6

Topology and Continuity

The structure of Markov chains is essentially probabilistic, as we have described it so far. In examining the stability properties of Markov chains, the context we shall most frequently use is also a probabilistic one: in Part II, stability properties such as recurrence or regularity will be defined as certain return to sets of positive ψ-measure, or as finite mean return times to petite sets, and so forth.

Yet for many chains, there is more structure than simply a σ-field and a probability kernel available, and the expectation is that any topological structure of the space will play a strong role in defining the behavior of the chain. In particular, we are used thinking of specific classes of sets in $\mathbb{R}^n$ as having intuitively reasonable properties.

When there is a topology, compact sets are thought of in some sense as manageable sets, having the same sort of properties as a finite set on a countable space; and so we could well expect "stable" chains to spend the bulk of their time in compact sets. Indeed, we would expect compact sets to have the sort of characteristics we have identified, and will identify, for small or petite sets.

Conversely, open sets are "non-negligible" in some sense, and if the chain is irreducible we might expect it at least to visit all open sets with positive probability. This indeed forms one alternative definition of "irreducibility".

In this, the first chapter in which we explicitly introduce topological considerations, we will have, as our two main motivations, the desire to link the concept of ψ-irreducibility with that of open set irreducibility and the desire to identify compact sets as petite.

The major achievement of the chapter lies in identifying a topological condition on the transition probabilities which achieves both of these goals, utilizing the sampled chain construction we have just considered in Section 5.5.1.

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

$$\liminf_{y \to x} h(y) \geq h(x), \qquad x \in \mathsf{X} :$$

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

We will use the following continuity properties of the transition kernel, couched in terms of lower semicontinuous functions, to define classes of chains with suitable topological properties.

Feller chains, continuous components and T-chains

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

(ii) If a is a sampling distribution and there exists a substochastic transition kernel T satisfying

$$K_a(x, A) \geq T(x, A), \qquad x \in \mathsf{X},\ A \in \mathcal{B}(\mathsf{X}),$$

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

(iii) If $\boldsymbol{\Phi}$ is a Markov chain for which there exists a sampling distribution a such that K_a possesses a continuous component T, with $T(x, \mathsf{X}) > 0$ for all x, then $\boldsymbol{\Phi}$ is called a *T-chain.*

We will prove as one highlight of this section

Theorem 6.0.1 **(i)** *If $\boldsymbol{\Phi}$ is a T-chain and $L(x, O) > 0$ for all x and all open sets $O \in \mathcal{B}(\mathsf{X})$ then $\boldsymbol{\Phi}$ is ψ-irreducible.*

(ii) *If every compact set is petite then $\boldsymbol{\Phi}$ is a T-chain; and conversely, if $\boldsymbol{\Phi}$ is a ψ-irreducible T-chain then every compact set is petite.*

(iii) *If $\boldsymbol{\Phi}$ is a ψ-irreducible Feller chain such that* $\operatorname{supp}\psi$ *has non-empty interior, then $\boldsymbol{\Phi}$ is a ψ-irreducible T-chain.*

Proof Proposition 6.2.2 proves (i); (ii) is in Theorem 6.2.5; (iii) is in Theorem 6.2.9. □

In order to have any such links as those in Theorem 6.0.1 between the measure-theoretic and topological properties of a chain, it is vital that there be at least a minimal adaptation of the dynamics of the chain to the topology of the space on which it lives.

For consider the chain on $[0, 1]$ with transition law for $x \in [0, 1]$ given by

$$P(n^{-1}, (n+1)^{-1}) = 1 - \alpha_n, \qquad P(n^{-1}, 0) = \alpha_n,\ n \in \mathbb{Z}_+; \tag{6.1}$$

$$P(x, 1) = 1, \qquad x \neq n^{-1}, \qquad n \geq 1. \tag{6.2}$$

This chain fails to visit most open sets, although it is definitely irreducible provided $\alpha_n > 0$ for all n: and although it never leaves a compact set, it is clearly unstable in an obvious way if $\sum_n \alpha_n < \infty$, since then it moves monotonically down the sequence $\{n^{-1}\}$ with positive probability.

Of course, the dynamics of this chain are quite wrong for the space on which we have embedded it: its structure is adapted to the normal topology on the integers, not to that on the unit interval or the set $\{n^{-1}, n \in \mathbb{Z}_+\}$. The Feller property obviously fails at $\{0\}$, as does any continuous component property if $\alpha_n \to 0$.

This is a trivial and pathological example, but one which proves valuable in exhibiting the need for the various conditions we now consider, which do link the dynamics to the structure of the space.

6.1 Feller Properties and Forms of Stability

6.1.1 Weak and strong Feller chains

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

$$Ph(x) = \int P(x, dy)h(y), \qquad x \in \mathsf{X}. \tag{6.3}$$

Suppose that X is a (locally compact separable metric) topological space, and let us denote the class of bounded continuous functions from X to $\mathbb{R}$ by $\mathcal{C}(\mathsf{X})$.

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

That this is consistent with the definition above follows from

Proposition 6.1.1 **(i)** *The transition kernel $P\mathbb{1}_O$ is lower semicontinuous for every open set $O \in \mathcal{B}(\mathsf{X})$ (that is, $\boldsymbol{\Phi}$ is weak Feller) if and only if P maps $\mathcal{C}(\mathsf{X})$ to $\mathcal{C}(\mathsf{X})$; and P maps all bounded measurable functions to $\mathcal{C}(\mathsf{X})$ (that is, $\boldsymbol{\Phi}$ is strong Feller) if and only if the function $P\mathbb{1}_A$ is lower semicontinuous for every set $A \in \mathcal{B}(\mathsf{X})$.*

(ii) *If the chain is weak Feller then for any closed set $C \subset \mathsf{X}$ and any non-decreasing function $m: \mathbb{Z}_+ \to \mathbb{Z}_+$ the function $\mathsf{E}_x[m(\tau_C)]$ is lower semicontinuous in x. Hence for any closed set $C \subset \mathsf{X}$, $r > 1$ and $n \in \mathbb{Z}_+$ the functions*

$$\mathsf{P}_x\{\tau_C \geq n\} \qquad \mathsf{E}_x[\tau_C] \qquad \text{and} \quad \mathsf{E}_x[r^{\tau_C}]$$

are lower semicontinuous.

(iii) *If the chain is weak Feller then for any open set $O \subset \mathsf{X}$, the function $\mathsf{P}_x\{\tau_O \leq n\}$ and hence also the functions $K_a(x, O)$ and $L(x, O)$ are lower semicontinuous.*

PROOF To prove (i), suppose that $\boldsymbol{\Phi}$ is Feller, so that $P\mathbb{1}_O$ is lower semicontinuous for any open set O. Choose $f \in \mathcal{C}(\mathsf{X})$, and assume initially that $0 \leq f(x) \leq 1$ for all x. For $N \geq 1$ define the Nth approximation to f as

$$f_N(x) := \frac{1}{N} \sum_{k=1}^{N-1} \mathbb{1}_{O_k}(x)$$

where $O_k = \{x : f(x) > k/N\}$. It is easy to see that $f_N \uparrow f$ as $N \uparrow \infty$, and by assumption Pf_N is lower semicontinuous for each N. By monotone convergence, $Pf_N \uparrow Pf$ as $N \uparrow \infty$, and hence by Theorem D.4.1 the function Pf is lower semicontinuous. Identical reasoning shows that the function $P(1 - f) = 1 - Pf$, and hence also $-Pf$, is lower semicontinuous. Applying Theorem D.4.1 once more we see that the function Pf is continuous whenever f is continuous with $0 \leq f \leq 1$.

By scaling and translation it follows that Pf is continuous whenever f is bounded and continuous.

