}(A)$. Therefore, $\mu = \mu^o_{\boldsymbol{\alpha}}$ and $\mu^o_{\boldsymbol{\alpha}}$ is the unique (sub)invariant measure.

(*iii*) If $\mu^o_{\boldsymbol{\alpha}}$ is finite, then it is invariant (by Proposition 19(*iv*) on page 45). Also by the construction of $\mu^o_{\boldsymbol{\alpha}}$ we have

$$\mu^o_{\boldsymbol{\alpha}}(\mathcal{X}) = \sum_{n=1}^{\infty} {}_{\boldsymbol{\alpha}}P^n(\boldsymbol{\alpha}, \mathcal{X}) = \sum_{n=1}^{\infty} P_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}} \geq n) = \mathbb{E}_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}}).$$

Therefore an invariant probability measure exists if and only if the mean return time to $\boldsymbol{\alpha}$ is finite. In the above the first equality holds by definition. The second equality holds because

$${}_{\boldsymbol{\alpha}}P^n(\boldsymbol{\alpha}, \mathcal{X}) := P_{\boldsymbol{\alpha}}(\Phi_n \in \mathcal{X}; \tau_{\boldsymbol{\alpha}} \geq n) = P_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}} \geq n).$$

The third equality holds because

$$\sum_{n=1}^{\infty} P_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}} \geq n) = \sum_{n=1}^{\infty} \mathbb{E}_{\boldsymbol{\alpha}}[\mathbb{I}(\tau_{\alpha} \geq n)] = \mathbb{E}_{\boldsymbol{\alpha}} \left[\sum_{n=1}^{\infty} \mathbb{I}(\tau_{\alpha} \geq n) \right] = \mathbb{E}_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}}).$$

□

If $\mathbf{\Phi}$ is recurrent, then we will use π to denote the unique invariant measure. That is $\pi = \mu^0_{\boldsymbol{\alpha}}$.

This leads us to the following criterion for positivity.

Theorem 16 (Kac's Theorem) *If $\mathbf{\Phi}$ is ψ-irreducible and admits an atom $\boldsymbol{\alpha} \in \mathcal{B}^+(\mathcal{X})$ (an accessible atom), then $\mathbf{\Phi}$ is positive recurrent if and only if $\mathbb{E}_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}}) < \infty$. If π is the invariant probability measure for $\mathbf{\Phi}$, then*

$$\pi(\boldsymbol{\alpha}) = \mathbb{E}^{-1}_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}}).$$

Proof:
($\Leftarrow$) Suppose $\mathbb{E}_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}}) < \infty$. Then $L(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = 1$. Why?

Suppose not.

$$L(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = P_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}} < \infty) < 1 \implies P_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}} = \infty) > 0 \implies \mathbb{E}_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}}) = \infty (\Rightarrow\Leftarrow).$$

But then $\mathbf{\Phi}$ is recurrent by Lemma 9. Let $\boldsymbol{\alpha}$ be an accessible atom and let $\mu^o_{\boldsymbol{\alpha}}$ be the unique invariant measure (guaranteed by the last Theorem). Then $\mu^o_{\boldsymbol{\alpha}}(\mathcal{X}) = \mathbb{E}_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}}) < \infty$, so that π is a finite measure and thus $\mathbf{\Phi}$ is positive.

($\Rightarrow$) Suppose $\mathbf{\Phi}$ is positive recurrent. Let π be its unique, (finite), invariant probability measure. Then $\mathbb{E}_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}}) = \mu^o_{\boldsymbol{\alpha}}(\mathcal{X}) < \infty$.

