Definition 6.81. A sequence $X_0, X_1, X_2, \ldots$ is *α-mixing* if

$$\alpha_n = \sup_{A,B} |P(X_n \in A, X_0 \in B) - P(X_n \in A)P(X_0 \in B)| \tag{6.50}$$

goes to 0 when n goes to infinity.

So, we see that an α-mixing sequence will tend to "look independent" if variables are far enough apart. As a result, we would expect that Theorem 6 is a consequence of α-mixing. This is, in fact, the case, as every positive recurr aperiodic Markov chain is α-mixing (Rosenblatt 1971, Section VII.3), and if Markov chain is stationary and α-mixing, the covariances go to zero (Billing 1995, Section 27).

However, for a Central Limit Theorem, we need even more. Not only must Markov chain be α-mixing, but we need the coefficient α_n to go to 0 fast eno that is, we need the dependence to go away fast enough. One version of a Ma chain Central Limit Theorem is the following (Billingsley 1995, Section 27).

Theorem 6.82. *Suppose that the Markov chain (X_n) is stationary and α-m with $\alpha_n = \mathcal{O}(n^{-5})$ and that $\mathbb{E}[X_n] = 0$ and $\mathbb{E}[X_n^{12}] < \infty$. Then,*

$$\sigma^2 = \lim_{n\to\infty} n \operatorname{var} \bar{X}_n < \infty$$

and if $\sigma^2 > 0$, $\sqrt{n}\bar{X}_n$ tends in law to $\mathcal{N}(0, \sigma^2)$.

This theorem is not very useful because the condition on the mixing coef is very hard to verify. (Billingsley 1995 notes that the conditions are stronger needed, but are imposed to avoid technical difficulties in the proof.) Others worked hard to get the condition in a more accessible form and have exploit relationship between mixing and ergodicity. Informally, if (6.50) goes to 0, di through by $P(X_0 \in B)$ we expect that

$$|P(X_n \in A | X_0 \in B) - P(X_n \in A)| \to 0, \tag{6.51}$$

which looks quite similar to the assumption that the Markov chain is ergodic corresponds, in fact, to a stronger type of mixing called *β-mixing*. See Bradley We actually need something stronger (see Problem 7.6), like uniform erg where there are constants M and $r < 1$ such $|P(X_n \in A | X_0 \in B) - P(X_n$ Mr^n. Tierney (1994) presents the following Central Limit Theorem.

Theorem 6.83. *Let (X_n) be a stationary uniformly ergodic Markov chain. function $h(\cdot)$ satisfying* var $h(X_i) = \sigma^2 < \infty$, *there exists a real number τ_h s*

$$\sqrt{n}\,\frac{\sum_{i=1}^n h(X_i)/n - \mathbb{E}[h(X)]}{\tau_h} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Other versions of the Central Limit Theorem exist. See, for example (1994), who surveys other mixing conditions and their connections with th Limit Theorem.

Again, we refer to Eaton (1992) for extensions, examples, and comments on this result. Note, however, that the verification of the recurrence of the Markov chain $(\theta^{(n)})$ is much easier than the determination of the lower bound of $\Delta(h)$. Hobert and Robert (1999) consider the potential of using the dual chain based on the kernel

$$K'(x,y) = \int_\Theta f(y|\theta)\pi(\theta|x)d\theta \tag{6.48}$$

(see Problem 6.64) and derive admissibility results for various distributions of interest.

6.9.3 Alternative Convergence Conditions

Athreya et al. (1996) present a careful development of the basic limit theorems for Markov chains, with conditions stated that are somewhat more accessible in Markov chain Monte Carlo uses, rather than formal probabilistic properties.

Consider a time-homogeneous Markov chain (X_n) where f is the invariant density and $f_k(\cdot|\cdot)$ is the conditional density of X_k given X_0. So, in particular, $f_1(\cdot|\cdot)$ is the transition kernel. For a basic limit theorem such as Theorem 6.51, there are two conditions that are required on the transition kernel, both of which have to do with the ability of the Markov chain to visit all sets A. Assume that the transition kernel satisfies[5]: There exists a set A such that

(i) $\sum_{k=1}^{\infty} \int_A f_k(x|x_0)\, d\mu(x) > 0$ for all x_0,
(ii) $\inf_{x,y\in A} f_1(y|x) > 0$.

A set A satisfying (i) is called *accessible*, which means that from anywhere in the state space there is positive probability of eventually entering A. Condition (ii) is essentially a minorization condition. The larger the set A, the easier it is to verify (i) and the harder it is to verify (ii) . These two conditions imply that the chain is irreducible and aperiodic. It is possible to weaken (ii) to a condition that involves f_k for some $k \geq 1$; see Athreya et al. (1996).

The limit theorem of Athreya et al. (1996) can be stated as follows.