Conversely, if P maps $\mathcal{C}(\mathsf{X})$ to itself, and O is an open set then by Theorem D.4.1 there exist continuous positive functions f_N such that $f_N(x) \uparrow \mathbb{1}_O(x)$ for each x as $N \uparrow \infty$. By monotone convergence $P\mathbb{1}_O = \lim P f_N$, which by Theorem D.4.1 implies that $P\mathbb{1}_O$ is lower semicontinuous.

A similar argument shows that P is strong Feller if and only if the function $P\mathbb{1}_A$ is lower semicontinuous for every set $A \in \mathcal{B}(\mathsf{X})$.

We next prove (ii). By definition of τ_C we have $\mathsf{P}_x\{\tau_C = 0\} = 0$, and hence without loss of generality we may assume that $m(0) = 0$. For each $i \geq 1$ define $\Delta_m(i) := m(i) - m(i-1)$, which is non-negative since m is non-increasing. By a change of summation,

$$\begin{aligned} \mathsf{E}[m(\tau_C)] &= \sum_{k=1}^{\infty} m(k)\mathsf{P}_x\{\tau_C = k\} \\ &= \sum_{k=1}^{\infty}\sum_{i=1}^{k} \Delta_m(i)\mathsf{P}_x\{\tau_C = k\} \\ &= \sum_{i=1}^{\infty} \Delta_m(i)\mathsf{P}_x\{\tau_C \geq i\} \end{aligned}$$

Since by assumption $\Delta_m(k) \geq 0$ for each $k > 0$, the proof of (ii) will be complete once we have shown that $\mathsf{P}_x\{\tau_C \geq k\}$ is lower semicontinuous in x for all k.

Since C is closed and hence $\mathbb{1}_{C^c}(x)$ is lower semicontinuous, by Theorem D.4.1 there exist positive continuous functions f_i, $i \geq 1$, such that $f_i(x) \uparrow \mathbb{1}_{C^c}(x)$ for each $x \in \mathsf{X}$.

Extend the definition of the kernel I_A, given by

$$I_A(x, B) = \mathbb{1}_{A\cap B}(x),$$

by writing for any positive function g

$$I_g(x, B) := g(x)\mathbb{1}_B(x).$$

Then for all $k \in \mathbb{Z}_+$,

$$\mathsf{P}_x\{\tau_C \geq k\} = (PI_{C^c})^{k-1}(x, \mathsf{X}) = \lim_{i\to\infty} (PI_{f_i})^{k-1}(x, \mathsf{X}).$$

It follows from the Feller property that $\{(PI_{f_i})^{k-1}(x, \mathsf{X}) : i \geq 1\}$ is an increasing sequence of continuous functions and, again by Theorem D.4.1, this shows that $\mathsf{P}_x\{\tau_C \geq k\}$ is lower semicontinuous in x, completing the proof of (ii).

Result (iii) is similar, and we omit the proof.

□

Many chains satisfy these continuity properties, and we next give some important examples.

Weak Feller chains: the nonlinear state space models One of the simplest examples of a weak Feller chain is the quite general nonlinear state space model NSS(F).

Suppose conditions (NSS1) and (NSS2) are satisfied, so that $\mathbf{X} = \{X_n\}$, where

$$X_k = F(X_{k-1}, W_k),$$

for some smooth (C^∞) function $F: \mathsf{X} \times \mathbb{R}^p \to \mathsf{X}$, where X is an open subset of $\mathbb{R}^n$; and the random variables $\{W_k\}$ are a disturbance sequence on $\mathbb{R}^p$.

Proposition 6.1.2 *The NSS(F) model is always weak Feller.*

PROOF We have by definition that the mapping $x \to F(x, w)$ is continuous for each fixed $w \in \mathbb{R}$. Thus whenever $h: \mathsf{X} \to \mathbb{R}$ is bounded and continuous, $h \circ F(x, w)$ is also bounded and continuous for each fixed $w \in \mathbb{R}$. It follows from the Dominated Convergence Theorem that

$$\begin{aligned} Ph\,(x) &= \mathsf{E}[h(F\,(x, W))] \\ &= \int \Gamma(dw) h \circ F(x, w) \end{aligned} \tag{6.4}$$

is a continuous function of $x \in \mathsf{X}$. □

This simple proof of weak continuity can be emulated for many models. It implies that this aspect of the topological analysis of many models is almost independent of the random nature of the inputs. Indeed, we could rephrase Proposition 6.1.2 as saying that since the associated control model CM(F) is a continuous function of the state for each fixed control sequence, the stochastic nonlinear state space model NSS(F) is weak Feller.

We shall see in Chapter 7 that this reflection of deterministic properties of CM(F) by NSS(F) is, under appropriate conditions, a powerful and exploitable feature of the nonlinear state space model structure.

Weak and strong Feller chains: the random walk The difference between the weak and strong Feller properties is graphically illustrated in

Proposition 6.1.3 *The unrestricted random walk is always weak Feller, and is strong Feller if and only if the increment distribution Γ is absolutely continuous with respect to Lebesgue measure μ^{Leb} on $\mathbb{R}$.*

PROOF Suppose that $h \in \mathcal{C}(\mathsf{X})$: the structure (3.35) of the transition kernel for the random walk shows that

$$\begin{aligned} Ph\,(x) &= \int_{\mathbb{R}} h(y)\Gamma(dy - x) \\ &= \int_{\mathbb{R}} h(y + x)\Gamma(dy) \end{aligned} \tag{6.5}$$

and since h is bounded and continuous, Ph is also bounded and continuous, again from the Dominated Convergence Theorem. Hence $\boldsymbol{\Phi}$ is always weak Feller, as we also know from Proposition 6.1.2.

Suppose next that Γ possesses a density γ with respect to μ^{Leb} on $\mathbb{R}$. Taking h in (6.5) to be any bounded function, we have

$$Ph\,(x) = \int_{\mathbb{R}} h(y)\gamma(y - x)\, dy; \tag{6.6}$$

but now from Lemma D.4.3 it follows that the convolution $Ph\,(x) = \gamma * h$ is continuous, and the chain is strong Feller.

Conversely, suppose the random walk is strong Feller. Then for any B such that $\Gamma(B) = \delta > 0$, by the lower semicontinuity of $P(x, B)$ there exists a neighborhood O of $\{0\}$ such that

$$P(x, B) \geq P(0, B)/2 = \Gamma(B)/2 = \delta/2, \qquad x \in O. \tag{6.7}$$

By Fubini's Theorem and the translation invariance of μ^{Leb} we have for any $A \in \mathcal{B}(\mathsf{X})$

$$\begin{aligned} \int_{\mathbb{R}} \mu^{\text{Leb}}(dy)\Gamma(A-y) &= \int_{\mathbb{R}} \mu^{\text{Leb}}(dy) \int_{\mathbb{R}} \mathbb{1}_{A-y}(x)\Gamma(dx) \\ &= \int_{\mathbb{R}} \Gamma(dx) \int_{\mathbb{R}} \mathbb{1}_{A-x}(y)\mu^{\text{Leb}}(dy) \\ &= \mu^{\text{Leb}}(A) \end{aligned} \tag{6.8}$$

since $\Gamma(\mathbb{R}) = 1$. Thus we have in particular from (6.7) and (6.8)

$$\begin{aligned} \mu^{\text{Leb}}(B) &= \int_{\mathbb{R}} \mu^{\text{Leb}}(dy)\Gamma(B-y) \\ &\geq \int_{O} \mu^{\text{Leb}}(dy)\Gamma(B-y) \\ &\geq \delta\mu^{\text{Leb}}(O)/2 \end{aligned} \tag{6.9}$$

and hence $\mu^{\text{Leb}} \succ \Gamma$ as required. □

6.1.2 Strong Feller chains and open set irreducibility

Our first interest in chains on a topological space lies in identifying their accessible sets.