Now

$$\pi(\boldsymbol{\alpha}) = \frac{\mu^o_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})}{\mu^o_{\boldsymbol{\alpha}}(\mathcal{X})} = \frac{U_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}, \boldsymbol{\alpha})}{\mu^o_{\boldsymbol{\alpha}}(\mathcal{X})} = \frac{L(\boldsymbol{\alpha}, \boldsymbol{\alpha})}{\mu^o_{\boldsymbol{\alpha}}(\mathcal{X})} = \frac{1}{\mathbb{E}_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}})}.$$

□

1.6.3 The existence of an invariant measure—recurrent chains

Proposition 20 *Let $\mathbf{\Phi}$ be a strongly aperiodic Markov chain and let $\check{\mathbf{\Phi}}$ denote the split chain. Then*

(i) if the measure $\check{\pi}$ is invariant for $\check{\mathbf{\Phi}}$ then the measure π on $\mathcal{B}(\mathcal{X})$ defined by

$$\pi(A) = \check{\pi}(A_0 \cup A_1), \quad A \in \mathcal{B}(\mathcal{X}),$$

is invariant for $\mathbf{\Phi}$ and $\check{\pi} = \pi^*$.

(ii) If μ is any subinvariant measure for $\mathbf{\Phi}$ then μ^* is subinvariant for $\check{\mathbf{\Phi}}$, and if μ is invariant then so is μ^*.

Proof: Follows from the development of the Nummelin splitting technique back in section 1.5.1, page 11. We first show that $\check{\pi} = \pi^*$. First, by equations (3), (4) and (5) on page 14, $\check{P}(x_i, \cdot) = P^*(x_i, \cdot)$. By the linearity of the splitting and by the invariance of $\check{\pi}$, for any $A \in \mathcal{B}(\mathcal{X})$,

$$\begin{aligned}\check{\pi}(A_0 \cup A_1) &= \int_{\check{\mathcal{X}}} \check{\pi}(dx_i)\check{P}(x_i, A_0 \cup A_1) = \int_{\check{\mathcal{X}}} \check{\pi}(dx_i)P^*(x_i, A_0 \cup A_1) \\ &= \left(\int_{\check{\mathcal{X}}} \check{\pi}(dx_i)P(x_i, \cdot)\right)^* (A_0 \cup A_1) = \pi_0^*(A_0 \cup A_1).\end{aligned}$$

Therefore $\check{\pi} = \pi_0^*$ where $\pi_0(\cdot) = \int_{\check{\mathcal{X}}} \check{\pi}(dx_i)P(x_i, \cdot)$. Now, by the definition of π above, we have

$$\pi(A) = \check{\pi}(A_0 \cup A_1) = \pi_0^*(A_0 \cup A_1) = \pi_0(A)$$

implies

$$\pi^*(A_0 \cup A_1) = \pi_0^*(A_0 \cup A_1) = \check{\pi}(A_0 \cup A_1).$$

Hence, $\pi^* = \check{\pi}$. Now we show that π is an invariant measure for $\mathbf{\Phi}$. For any $A \in \mathcal{B}(\mathcal{X})$ we have by the invariance of π^*

$$\pi(A) = \pi^*(A_0 \cup A_1) = \pi_0^*(A_0 \cup A_1) = \left(\int_{\check{\mathcal{X}}} \pi^*(dx_i)\check{P}(x, \cdot)\right)^* (A_0 \cup A_1) = \int_{\mathcal{X}} \pi(dx)P(x, A).$$

Now we prove part (ii). If μ is subinvariant, we have

$$\mu^*(A_0 \cup A_1) = \mu(A) \geq \int_{\mathcal{X}} \mu(dx)P(x, A) = \int_{\check{\mathcal{X}}} \mu^*(dx_i)\check{P}(x_i, A_0 \cup A_1),$$

where the second equality follows from Theorem 2(i). If μ is strictly invariant, then so must be μ^*. □

So now we have that a strongly aperiodic Markov chain admits an invariant measure if the split chain does. This leads us to

Proposition 21 *If $\mathbf{\Phi}$ is recurrent and strongly aperiodic, then $\mathbf{\Phi}$ admits a unique (up to constant multiples) subinvariant measure that is invariant.*