Theorem 6.80. *Suppose that the Markov chain (X_n) has invariant density $f(\cdot)$ and transition kernel $f_1(\cdot|\cdot)$ that satisfies Conditions (i) and (ii). Then*

$$\lim_{k\to\infty} \sup_A \left| \int_A f_k(x|x_0)dx - \int_A f(x)dx \right| = 0 \tag{6.49}$$

for f almost all x_0.

6.9.4 Mixing Conditions and Central Limit Theorems

In Section 6.7.2, we established a Central Limit Theorem using regeneration, which allowed us to use a typical independence argument. Other conditions, known as *mixing conditions*, can also result in a Central Limit Theorem. These mixing conditions guarantee that the dependence in the Markov chain decreases fast enough, and variables that are far enough apart are close to being independent. Unfortunately, these conditions are usually quite difficult to verify. Consider the property of *α-mixing* (Billingsley 1995, Section 27).

[5] The conditions stated here are weaker than those given in the first edition; we thank Hani Doss for showing us this improvement.

6.9.2 Eaton's Admissibility Condition

Eaton (1992) exhibits interesting connections, similar to Brown (1971), between the admissibility of an estimator and the recurrence of an associated Markov chain. The problem considered by Eaton (1992) is to determine whether, for a *bounded* function $g(\theta)$, a generalized Bayes estimator associated with a prior measure π is admissible under quadratic loss. Assuming that the posterior distribution $\pi(\theta|x)$ is well defined, he introduces the transition kernel

$$K(\theta,\eta) = \int_{\mathcal{X}} \pi(\theta|x) f(x|\eta)\, dx, \tag{6.46}$$

which is associated with a Markov chain $(\theta^{(n)})$ generated as follows: The transition from $\theta^{(n)}$ to $\theta^{(n+1)}$ is done by generating first $x \sim f(x|\theta^{(n)})$ and then $\theta^{(n+1)} \sim \pi(\theta|x)$. (Most interestingly, this is also a kernel used by Markov Chain Monte Carlo methods, as shown in Chapter 9.) Note that the prior measure π is an invariant measure for the chain $(\theta^{(n)})$. For every measurable set C such that $\pi(C) < +\infty$, consider

$$V(C) = \left\{h \in \mathcal{L}^2(\pi); h(\theta) \geq 0 \text{ and } h(\theta) \geq 1 \text{ when } \theta \in C\right\}$$

and

$$\Delta(h) = \int\int \left\{h(\theta) - h(\eta)\right\}^2 K(\theta,\eta)\pi(\eta)\, d\theta\, d\eta.$$

The following result then characterizes admissibility for *all bounded functions* in terms of Δ and $V(C)$ (that is, independently of the estimated functions g).

Theorem 6.78. *If for every C such that $\pi(C) < +\infty$,*

$$\inf_{h \in V(C)} \Delta(h) = 0, \tag{6.47}$$

then the Bayes estimator $\mathbb{E}^{\pi}[g(\theta)|x]$ is admissible under quadratic loss for every bounded function g.

This result is obviously quite general but only mildly helpful in the sense that the practical verification of (6.47) for every set C can be overwhelming. Note also that (6.47) always holds when π is a proper prior distribution since $h \equiv 1$ belongs to $\mathcal{L}^2(\pi)$ and $\Delta(1) = 0$ in this case. The extension then considers approximations of 1 by functions in $V(C)$. Eaton (1992) exhibits a connection with the Markov chain $(\theta^{(n)})$, which gives a condition equivalent to Theorem 6.78. First, for a given set C, a stopping rule τ_C is defined as the first integer $n > 0$ such that $(\theta^{(n)})$ belongs to C (and $+\infty$ otherwise), as in Definition 6.10.

Theorem 6.79. *For every set C such that $\pi(C) < +\infty$,*

$$\inf_{h \in V(C)} \Delta(h) = \int_C \left\{1 - P(\tau_C < +\infty | \theta^{(0)} = \eta)\right\} \pi(\eta)\, d\eta.$$

Therefore, the generalized Bayes estimators of bounded functions of θ are admissible if and only if the associated Markov chain $(\theta^{(n)})$ is recurrent.

$$\begin{aligned}\Delta V(x) &\le -(1-|\theta|)\,V(x) + \mathbb{E}[|\varepsilon_1|] + 1 - |\theta| \\ &= -(1-|\theta|)\,\gamma V(x) + \mathbb{E}[|\varepsilon_1|] + 1 - |\theta| - (1-\gamma)(1-|\theta|)\,V(x) \\ &\le -\beta V(x) + b\,\mathbb{I}_C(x)\end{aligned}$$

for $\beta = (1-|\theta|)\,\gamma$, $b = \mathbb{E}[|\varepsilon_1|] + 1 - |\theta|$, and C equal to

$$C = \{x;\ V(x) < (\mathbb{E}[|\varepsilon_1|] + 1 - |\theta|)/(1-\gamma)(1-|\theta|)\}\ ,$$

if $|\theta| < 1$ and $\mathbb{E}[|\varepsilon_1|] < +\infty$. These conditions thus imply geometric ergodicity for AR(1) models. ‖

Meyn and Tweedie (1994) propose, in addition, *explicit* evaluations of convergence rates r as well as explicit bounds R in connection with drift conditions (6.42), but the geometric convergence is evaluated under a norm induced by the very function V satisfying (6.42), which makes the result somewhat artificial.