> Open set irreducibility
>
> **(i)** A point $x \in \mathsf{X}$ is called *reachable* if for every open set $O \in \mathcal{B}(\mathsf{X})$ containing x (i.e. for every neighborhood of x)
>
> $$\sum_n P^n(y, O) > 0, \qquad y \in \mathsf{X}.$$
>
> **(ii)** The chain $\boldsymbol{\Phi}$ is called *open set irreducible* if every point is reachable.

We will use often the following result, which is a simple consequence of the definition of support.

Lemma 6.1.4 *If $\boldsymbol{\Phi}$ is ψ-irreducible then x^* is reachable if and only if $x^* \in \operatorname{supp}(\psi)$.*

Proof If $x^* \in \text{supp}(\psi)$ then, for any open set O containing x^*, we have $\psi(O) > 0$ by the definition of the support. By ψ-irreducibility it follows that $L(x, O) > 0$ for all x, and hence x^* is reachable.

Conversely, suppose that $x^* \notin \text{supp}(\psi)$, and let $O = \text{supp}(\psi)^c$. The set O is open by the definition of the support, and contains the state x^*. By Proposition 4.2.3 there exists an absorbing, full set $A \subseteq \text{supp}(\psi)$. Since $L(x, O) = 0$ for $x \in A$ it follows that x^* is not reachable. □

It is easily checked that open set irreducibility is equivalent to irreducibility when the state space of the chain is countable and is equipped with the discrete topology.

The open set irreducibility definition is conceptually similar to the ψ-irreducibility definition: they both imply that "large" sets can be reached from every point in the space. In the ψ-irreducible case large sets are those of positive ψ-measure, whilst in the open set irreducible case, large sets are open non-empty sets.

In this book our focus is on the property of ψ-irreducibility as a fundamental structural property. The next result, despite its simplicity, begins to link that property to the properties of open-set irreducible chains.

Proposition 6.1.5 *If $\boldsymbol{\Phi}$ is a strong Feller chain, and X contains one reachable point x^*, then $\boldsymbol{\Phi}$ is ψ-irreducible, with $\psi = P(x^*, \cdot\,)$.*

Proof Suppose A is such that $P(x^*, A) > 0$. By lower semicontinuity of $P(\cdot\,, A)$, there is a neighborhood O of x^* such that $P(z, A) > 0, z \in O$. Now, since x^* is reachable, for any $y \in \mathsf{X}$, we have for some n

$$P^{n+1}(y, A) \geq \int_O P^n(y, dz) P(z, A) > 0 \tag{6.10}$$

which is the result. □

This gives trivially

Proposition 6.1.6 *If $\boldsymbol{\Phi}$ is an open set irreducible strong Feller chain, then $\boldsymbol{\Phi}$ is a ψ-irreducible chain.* □

We will see below in Proposition 6.2.2 that this strong Feller condition, which (as is clear from Proposition 6.1.3) may be unsatisfied for many models, is not needed in full to get this result, and that Proposition 6.1.5 and Proposition 6.1.6 hold for T-chains also.

There are now two different approaches we can take in connecting the topological and continuity properties of Feller chains with the stochastic or measure-theoretic properties of the chain. We can either weaken the strong Feller property by requiring in essence that it only hold partially; or we could strengthen the weak Feller condition whilst retaining its essential flavor.

It will become apparent that the former, T-chain, route is usually far more productive, and we move on to this next. A strengthening of the Feller property to give *e-chains* will then be developed in Section 6.4.

6.2 T-chains

6.2.1 T-chains and open set irreducibility

The calculations for NSS(F) models and random walks show that the majority of the chains we have considered to date have the weak Feller property.

However, we clearly need more than just the weak Feller property to connect measure-theoretic and topological irreducibility concepts: every random walk is weak Feller, and we know from Section 4.3.3 that any chain with increment measure concentrated on the rationals enters every open set but is not ψ-irreducible.

Moving from the weak to the strong Feller property is however excessive. Using the ideas of sampled chains introduced in Section 5.5.1 we now develop properties of the class of T-chains, which we shall find includes virtually all models we will investigate, and which appears almost ideally suited to link the general space attributes of the chain with the topological structure of the space.

The T-chain definition describes a class of chains which are not totally adapted to the topology of the space, in that the strongly continuous kernel T, being only a "component" of P, may ignore many discontinuous aspects of the motion of $\boldsymbol{\Phi}$: but it does ensure that the chain is not completely singular in its motion, with respect to the normal topology on the space, and the strong continuity of T links set-properties such as ψ-irreducibility to the topology in a way that is not natural for weak continuity.

We illustrate precisely this point now, with the analogue of Proposition 6.1.5.

Proposition 6.2.1 *If $\boldsymbol{\Phi}$ is a T-chain, and X contains one reachable point x^*, then $\boldsymbol{\Phi}$ is ψ-irreducible, with $\psi = T(x^*, \cdot\,)$.*

PROOF Let T be a continuous component for K_a: since T is everywhere non-trivial, we must have in particular that $T(x^*, \mathsf{X}) > 0$. Suppose A is such that $T(x^*, A) > 0$. By lower semicontinuity of $T(\,\cdot\,, A)$, there is a neighborhood O of x^* such that $T(w, A) > 0, w \in O$. Now, since x^* is reachable, for any $y \in \mathsf{X}$, we have from Proposition 5.5.2

$$\begin{aligned} K_{a_\varepsilon * a}(y, A) &\geq \int_O K_{a_\varepsilon}(y, dw) K_a(w, A) \\ &\geq \int_O K_{a_\varepsilon}(y, dw) T(w, A) \\ &> 0 \end{aligned} \tag{6.11}$$

which is the result. □

This result has, as a direct but important corollary

Proposition 6.2.2 *If $\boldsymbol{\Phi}$ is an open set irreducible T-chain, then $\boldsymbol{\Phi}$ is a ψ-irreducible T-chain.* □

6.2.2 T-chains and petite sets

When the Markov chain $\boldsymbol{\Phi}$ is ψ-irreducible, we know that there always exists at least one petite set. When X is topological, it turns out that there is a perhaps surprisingly direct connection between the existence of petite sets and the existence of continuous components.

In the next two results we show that the existence of sufficient open petite sets implies that Φ is a T-chain.

Proposition 6.2.3 *If an open ν_a-petite set A exists, then K_a possesses a continuous component non-trivial on all of A.*

PROOF Since A is ν_a-petite, by definition we have

$$K_a(\,\cdot\,,\,\cdot\,) \geq \mathbb{1}_A(\,\cdot\,)\nu\{\,\cdot\,\}.$$

Now set $T(x,B) := \mathbb{1}_A(x)\nu(B)$: this is certainly a component of K_a, non-trivial on A. Since A is an open set its indicator function is lower semicontinuous; hence T is a continuous component of K_a. □

Using such a construction we can build up a component which is non-trivial everywhere, if the space X is sufficiently rich in petite sets. We need first

Proposition 6.2.4 *Suppose that for each $x \in \mathsf{X}$ there exists a probability distribution a_x on $\mathbb{Z}_+$ such that K_{a_x} possesses a continuous component T_x which is non-trivial at x. Then Φ is a T-chain.*

PROOF For each $x \in \mathsf{X}$, let O_x denote the set

$$O_x = \{y \in \mathsf{X} : T_x(y, \mathsf{X}) > 0\}.$$

which is open since $T_x(\,\cdot\,, \mathsf{X})$ is lower semicontinuous. Observe that by assumption, $x \in O_x$ for each $x \in \mathsf{X}$.

By Lindelöf's Theorem D.3.1 there exists a countable subcollection of sets $\{O_i : i \in \mathbb{Z}_+\}$ and corresponding kernels T_i and K_{a_i} such that $\bigcup O_i = \mathsf{X}$. Letting

$$T = \sum_{k=1}^{\infty} 2^{-k} T_k \qquad \text{and} \qquad a = \sum_{k=1}^{\infty} 2^{-k} a_k,$$

it follows that $K_a \geq T$, and hence satisfies the conclusions of the proposition. □

We now get a virtual equivalence between the T-chain property and the existence of compact petite sets.