Proof: Assume that $\mathbf{\Phi}$ is strongly aperiodic. Split the chain. Let $\check{\mathbf{\Phi}}$ denote the split chain. Since $\mathbf{\Phi}$ is recurrent, then $\check{\mathbf{\Phi}}$ is also recurrent by Proposition 15. But $\check{\mathbf{\Phi}}$ contains an accessible

atom, thus Theorem 15 guarantees that $\check{\mathbf{\Phi}}$ has a unique subinvariant measure $\check{\pi}$ that is invariant. Therefore π is invariant for Φ by Proposition 20. Now uniqueness. Suppose μ is another subinvariant measure for $\mathbf{\Phi}$. Then μ^* is invariant for $\check{\mathbf{\Phi}}$ (Proposition 20). But this measure is unique (up to a constant) by Proposition 15. Thus $\mu^* = c\check{\pi}$. But, by the linearity of μ^* and $\check{\pi}$, $\mu = c\pi$. □

Now we use the fact that the resolvent chain is strongly aperiodic to lift this result to the whole chain, even when the chain is not strongly aperiodic.

Lemma 10 *For any $\epsilon \in (0,1)$, a measure π is invariant with respect to the resolvent K_ϵ if and only if it is invariant with respect to P (the one step transition probability kernel for the original chain).*

Proof: We showed earlier that if π is invariant for $\mathbf{\Phi}$, then π is invariant for the K_ϵ-chain. Now, suppose π is invariant for the K_ϵ-chain.

$$\begin{aligned}
\int_{\mathcal{X}} \pi(dx)P(x,A) &= \int_{\mathcal{X}} \left[\int_{\mathcal{X}} \pi(dz)K_\epsilon(z,dx)\right] P(x,A) \\
&= \int_{\mathcal{X}} \left[\int_{\mathcal{X}} \pi(dz)(1-\epsilon)\sum_{k=0}^{\infty}\epsilon^k P^k(z,dx)\right] P(x,A) \\
&= (1-\epsilon)\int_{\mathcal{X}} \pi(dz)\sum_{k=0}^{\infty}\epsilon^k \int_{\mathcal{X}} P^k(z,dx)P(x,A) \\
&= (1-\epsilon)\epsilon^{-1}\int_{\mathcal{X}} \pi(dz)\sum_{k=0}^{\infty}\epsilon^{k+1}P^{k+1}(z,A) \\
&= (1-\epsilon)\epsilon^{-1}\int_{\mathcal{X}} \pi(dz)\left[\sum_{k=0}^{\infty}\epsilon^k P^k(z,A) - P^0(z,A)\right] \\
&= \epsilon^{-1}\int_X \pi(dz)K_\epsilon(z,A) - (1-\epsilon)\epsilon^{-1}\pi(A) \\
&= \epsilon^{-1}\pi(A) - (1-\epsilon)\epsilon^{-1}\pi(A) = \pi(A).
\end{aligned}$$

□

Using this lemma we can then prove

Proposition 22 *If $\mathbf{\Phi}$ is recurrent then $\mathbf{\Phi}$ has a unique (up to constant multiples) subinvariant measure that is invariant.*

Proof: We know that the K_ϵ chain is strongly aperiodic (Proposition 10). Since $\mathbf{\Phi}$ is recurrent, we know that the K_ϵ -chain is recurrent as well (Proposition 16). By Proposition 21, we can let π denote the unique invariant measure for the K_ϵ-chain. But then, by Lemma 10, π is invariant for $\mathbf{\Phi}$.

Now suppose that μ is subinvariant for $\mathbf{\Phi}$. But then μ is subinvariant for the K_ϵ-chain. Hence, there is a $c > 0$ such that $\mu = c\pi$ and so π is unique up to a constant. □

And if $\mathbf{\Phi}$ is aperiodic, we can show that it admits an invariant measure by using its skeletons.

Theorem 17 *Suppose that $\mathbf{\Phi}$ is ψ-irreducible and aperiodic. Then, for each m, a measure π is invariant for the m-skeleton if and only if it is invariant for the $\mathbf{\Phi}$.*

Hence, under aperiodicity, the chain $\mathbf{\Phi}$ is positive if and only if each of the m-skeletons is positive.

