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.