Theorem 6.2.5 **(i)** *If every compact set is petite, then Φ is a T-chain.*

(ii) *Conversely, if Φ is a ψ-irreducible T-chain then every compact set is petite, and consequently if Φ is an open set irreducible T-chain then every compact set is petite.*

PROOF Since X is σ-compact, there is a countable covering of open petite sets, and the result (i) follows from Proposition 6.2.3 and Proposition 6.2.4.

Now suppose that Φ is ψ-irreducible, so that there exists some petite $A \in \mathcal{B}^+(\mathsf{X})$, and let K_a have an everywhere non-trivial continuous component T.

By irreducibility $K_{a_\varepsilon}(x, A) > 0$, and hence from (5.46)

$$K_{a * a_\varepsilon}(x, A) = K_a K_{a_\varepsilon}(x, A) \geq T K_{a_\varepsilon}(x, A) > 0$$

for all $x \in \mathsf{X}$.

The function $TK_{a_\varepsilon}(\,\cdot\,, A)$ is lower semicontinuous and positive everywhere on X. Hence $K_{a*a_\varepsilon}(x, A)$ is uniformly bounded from below on compact subsets of X. Proposition 5.2.4 completes the proof that each compact set is petite.

The fact that we can weaken the irreducibility condition to open-set irreducibility follows from Proposition 6.2.2. □

The following factorization, which generalizes Proposition 5.5.5, further links the continuity and petiteness properties of T-chains.

Proposition 6.2.6 *If $\boldsymbol{\Phi}$ is a ψ-irreducible T-chain, then there is a sampling distribution b, an everywhere strictly positive, continuous function $s'\colon \mathsf{X} \to \mathbb{R}$, and a maximal irreducibility measure ψ_b such that*

$$K_b(x, B) \geq s'(x)\psi_b(B), \qquad x \in \mathsf{X},\ B \in \mathcal{B}(\mathsf{X}).$$

PROOF If T is a continuous component of K_a, then we have from Proposition 5.5.5 (iii),

$$\begin{aligned} K_{a*c}(x, B) &\geq \int K_a(x, dy)s(y)\,\psi_c(B) \\ &\geq T(x, s)\psi_c(B) \end{aligned}$$

The function $T(\,\cdot\,, s)$ is positive everywhere and lower semicontinuous, and therefore it dominates an everywhere positive continuous function s'; and we can take $b = a * c$ to get the required properties. □

6.2.3 Feller chains, petite sets, and T-chains

We now investigate the existence of compact petite sets when the chain satisfies only the (weak) Feller continuity condition. Ultimately this leads to an auxiliary condition, satisfied by very many models in practice, under which a weak Feller chain is also a T-chain.

We first require the following lemma for petite sets for Feller chains.

Lemma 6.2.7 *If $\boldsymbol{\Phi}$ is a ψ-irreducible Feller chain, then the closure of every petite set is petite.*

PROOF By Proposition 5.2.4 and Proposition 5.5.4 and regularity of probability measures on $\mathcal{B}(\mathsf{X})$ (i.e. a set $A \in \mathcal{B}(\mathsf{X})$ may be approximated from within by compact sets), the set A is petite if and only if there exists a probability a on $\mathbb{Z}_+$, $\delta > 0$, and a compact petite set $C \subset \mathsf{X}$ such that

$$K_a(x, C) \geq \delta, \qquad x \in A.$$

By Proposition 6.1.1 the function $K_a(x, C)$ is upper semicontinuous when C is compact. Thus we have

$$\inf_{x \in \bar{A}} K_a(x, C) = \inf_{x \in A} K_a(x, C)$$

and this shows that the closure of a petite set is petite. □

It is now possible to define auxiliary conditions under which all compact sets are petite for a Feller chain.

Proposition 6.2.8 *Suppose that Φ is ψ-irreducible. Then all compact subsets of* X *are petite if either:*

(i) *Φ has the Feller property and an open ψ-positive petite set exists; or*

(ii) *Φ has the Feller property and* $\operatorname{supp}\psi$ *has non-empty interior.*

PROOF To see (i), let A be an open petite set of positive ψ-measure. Then $K_{a_\varepsilon}(\,\cdot\,, A)$ is lower semicontinuous and positive everywhere, and hence bounded from below on compact sets. Proposition 5.5.4 again completes the proof.

To see (ii), let A be a ψ-positive petite set, and define

$$A_k := \text{ closure } \{x : K_{a_\varepsilon}(x, A) \geq 1/k\} \cap \operatorname{supp}\psi.$$

By Proposition 5.2.4 and Lemma 6.2.7, each A_k is petite. Since $\operatorname{supp}\psi$ has non-empty interior it is of the second category, and hence there exists $k \in \mathbb{Z}_+$ and an open set $O \subset A_k \subset \operatorname{supp}\psi$. The set O is an open ψ-positive petite set, and hence we may apply (i) to conclude (ii). □

A surprising, and particularly useful, conclusion from this cycle of results concerning petite sets and continuity properties of the transition probabilities is the following result, showing that Feller chains are in many circumstances also T-chains. We have as a corollary of Proposition 6.2.8 (ii) and Proposition 6.2.5 (ii) that

Theorem 6.2.9 *If a ψ-irreducible chain Φ is weak Feller and if* $\operatorname{supp}\psi$ *has nonempty interior then Φ is a T-chain.* □

These results indicate that the Feller property, which is a relatively simple condition to verify in many applications, provides some strong consequences for ψ-irreducible chains.

Since we may cover the state space of a ψ-irreducible Markov chain by a countable collection of petite sets, and since by Lemma 6.2.7 the closure of a petite set is itself petite, it might seem that Theorem 6.2.9 could be strengthened to provide an open covering of X by petite sets without additional hypotheses on the chain. It would then follow by Theorem 6.2.5 that any ψ-irreducible Feller chain is a T-chain.

Unfortunately, this is not the case, as is shown by the following counterexample. Let $\mathsf{X} = [0, 1]$ with the usual topology, let $0 < |\alpha| < 1$, and define the Markov transition function P for $x > 0$ by

$$P(x, \{0\}) = 1 - P(x, \{\alpha x\}) = x$$

We set $P(0, \{0\}) = 1$. The transition function P is Feller and δ_0-irreducible. But for any $n \in \mathbb{Z}_+$ we have

$$\lim_{x \to 0} \mathsf{P}_x(\tau_{\{0\}} \geq n) = 1,$$

from which it follows that there does not exist an open petite set containing the point $\{0\}$.

Thus we have constructed a ψ-irreducible Feller chain on a compact state space which is not a T-chain.

6.3 Continuous Components For Specific Models

For a very wide range of the irreducible examples we consider, the support of the irreducibility measure does indeed have non-empty interior under some "spread-out" type of assumption. Hence weak Feller chains, such as the entire class of nonlinear models, will have all of the properties of the seemingly much stronger T-chain models provided they have an appropriate irreducibility structure.

We now identify a number of other examples of T-chains more explicitly.

6.3.1 Random walks

Suppose $\boldsymbol{\Phi}$ is random walk on a half-line. We have already shown that provided the increment distribution Γ provides some probability of negative increments then the chain is δ_0-irreducible, and moreover all of the sets $[0, c]$ are small sets.

Thus all compact sets are small and we have immediately from Theorem 6.2.5

Proposition 6.3.1 *The random walk on a half line with increment measure Γ is always a ψ-irreducible T-chain provided that $\Gamma(-\infty, 0) > 0$.* □

Exactly the same argument for a storage model with general state-dependent release rule $r(x)$, as discussed in Section 2.4.4, shows these models to be δ_0-irreducible T-chains when the integral $R(x)$ of (2.33) is finite for all x.