Proof: Follows along the same sort of reasoning as for the K_ϵ-chain. First, If π is invariant for $\mathbf{\Phi}$, then it is invariant for $\mathbf{\Phi}^m$ (we showed this on page 44). Suppose π_m is invariant for $\mathbf{\Phi}^m$. By aperiodicity, the measure π_m is unique up to constant multiples for $\mathbf{\Phi}^m$. Write

$$\pi(A) = m^{-1} \sum_{k=0}^{m-1} \int_{\mathcal{X}} \pi_m(dx) P^k(x, A), \quad A \in \mathcal{B}(\mathcal{X}).$$

But then,

$$\begin{aligned}
\int_{\mathcal{X}} \pi(dx) P(x, A) &= \int_{\mathcal{X}} \left[m^{-1} \sum_{k=0}^{m-1} \int_{\mathcal{X}} \pi_m(dz) P^k(z, dx) \right] P(x, A) \\
&= m^{-1} \sum_{k=0}^{m-1} \int_{\mathcal{X}} \pi_m(dz) P^{k+1}(z, A) \\
&= m^{-1} \sum_{k=0}^{m} \int_{\mathcal{X}} \pi_m(dz) \left[P^k(z, A) - P^0(z, A) \right] \\
&= m^{-1} \left[\sum_{k=0}^{m} \int_{\mathcal{X}} \pi_m(dz) P^k(z, A) - \pi_m(A) \right] \\
&= m^{-1} \left[\sum_{k=0}^{m-1} \int_{\mathcal{X}} \pi_m(dz) P^k(z, A) + \int_{\mathcal{X}} \pi_m(dz) P^m(z, A) - \pi_m(A) \right] \\
&= m^{-1} \sum_{k=0}^{m-1} \int_{\mathcal{X}} \pi_m(dz) P^k(z, A) = \pi(A).
\end{aligned}$$

Hence, π is an invariant measure for $\mathbf{\Phi}$. But then it is also an invariant measure for $\mathbf{\Phi}^m$. By the uniqueness of π_m, there is a $c > 0$ such that $\pi = c\pi_m$. But π is invariant for $\mathbf{\Phi}^j$ for all j. Therefore

$$\begin{aligned}
\pi(A) &= m^{-1} \sum_{k=0}^{m-1} \int_{\mathcal{X}} \pi_m(dx) P^k(x, A) \\
&= c^{-1} m^{-1} \sum_{k=0}^{m-1} \int_{\mathcal{X}} \pi(dx) P^k(x, A) \\
&= c^{-1} m^{-1} \sum_{k=0}^{m-1} \pi(A) = c^{-1} \pi(A).
\end{aligned}$$

Thus $c\pi(A) = \pi(A) = c\pi_m(A)$ so that $\pi(A) = \pi_m(A)$. □.

Ergodicity

In the previous sections we developed stability structures on Markov chains to ensure the existence of invariant measures on $\mathcal{B}(\mathcal{X})$ defined by the relation

$$\pi(A) = \int_{\mathcal{X}} \pi(dx) P(x, A).$$

Now, given the existence of an invariant measure, we will explore the question of when do the n-step transition probabilities converge in a suitable manner to π. For MCMC theory we will be interested in positive chains—those chains that admit a finite invariant measure that may be renormalized to an invariant probability measure.

Modes of convergence

When speaking of ergodicity we are referring to convergence of the chain in terms of its transition probabilities. Convergence of the chain through its sample paths leads to strong laws and to central limit theorems. To develop ergodic convergence we need the

Definition 33 (Total variation norm) *If μ is a signed measure on $\mathcal{B}(\mathcal{X})$, then the* total variation norm $||\mu||_{TV}$ *is defined as*

$$||\mu||_{TV} = \sup_{f:|f|\leq 1} |\mu(f)| = \sup_{A \in \mathcal{B}(\mathcal{X})} \mu(A) - \inf_{A \in \mathcal{B}(\mathcal{X})} \mu(A).$$

We need to define what is meant by $\mu(f)$ for a measure μ and a function f.

$$\mu(f) := \int_{\mathcal{X}} \mu(dx) f(x)$$

The key limit of interest is

$$\lim_{n\to\infty} ||P^n(x,\cdot) - \pi||_{TV} = 2 \lim_{n\to\infty} \sup_{A\in\mathcal{B}(\mathcal{X})} |P^n(x,A) - \pi(A)| = 0. \tag{16}$$

1.6.4 Independence of initial and limiting distributions

A desirable feature of the limiting behavior of the transition probabilities is that it is independent of the initial starting point or initial distribution. We will find conditions under which this is the case. In fact, there are situations where if the chain starts with an initial distribution that is "to close to infinity" then it may never reach the expected stationary distribution. Fortunately for us, this is not the case for positive Harris chains.

1.6.5 Ergodic chains

The term *ergodic* will refer to chains where the limit in (16) holds as time approaches infinity, rather than as time approaches infinity through some subsequence.

The development follows the same lines as the development of recurrent behavior: show ergodicity holds for positive Harris chains that possess an accessible atom. Generalize to positive Harris chains that do not possess an accessible atom by using the Nummelin splitting criterion.

1.6.6 Chains with atoms

Recall that a chain $\mathbf{\Phi}$ is called positive Harris if it is ψ-irreducible, Harris recurrent and admits an invariant probability measure, π.

Definition 34 (Ergodic atoms) *If $\mathbf{\Phi}$ is positive Harris, an atom $\boldsymbol{\alpha} \in \mathcal{B}^+(\mathcal{X})$ is called* ergodic *if it satifies*

$$\lim_{n\to\infty} |P^n(\boldsymbol{\alpha},\boldsymbol{\alpha}) - \pi(\boldsymbol{\alpha})| = 0.$$

Note that in the positive Harris case, an atom can be ergodic only if the chain is aperiodic.

Proposition 23 *If* $\mathbf{\Phi}$ *is positive recurrent and aperiodic, then any atom* $\boldsymbol{\alpha} \in \mathcal{B}^+(\mathcal{X})$ *is ergodic.*

Proposition 24 *If* $\mathbf{\Phi}$ *is positive Harris and admits an accessible ergodic atom (and hence is aperiodic), then for any initial measure* λ*,*

$$\lim_{n\to\infty} \left|\left| \int_{\mathcal{X}} \lambda(dx) P^n(x, \cdot) - \pi \right|\right|_{TV} = 0.$$

Given these propositions, we have the following

Theorem 18 *If* $\mathbf{\Phi}$ *is positive Harris and strongly aperiodic, then for any initial measure* λ*,*

$$\lim_{n\to\infty} \left|\left| \int_{\mathcal{X}} \lambda(dx) P^n(x, \cdot) - \pi \right|\right|_{TV} = 0.$$

Outline of proof:

> Since $\mathbf{\Phi}$ is positive Harris and strongly aperiodic, we know the split chain $\check{\mathbf{\Phi}}$ is also positive Harris and strongly aperiodic. We also know that $\check{\mathbf{\Phi}}$ possesses an accessible atom in $\mathcal{B}^+(\mathcal{X})$ and hence this atom is ergodic by Proposition 23. Then by Proposition 24 we know that $\check{P}^n$ converges in total variation norm to $\check{\pi}$. We can then transfer this convergence to the convergence of P^n to π in total variation norm by using the fact that the original chain is the marginal chain of the split chain. and π is the marginal measure of $\check{\pi}$.

This leads us directly to

Theorem 19 *If* $\mathbf{\Phi}$ *is positive Harris and aperiodic, then for every initial measure* λ

$$\lim_{n\to\infty} \left|\left| \int_{\mathcal{X}} \lambda(dx) P^n(x, \cdot) - \pi \right|\right|_{TV} = 0.$$

Outline of proof:

> Since $\mathbf{\Phi}$ is positive Harris and aperiodic, we know that there exists an $m > 1$ such that the m-skeleton chain $\mathbf{\Phi}^m$ is strongly aperiodic and positive Harris. Therefore P^{mn} converges to π by the previous theorem. The results can then be shown to carry over to the original chain.

This last theorem is the main result that we are after. For a chain that admits an invariant probability measure, π, and is aperiodic and Harris recurrent, the n-step transition probability kernels converge in total variation norm to a stationary distribution which happens to be π, and this is true for any initial distribution. The ramification of this for MCMC theory is the following: given we have run the chain long enough that P^n will be approximately equal to π for all $n > N$.