Thus the virtual equivalence of the petite compact set condition and the T-chain condition provides an easy path to showing the existence of continuous components for many models with a real atom in the space.

Assessing conditions for non-atomic chains to be T-chains is not quite as simple in general. However, we can describe exactly what the continuous component condition defining T-chains means in the case of the random walk. Recall that the random walk is called spread-out if some convolution power Γ^{n*} is non-singular with respect to μ^{Leb} on $\mathbb{R}$.

Proposition 6.3.2 *The unrestricted random walk is a T-chain if and only if it is spread out.*

PROOF If Γ is spread out then for some M, and some positive function γ, we have

$$P^M(x, A) = \Gamma^{M*}(A - x) \geq \int_{A-x} \gamma(y)dy := T(x, A)$$

and exactly as in the proof of Proposition 6.1.3, it follows that T is strong Feller: the spread-out assumption ensures that $T(x, \mathsf{X}) > 0$ for all x, and so by choosing the sampling distribution as $a = \delta_M$ we find that $\boldsymbol{\Phi}$ is a T-chain.

The converse is somewhat harder, since we do not know *a priori* that when $\boldsymbol{\Phi}$ is a T-chain, the component T can be chosen to be translation invariant. So let us assume that the result is false, and choose A such that $\mu^{\mathrm{Leb}}(A) = 0$ but $\Gamma^{n*}(A) = 1$ for every n. Then $\Gamma^{n*}(A^c) = 0$ for all n and so for the sampling distribution a associated with the component T,

$$T(0, A^c) \leq K_a(0, A^c) = \sum_n \Gamma^{n*}(A^c)a(n) = 0.$$

The non-triviality of the component T thus ensures $T(0,A) > 0$, and since $T(x,A)$ is lower semicontinuous, there exists a neighborhood O of $\{0\}$ and a $\delta > 0$ such that $T(x,A) \geq \delta > 0$, $x \in O$.

Since T is a component of K_a, this ensures

$$K_a(x,A) \geq \delta > 0, \qquad x \in O.$$

But as in (6.8) by Fubini's Theorem and the translation invariance of μ^{Leb} we have

$$\begin{aligned}\mu^{\mathrm{Leb}}(A) &= \int_{\mathbb{R}} \mu^{\mathrm{Leb}}(dy)\Gamma^{n*}(A-y) \\ &= \int_{\mathbb{R}} \mu^{\mathrm{Leb}}(dy)P^n(y,A).\end{aligned} \tag{6.12}$$

Multiplying both sides of (6.12) by $a(n)$ and summing gives

$$\begin{aligned}\mu^{\mathrm{Leb}}(A) &= \int_{\mathbb{R}} \mu^{\mathrm{Leb}}(dy)K_a(y,A) \\ &\geq \int_O \mu^{\mathrm{Leb}}(dy)K_a(y,A) \\ &\geq \delta\mu^{\mathrm{Leb}}(O)\end{aligned} \tag{6.13}$$

and since $\mu^{\mathrm{Leb}}(O) > 0$, we have a contradiction. □

This example illustrates clearly the advantage of requiring only a continuous component, rather than the Feller property for the chain itself.

6.3.2 Linear models as T-chains

Proposition 6.3.2 implies that the random walk model is a T-chain whenever the distribution of the increment variable $\mathbf{W}$ is sufficiently rich that, from each starting point, the chain does not remain in a set of zero Lebesgue measure.

This property, that when the set of reachable states is appropriately large the model is a T-chain, carries over to a much larger class of processes, including the linear and nonlinear state space models.

Suppose that $\mathbf{X}$ is a LSS(F,G)model, defined as usual by $X_{k+1} = FX_k + GW_{k+1}$. By repeated substitution in (LSS1) we obtain for any $m \in \mathbb{Z}_+$,

$$X_m = F^m X_0 + \sum_{i=0}^{m-1} F^i G W_{m-i} \tag{6.14}$$

To obtain a continuous component for the LSS(F,G) model, our approach is similar to that in deriving its irreducibility properties in Section 4.4. We require that the set of possible reachable states be large for the associated deterministic linear control system, and we also require that the set of reachable states remain large when the control sequence $\mathbf{u}$ is replaced by the random disturbance $\mathbf{W}$. One condition sufficient to ensure this is

Non-singularity Condition for the LSS(F,G) Model

(LSS4) The distribution Γ of the random variable W is non-singular with respect to Lebesgue measure, with non-trivial density γ_w.

Using (6.14) we now show that the n-step transition kernel itself possesses a continuous component provided, firstly, Γ is non-singular with respect to Lebesgue measure and secondly, the chain $\mathbf{X}$ can be driven to a sufficiently large set of states in $\mathbb{R}^n$ through the action of the disturbance process $\mathbf{W} = \{W_k\}$ as described in the last term of (6.14). This second property is a consequence of the controllability of the associated model LCM(F,G).

In Chapter 7 we will show that this construction extends further to more complex nonlinear models.

Proposition 6.3.3 *Suppose the deterministic control model LCM(F,G) on $\mathbb{R}^n$ satisfies the controllability condition (LCM3), and the associated LSS(F,G) model $\mathbf{X}$ satisfies the nonsingularity condition(LSS4).*

Then the n-skeleton possesses a continuous component which is everywhere non-trivial, so that $\mathbf{X}$ is a T-chain.

Proof We will prove this result in the special case where W is a scalar. The general case with $W \in \mathbb{R}^p$ is proved using the same methods as in the case where $p = 1$, but much more notation is needed for the required change of variables [174].

Let f denote an arbitrary positive function on $\mathsf{X} = \mathbb{R}^n$. From (6.14) together with non-singularity of the disturbance process $\mathbf{W}$ we may bound the conditional mean of $f(\Phi_n)$ as follows:

$$\begin{aligned} P^n f\,(x_0) &= \mathsf{E}[f(F^n x_0 + \sum_{i=0}^{n-1} F^i G W_{n-i})] \qquad (6.15)\\ &\geq \int \cdots \int f(F^n x_0 + \sum_{i=0}^{n-1} F^i G w_{n-i})\, \gamma_w(w_1) \cdots \gamma_w(w_n)\, dw_1 \ldots dw_n. \end{aligned}$$

Letting C_n denote the controllability matrix in (4.13) and defining the vector valued random variable $\vec{W}_n = (W_1, \ldots, W_n)^\top$, we define the kernel T as

$$Tf\,(x) := \int f(F^n x + C_n \vec{w}_n)\, \gamma_{\vec{w}}(\vec{w}_n)\, d\vec{w}_n.$$

We have $T(x, \mathsf{X}) = \{\int \gamma_w(x)\, dx\}^n > 0$, which shows that T is everywhere non-trivial; and T is a component of P^n since (6.15) may be written in terms of T as

$$P^n f\,(x_0) \geq \int f(F^n x_0 + C_n \vec{w}_n)\, \gamma_{\vec{w}}(\vec{w}_n)\, d\vec{w}_n = Tf(x_0). \qquad (6.16)$$

Let $|C_n|$ denote the determinant of C_n, which is non-zero since the pair (F, G) is controllable. Making the change of variables

$$\vec{v}_n = C_n \vec{w}_n \qquad d\vec{v}_n = |C_n| d\vec{w}_n$$

in (6.16) allows us to write

$$Tf\,(x_0) = \int f(F^n x_0 + \vec{v}_n) \gamma_{\vec{w}}(C_n^{-1}\vec{v}_n)|C_n|^{-1}\, d\vec{v}_n.$$

By Lemma D.4.3 and the Dominated Convergence Theorem, the right hand side of this identity is a continuous function of x_0 whenever f is bounded. This combined with (6.16) shows that T is a continuous component of P^n. □

In particular this shows that the ARMA process (ARMA1) and any of its variations may be modeled as a T-chain if the noise process **W** is sufficiently rich with respect to Lebesgue measure, since they possess a controllable realization from Proposition 4.4.2.

In general, we can also obtain a T-chain by restricting the process to a controllable subspace of the state space in the manner indicated after Proposition 4.4.3.

6.3.3 Linear models as ψ-irreducible T-chains

We saw in Proposition 4.4.3 that a controllable LSS(F,G) model is ψ-irreducible (with ψ equivalent to Lebesgue measure) if the distribution Γ of W is Gaussian. In fact, under the conditions of that result, the process is also strong Feller, as we can see from the exact form of (4.18). Thus the controllable Gaussian model is a ψ-irreducible T-chain, with ψ specifically identified and the "component" T given by P itself.

In Proposition 6.3.3 we weakened the Gaussian assumption and still found conditions for the LSS(F,G) model to be a T-chain. We need extra conditions to retain ψ-irreducibility.

Now that we have developed the general theory further we can also use substantially weaker conditions on **W** to prove the chain possesses a reachable state, and this will give us the required result from Section 6.2.1. We introduce the following condition on the matrix F used in (LSS1):

Eigenvalue condition for the LSS(F,G) model

(LSS5) The eigenvalues of F fall within the open unit disk in $\mathbb{C}$.

We will use the following lemma to control the growth of the models below.

Lemma 6.3.4 *Let $\rho(F)$ denote the modulus of the eigenvalue of F of maximum modulus, where F is an $n \times n$ matrix. Then for any matrix norm $\|\cdot\|$ we have the limit*

$$\log(\rho(F)) = \lim_{n\to\infty} \frac{1}{n} \log(\|F^n\|). \tag{6.17}$$

PROOF The existence of the limit (6.17) follows from the Jordan Decomposition and is a standard result from linear systems theory: see [39] or Exercises 2.I.2 and 2.I.5 of [69] for details. □

A consequence of Lemma 6.3.4 is that for any constants $\underline{\rho}, \overline{\rho}$ satisfying $\underline{\rho} < \rho(F) < \overline{\rho}$, there exists $c > 1$ such that

$$c^{-1}\underline{\rho}^n \leq \|F^n\| \leq c\overline{\rho}^n. \tag{6.18}$$

Hence for the linear state space model, under the eigenvalue condition (LSS5), the convergence $F^n \to 0$ takes place at a geometric rate. This property is used in the following result to give conditions under which the linear state space model is irreducible.

Proposition 6.3.5 *Suppose that the LSS(F,G) model* $\mathbf{X}$ *satisfies the density condition (LSS4) and the eigenvalue condition (LSS5), and that the associated control system LCM(F,G) is controllable.*

Then $\mathbf{X}$ *is a ψ-irreducible T-chain and every compact subset of* X *is small.*

PROOF We have seen in Proposition 6.3.3 that the linear state space model is a T-chain under these conditions. To obtain irreducibility we will construct a reachable state and use Proposition 6.2.1.

Let $w^\star$ denote any element of the support of the distribution Γ of W, and let

$$x^\star = \sum_{k=0}^{\infty} F^k G w^\star.$$

If in (1.4), the control $u_k = w^\star$ for all k, then the system x_k converges to $x^\star$ uniformly for initial conditions in compact subsets of X.

By (pointwise) continuity of the model, it follows that for any bounded set $A \subset \mathsf{X}$ and open set O containing $x^\star$, there exists $\varepsilon > 0$ sufficiently small and $N \in \mathbb{Z}_+$ sufficiently large such that $x_N \in O$ whenever $x_0 \in A$, and $u_i \in w^\star + \varepsilon B$, for $1 \leq i \leq N$, where B denotes the open unit ball centered at the origin in X. Since $w^\star$ lies in the support of the distribution of W_k we can conclude that $P^N(x_0, O) \geq \Gamma(w^\star + \varepsilon B)^N > 0$ for $x_0 \in A$.

Hence $x^\star$ is reachable, which by Proposition 6.2.1 and Proposition 6.3.3 implies that $\boldsymbol{\Phi}$ is ψ-irreducible for some ψ.

We now show that all bounded sets are small, rather than merely petite. Proposition 6.3.3 shows that P^n possesses a strong Feller component T. By Theorem 5.2.2 there exists a small set C for which $T(x^\star, C) > 0$ and hence, by the Feller property, an open set O containing $x^\star$ exists for which

$$\inf_{x \in O} T(x, C) > 0.$$

By Proposition 5.2.4 O is also a small set. If A is a bounded set, then we have already shown that $A \overset{\delta_M}{\rightsquigarrow} O$ for some N, so applying Proposition 5.2.4 once more we have the desired conclusion that A is small. □

6.3.4 The first-order SETAR model

Results for nonlinear models are not always as easy to establish. However, for simple models similar conditions on the noise variables establish similar results. Here we consider the first-order SETAR models, which are defined as piecewise linear models satisfying

$$X_n = \phi(j) + \theta(j)X_{n-1} + W_n(j), \qquad X_{n-1} \in R_j$$

where $-\infty = r_0 < r_1 < \cdots < r_M = \infty$ and $R_j = (r_{j-1}, r_j]$; for each j, the noise variables $\{W_n(j)\}$ form an i.i.d. zero-mean sequence independent of $\{W_n(i)\}$ for $i \neq j$. Throughout, $W(j)$ denotes a generic variable with distribution Γ_j.

In order to ensure that these models can be analyzed as T-chains we make the following additional assumption, analogous to those above.

(SETAR2) For each $j = 1, \cdots, M$, the noise variable $W(j)$ has a density positive on the whole real line.

Even though this model is not Feller, due to the possible presence of discontinuities at the boundary points $\{r_i\}$, we can establish

Proposition 6.3.6 *Under (SETAR1) and (SETAR2), the SETAR model is a φ-irreducible T-process with φ taken as Lebesgue measure μ^{Leb} on $\mathbb{R}$.*

PROOF The μ^{Leb}-irreducibility is immediate from the assumption of positive densities for each of the $W(j)$. The existence of a continuous component is less simple.

It is obvious from the existence of the densities that at any point in the interior of any of the regions R_i the transition function is strongly continuous. We do not necessarily have this continuity at the boundaries r_i themselves. However, as $x \uparrow r_i$ we have strong continuity of $P(x, \cdot\,)$ to $P(r_i, \cdot\,)$, whilst the limits as $x \downarrow r_i$ of $P(x, A)$ always exist giving a limit measure $P'(r_i, \cdot\,)$ which may differ from $P(r_i, \cdot\,)$.

If we take $T_i(x, \cdot\,) = \min(P'(r_i, \cdot\,), P(r_i, \cdot\,), P(x, \cdot\,))$ then T_i is a continuous component of P at least in some neighborhood of r_i; and the assumption that the densities of both $W(i), W(i+1)$ are positive everywhere guarantees that T_i is non-trivial.

But now we may put these components together using Proposition 6.2.4 and we have shown that the SETAR model is a T-chain. □

Clearly one can weaken the positive density assumption. For example, it is enough for the T-chain result that for each j the supports of $W(j) - \phi(j) - \theta(j)r_j$ and $W(j+1) - \phi(j+1) - \theta(j+1)r_j$ should not be distinct, whilst for the irreducibility one can similarly require only that the densities of $W(j) - \phi(j) - \theta(j)x$ exist in a fixed neighborhood of zero, for $x \in (r_{j-1}, r_j]$. For chains which do not for some structural reason obey (SETAR2) one would need to check the conditions on the support of the noise variables with care to ensure that the conclusions of Proposition 6.3.6 hold.

6.4 e-Chains

Now that we have developed some of the structural properties of T-chains that we will require, we move on to a class of Feller chains which also have desirable structural properties, namely e-chains.