In terms of the definition of the total variation norm

$$\lim_{n\to\infty} \left|\left| \int_{\mathcal{X}} \lambda(dx) P^n(x, \cdot) - \pi \right|\right|_{TV} = 0$$

is equivalent to

$$\lim_{n\to\infty} \sup_{|f|\leq 1} \left| \int_{\mathcal{X}} \lambda(dx) \int_{\mathcal{X}} P^n(x, dy) f(y) - \int_{\mathcal{X}} \pi(dy) f(y) \right| = 0.$$

which is equivalent to saying

$$\lim_{n\to\infty} \mathbb{E}_\lambda[f(\Phi_n)] = \int_{\mathcal{X}} \pi(dx) f(x) = \mathbb{E}_\pi[f(\Phi_0)]$$

for any initial measure λ and any bounded function f.

1.7 Sample paths and limit theorems

For any function g on $\mathcal{X}$, let

$$S_n(g) := \sum_{k=1}^{n} g(\Phi_k).$$

We are interested in a Law of Large Numbers (LLN) and a Central Limit Theorem (CLT).

Definition 35 (LLN) *The Law of Large Numbers holds for a function g if*

$$\lim_{n\to\infty} \frac{1}{n} S_n(g) = \pi(g) := \int_{\mathcal{X}} \pi(dx) g(x) = \mathbb{E}_\pi[g(\Phi_0)].$$

Definition 36 (CLT) *The Central Limit Theorem holds for a function g if there exists a constant $\gamma_g^2 \in (0, \infty)$ such that for each initial condition $x \in \mathcal{X}$*

$$\lim_{n\to\infty} P_x \left[(n\gamma_g^2)^{-1/2} S_n[g - \pi(g)] \leq t \right] = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

In other words, as $n \to \infty$,

$$(n\gamma_g^2)^{-1/2} S_n[g - \pi(g)] \xrightarrow{d} N(0, 1).$$

This will look more familiar if we write it all out:

$$(n\gamma_g^2)^{-1/2} S_n[g - \pi(g)] = (n\gamma_g^2)^{-1/2} \left(\sum_{k=1}^{n} g(\Phi_k) - n\mathbb{E}_\pi[g(\Phi_0)] \right) \xrightarrow{d} N(0,1)$$

$$\Longrightarrow$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g(\Phi_k) \stackrel{d}{=} N\left(\mathbb{E}_\pi[g(\Phi_0)], \gamma_g^2/n\right).$$

Theorem 20 *If $\mathbf{\Phi}$ is a positive Harris chain with invariant probability measure π, then the LLN holds for any g that is absolutely integrable with respect to π, i.e. such that $\pi(|g|) = \int |g| d\pi < \infty$.*

We will state a CLT for the case when the chain is also time reversible, although CLTs exist for more general chains as well.

1.7.1 Reversibility

Definition 37 *A Markov chain, that has reached steady state, with invariant measure π is called* reversible *if*

$$\Pr(\Phi_{n+1} \in A \mid \Phi_n = x) = \Pr(\Phi_n \in A \mid \Phi_{n+1} = x).$$

Definition 38 *A Markov chain with transition kernel P satisfies the* detailed balance condition *if there exists a measure π satisfying*

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx).$$

Theorem 21 *If $\mathbf{\Phi}$ is a Markov chain that satisfies detailed balance with a probability measure π, then*

(i) π is the invariant probability measure of the chain (i.e. the chain is positive) and

(ii) the chain is reversible.

Theorem 22 *If* $\mathbf{\Phi}$ *is* ψ*-irreducible, aperiodic, reversible and positive with invariant probability measure* π *and if*

$$0 < \gamma_g^2 := \mathbb{E}_\pi \left[(g(\Phi_0) - \pi(g))^2 \right] + 2 \sum_{k=1}^{\infty} \mathbb{E}_\pi \left[(g(\Phi_0) - \pi(g)) \, (g(\Phi_k) - \pi(g)) \right] < \infty,$$

then the CLT holds for the chain. Furthermore

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E}_\pi \left[S_n^2(g - \pi(g)) \right] = \gamma_g^2.$$

That is, γ_g^2 *coincides with the asymptotic variance.*