6.4.1 e-Chains and dynamical systems

The stability of weak Feller chains is naturally approached in the context of dynamical systems theory as introduced in the heuristic discussion in Chapter 1. Recall from Section 1.3.2 that the Markov transition function P gives rise to a deterministic map from $\mathcal{M}$, the space of probabilities on $\mathcal{B}(\mathsf{X})$, to itself, and we can construct on this basis a dynamical system $(P, \mathcal{M}, d)$, provided we specify a metric d, and hence also a topology, on $\mathcal{M}$.

To do this we now introduce the topology of weak convergence.

> Weak Convergence
>
> A sequence of probabilities $\{\mu_k : k \in \mathbb{Z}_+\} \subset \mathcal{M}$ converges *weakly* to $\mu_\infty \in \mathcal{M}$ (denoted $\mu_k \xrightarrow{\mathrm{w}} \mu_\infty$) if
>
> $$\lim_{k\to\infty} \int f\, d\mu_k = \int f\, d\mu_\infty$$
>
> for every $f \in \mathcal{C}(\mathsf{X})$.

Due to our restrictions on the state space X, the topology of weak convergence is induced by a number of metrics on $\mathcal{M}$; see Section D.5. One such metric may be expressed

$$d_m(\mu, \nu) = \sum_{k=0}^{\infty} |\int f_k\, d\mu - \int f_k\, d\nu| 2^{-k}, \qquad \mu, \nu \in \mathcal{M} \tag{6.19}$$

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

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

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

or equivalently, $\lim_{k\to\infty} Pf\,(x_k) = Pf\,(x_\infty)$ for all $f \in \mathcal{C}(\mathsf{X})$. That is, if the Markov transition function induces a continuous map on $\mathcal{M}$, then Pf must be continuous for any bounded continuous function f.

This is exactly the weak Feller property. Conversely, it is obvious that for any weak Feller Markov transition function P, the associated operator P on $\mathcal{M}$ is continuous. We have thus shown

Proposition 6.4.1 *The triple $(P, \mathcal{M}, d_m)$ is a dynamical system if and only if the Markov transition function P has the weak Feller property.* □

Although we do not get further immediate value from this result, since there do not exist a great number of results in the dynamical systems theory literature to be exploited in this context, these observations guide us to stronger and more useful continuity conditions.

Equicontinuity and e-Chains

The Markov transition function P is called *equicontinuous* if for each $f \in \mathcal{C}_c(\mathsf{X})$ the sequence of functions $\{P^k f : k \in \mathbb{Z}_+\}$ is equicontinuous on compact sets.

A Markov chain which possesses an equicontinuous Markov transition function will be called an *e-chain*.

There is one striking result which very largely justifies our focus on e-chains, especially in the context of more stable chains.

Proposition 6.4.2 *Suppose that the Markov chain $\boldsymbol{\Phi}$ has the Feller property, and that there exists a unique probability measure π such that for every x*

$$P^n(x, \cdot\,) \xrightarrow{\mathrm{w}} \pi. \tag{6.20}$$

Then $\boldsymbol{\Phi}$ is an e-chain.

Proof Since the limit in (6.20) is continuous (and in fact constant) it follows from Ascoli's Theorem D.4.2 that the sequence of functions $\{P^k f : k \in \mathbb{Z}_+\}$ is equicontinuous on compact subsets of X whenever $f \in \mathcal{C}(\mathsf{X})$. Thus the chain $\boldsymbol{\Phi}$ is an e-chain. □

Thus chains with good limiting behavior, such as those in Part III in particular, are forced to be e-chains, and in this sense the e-chain assumption is for many purposes a minor extra step after the original Feller property is assumed.

Recall from Chapter 1 that the dynamical system $(P, \mathcal{M}, d_m)$ is called stable in the sense of Lyapunov if for each measure $\mu \in \mathcal{M}$,

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

The following result creates a further link between classical dynamical systems theory, and the theory of Markov chains on topological state spaces. The proof is routine and we omit it.

Proposition 6.4.3 *The Markov chain is an e-chain if and only if the dynamical system $(P, \mathcal{M}, d_m)$ is stable in the sense of Lyapunov.*

6.4.2 e-Chains and tightness

Stability in the sense of Lyapunov is a useful concept when a stationary point for the dynamical system exists. If x^* is a stationary point and the dynamical system is stable in the sense of Lyapunov, then trajectories which start near x^* will stay near x^*, and this turns out to be a useful notion of stability.

For the dynamical system $(P, \mathcal{M}, d_m)$, a stationary point is an invariant probability: that is, a probability satisfying

$$\pi(A) = \int \pi(dx) P(x, A), \qquad A \in \mathcal{B}(\mathsf{X}). \tag{6.21}$$

Conditions for such an invariant measure π to exist are the subject of considerable study for ψ-irreducible chains in Chapter 10, and in Chapter 12 we return to this question for weak Feller chains and e-chains.

A more immediately useful concept is that of Lagrange stability. Recall from Section 1.3.2 that $(P, \mathcal{M}, d_m)$ is Lagrange stable if, for every $\mu \in \mathcal{M}$, the orbit of measures μP^k is a precompact subset of $\mathcal{M}$. One way to investigate Lagrange stability for weak Feller chains is to utilize the following concept, which will have much wider applicability in due course.

> Chains Bounded in Probability
>
> The Markov chain $\boldsymbol{\Phi}$ is called *bounded in probability* if for each initial condition $x \in \mathsf{X}$ and each $\varepsilon > 0$, there exists a compact subset $C \subset \mathsf{X}$ such that
>
> $$\liminf_{k \to \infty} \mathsf{P}_x\{\Phi_k \in C\} \geq 1 - \varepsilon.$$

Boundedness in probability is simply tightness for the collection of probabilities $\{P^k(x, \cdot\,) : k \geq 1\}$. Since it is well known [24] that a set of probabilities $\mathcal{A} \subset \mathcal{M}$ is tight if and only if $\mathcal{A}$ is precompact in the metric space $(\mathcal{M}, d_m)$ this proves

Proposition 6.4.4 *The chain $\boldsymbol{\Phi}$ is bounded in probability if and only if the dynamical system $(P, \mathcal{M}, d_m)$ is Lagrange stable.* □

For e-chains, the concepts of boundedness in probability and Lagrange stability also interact to give a useful stability result for a somewhat different dynamical system.

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

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

where $\{C_k\}$ is a sequence of open precompact sets whose union is equal to X. The associated metric d_c generates the topology of uniform convergence on compact subsets of X.

If P is a weak Feller kernel, then the mapping P on $\mathcal{C}(\mathsf{X})$ is continuous with respect to this norm, and in this case the triple $(P, \mathcal{C}(\mathsf{X}), d_c)$ is a dynamical system.

By Ascoli's Theorem D.4.2, $(P, \mathcal{C}(\mathsf{X}), d_c)$ will be Lagrange stable if and only if for each initial condition $f \in \mathcal{C}(\mathsf{X})$, the orbit $\{P^k f : k \in \mathbb{Z}_+\}$ is uniformly bounded, and equicontinuous on compact subsets of X. This fact easily implies

Proposition 6.4.5 *Suppose that $\boldsymbol{\Phi}$ is bounded in probability. Then $\boldsymbol{\Phi}$ is an e-chain if and only if the dynamical system $(P, \mathcal{C}(\mathsf{X}), d_c)$ is Lagrange stable.* □

To summarize, for weak Feller chains boundedness in probability and the equicontinuity assumption are, respectively, exactly the same as Lagrange stability and stability in the sense of Lyapunov for the dynamical system $(P, \mathcal{M}, d_m)$; and these stability conditions are both simultaneously satisfied if and only if the dynamical system $(P, \mathcal{M}, d_m)$ and its dual $(P, \mathcal{C}(\mathsf{X}), d_c)$ are simultaneously Lagrange stable.

These connections suggest that equicontinuity will be a useful tool for studying the limiting behavior of the distributions governing the Markov chain $\boldsymbol{\Phi}$, a belief which will be justified in the results in Chapter 12 and Chapter 18.

6.4.3 Examples of e-chains

The easiest example of an e-chain is the simple linear model described by (SLM1) and (SLM2).

If x and y are two initial conditions for this model, and the resulting sample paths are denoted $\{X_n(x)\}$ and $\{X_n(y)\}$ respectively for the same noise path, then by (SLM1) we have

$$X_{n+1}(x) - X_{n+1}(y) = \alpha(X_n(x) - X_n(y)) = \alpha^{n+1}(x - y). \tag{6.22}$$

If $|\alpha| \leq 1$, then this indicates that the sample paths should remain close together if their initial conditions are also close.

From this observation we now show that the simple linear model is an e-chain under the stability condition that $|\alpha| \leq 1$. Since the random walk on $\mathbb{R}$ is a special case of the simple linear model with $\alpha = 1$, this also implies that the random walk is also an e-chain.

Proposition 6.4.6 *The simple linear model defined by (SLM1) and (SLM2) is an e-chain provided that $|\alpha| \leq 1$.*

PROOF Let $f \in \mathcal{C}_c(\mathsf{X})$. By uniform continuity of f, for any $\varepsilon > 0$ we can find $\delta > 0$ so that $|f(x) - f(y)| \leq \varepsilon$ whenever $|x - y| \leq \delta$. It follows from (6.22) that for any $n \in \mathbb{Z}_+$, and any x, $y \in \mathbb{R}$ with $|x - y| \leq \delta$,

$$\begin{aligned} |P^{n+1} f(x) - P^{n+1} f(y)| &= |\mathsf{E}[f(X_{n+1}(x)) - f(X_{n+1}(y))]| \\ &\leq \mathsf{E}[|f(X_{n+1}(x)) - f(X_{n+1}(y))|] \\ &\leq \varepsilon, \end{aligned}$$

which shows that $\mathbf{X}$ is an e-chain. □

Equicontinuity is rather difficult to verify or rule out directly in general, especially before some form of stability has been established for the process. Although the equicontinuity condition may seem strong, it is surprisingly difficult to construct a natural example of a Feller chain which is not an e-chain. Indeed, our concentration on them is justified by Proposition 6.4.2 and this does provide an indirect way to verify that many Feller examples are indeed e-chains.

One example of a "non-e" chain is, however, provided by a "multiplicative random walk" on $\mathbb{R}_+$, defined by

$$X_{k+1} = \sqrt{X_k} W_{k+1}, \qquad k \in \mathbb{Z}_+, \tag{6.23}$$

where $\mathbf{W}$ is a disturbance sequence on $\mathbb{R}_+$ whose marginal distribution possesses a finite first moment. The chain is Feller since the right hand side of (6.23) is continuous in X_k. However, $\mathbf{X}$ is not an e-chain when $\mathbb{R}$ is equipped with the usual topology.

A complete proof of this fact requires more theory than we have so far developed, but we can give a sketch to illustrate what can go wrong.

When $X_0 \neq 0$, the process $\log X_k$, $k \in \mathbb{Z}_+$, is a version of the simple linear model described in Chapter 2, with $\alpha = \frac{1}{2}$. We will see in Section 10.5.4 that this implies that for any $X_0 = x_0 \neq 0$ and any bounded continuous function f,

$$P^k f\,(x_0) \to f_\infty, \qquad k \to \infty$$

where f_∞ is a constant. When $x_0 = 0$ we have that $P^k f\,(x_0) = f(x_0) = f(0)$ for all k.

From these observations it is easy to see that $\mathbf{X}$ is not an e-chain. Take $f \in \mathcal{C}_c(\mathsf{X})$ with $f(0) = 0$ and $f(x) \geq 0$ for all $x > 0$: we may assume without loss of generality that $f_\infty > 0$. Since the one-point set $\{0\}$ is absorbing we have $P^k(0, \{0\}) = 1$ for all k, and it immediately follows that $P^k f$ converges to a discontinuous function. By Ascoli's Theorem the sequence of functions $\{P^k f : k \in \mathbb{Z}_+\}$ cannot be equicontinuous on compact subsets of $\mathbb{R}_+$, which shows that $\mathbf{X}$ is not an e-chain.

However by modifying the topology on $\mathsf{X} = \mathbb{R}_+$ we do obtain an e-chain as follows. Define the topology on the strictly positive real line $(0, \infty)$ in the usual way, and define $\{0\}$ to be open, so that X becomes a disconnected set with two open components. Then, in this topology, $P^k f$ converges to a uniformly continuous function which is constant on each component of X. From this and Ascoli's Theorem it follows that $\mathbf{X}$ is an e-chain.

It appears in general that such pathologies are typical of "non-e" Feller chains, and this again reinforces the value of our results for e-chains, which constitute the more typical behavior of Feller chains.

6.5 Commentary

The weak Feller chain has been a basic starting point in certain approaches to Markov chain theory for many years. The work of Foguel [78, 80], Jamison [108, 109, 110], Lin [154], Rosenblatt [229] and Sine [241, 242, 243] have established a relatively rich theory based on this approach, and the seminal book of Dynkin [70] uses the Feller property extensively.

We will revisit this in much greater detail in Chapter 12, where we will also take up the consequences of the e-chain assumption: this will be shown to have useful attributes in the study of limiting behavior of chains.

The equicontinuity results here, which relate this condition to the dynamical systems viewpoint, are developed by Meyn [170]. Equicontinuity may be compared to *uniform stability* [108] or *regularity* [77]. Whilst e-chains have also been developed in detail, particularly by Rosenblatt [227], Jamison [108, 109] and Sine [241, 242] they do not have particularly useful connections with the ψ-irreducible chains we are about to explore, which explains their relatively brief appearance at this stage.

The concept of continuous components appears first in Pollard and Tweedie [216, 217], and some practical applications are given in Laslett et al [153]. The real exploitation of this concept really begins in Tuominen and Tweedie [269], from which we take Proposition 6.2.2. The connections between T-chains and the existence of compact petite sets is a recent result of Meyn and Tweedie [178].

In practice the identification of ψ-irreducible Feller chains as T-chains provided only that $\operatorname{supp}\psi$ has non-empty interior is likely to make the application of the results for such chains very much more common. This identification is new. The condition that $\operatorname{supp}\psi$ have non-empty interior has however proved useful in a number of associated areas in [217] and in Cogburn [53].

We note in advance here the results of Chapter 9 and Chapter 18, where we will show that a number of stability criteria for general space chains have "topological" analogues which, for T-chains, are exact equivalences. Thus T-chains will prove of on-going interest.

Finding criteria for chains to have continuity properties is a model-by-model exercise, but the results on linear and nonlinear systems here are intended to guide this process in some detail.

The assumption of a spread-out increment process, made in previous chapters for chains such as the unrestricted random walk, may have seemed somewhat arbitrary. It is striking therefore that this condition is both necessary and sufficient for random walk to be a T-chain, as in Proposition 6.3.2 which is taken from Tuominen and Tweedie [269]; they also show that this result extends to random walks on locally compact Haussdorff groups, which are T-chains if and only if the increment measure has some convolution power non-singular with respect to (right) Haar measure. These results have been extended to random walks on semi-groups by Högnas in [98].

In a similar fashion, the analysis carried out in Athreya and Pantula [14] shows that the simple linear model satisfying the eigenvalue condition (LSS5) is a T-chain if and only if the disturbance process is spread out. Chan et al [43] show in effect that for the SETAR model compact sets are petite under positive density assumptions, but the proof here is somewhat more transparent.

These results all reinforce the impression that even for the simplest possible models it is not possible to dispense with an assumption of positive densities, and we adopt it extensively in the models we consider from here on.