There also is an equivalent form of *uniform ergodicity* involving drift, namely that (X_n) is aperiodic and there exist a small set C, a bounded potential function $V \ge 1$, and constants $0 < b < \infty$ and $\beta > 0$ such that

$$\Delta V(x) \le -\beta V(x) + b\,\mathbb{I}_C(x)\,, \qquad x \in \mathcal{X}\,. \tag{6.44}$$

In a practical case (see, e.g., Example 12.6), this alternative to the conditions of Theorem 6.59 is often the most natural approach.

As mentioned after Theorem 6.64, there exist alternative versions of the Central Limit Theorem based on drift conditions. Assume that there exist a function $f \ge 1$, a finite potential function V, and a small set C such that

$$\Delta V(x) \le -f(x) + b\,\mathbb{I}_C(x), \qquad x \in \mathcal{X}, \tag{6.45}$$

and that $\mathbb{E}^\pi[V^2] < \infty$. This is exactly condition (6.44) above, with $f = V$, which implies that (6.45) holds for an uniformly ergodic chain.

Theorem 6.77. *If the ergodic chain (X_n) with invariant distribution π satisfies conditions (6.45), for every function g such that $|g| \le f$, then*

$$\begin{aligned}\gamma_g^2 &= \lim_{n\to\infty} n\mathbb{E}_\pi[S_n^2(\overline{g})] \\ &= \mathbb{E}_\pi[\overline{g}^2(x_0)] + 2\sum_{k=1}^{\infty} \mathbb{E}_\pi[\overline{g}(x_0)\overline{g}(x_k)]\end{aligned}$$

is non-negative and finite. If $\gamma_g > 0$, the Central Limit Theorem holds for $S_n(\overline{g})$. If $\gamma_g = 0$, $\sqrt{n}S_n(\overline{g})$ almost surely goes to 0.

This theorem is definitely relevant for convergence assessment of Markov chain Monte Carlo algorithms since, when $\gamma_g^2 > 0$, it is possible to assess the convergence of the ergodic averages $S_n(g)$ to the quantity of interest $\mathbb{E}^\pi[g]$. Theorem 6.77 also suggests how to implement this monitoring through renewal theory, as discussed in detail in Chapter 12.

(b) there exist a small set C and a positive number M_C such that

$$\sup_{x \in C} \mathbb{E}_x[\tau_C] \leq M_C \; ;$$

(c) there exist a small set C, a function V taking values in $\mathbb{R} \cup \{\infty\}$, and a positive real number b such that

$$\Delta V(x) \leq -1 + b \mathbb{I}_C(x) \; . \tag{6.40}$$

See Meyn and Tweedie (1993, Chapter 11) for a proof and discussion of these equivalences. (If there exists V finite and bounded on C which satisfies (6.40), the chain (X_n) is necessarily Harris positive.)

The notion of a *Kendall atom* introduced in Section 6.6.2 can also be extended to non-atomic chains by defining *Kendall sets* as sets A such that

$$\sup_{x \in A} \mathbb{E}_x \left[\sum_{k=0}^{\tau_A - 1} \kappa^k \right] < \infty \; . \tag{6.41}$$

with $\kappa > 1$. The existence of a Kendall set guarantees a *geometric drift* condition. If C is a Kendall set and if

$$V(x) = \mathbb{E}_x[\kappa^{\sigma_C}] \; ,$$

the function V satisfies

$$\Delta V(x) \leq -\beta V(x) + b \mathbb{I}_C(x) \tag{6.42}$$

with $\beta > 0$ and $0 < b < \infty$. This condition also guarantees geometric convergence for (X_n) in the following way.

Theorem 6.75. *For a ψ-irreducible and aperiodic chain (X_n) and a small Kendall set C, there exist $R < \infty$ and $r > 1$, $\kappa > 1$ such that*

$$\sum_{n=1}^{\infty} r^n \| K^n(x, \cdot) - \pi(\cdot) \| \leq R \, \mathbb{E}_x \left[\sum_{k=0}^{\tau_C} \kappa^k \right] < \infty \tag{6.43}$$

for almost every $x \in \mathcal{X}$.

The three conditions (6.41), (6.42) and (6.43) are, in fact, equivalent for ψ-irreducible aperiodic chains if A is a small set in (6.41) and if V is bounded from below by 1 in (6.42) (see Meyn and Tweedie 1993, pp. 354–355). The drift condition (6.42) is certainly the simplest to check in practice, even though the potential function V must be derived.

Example 6.76. (Continuation of Example 6.20) The condition $|\theta| < 1$ is necessary for the chain $X_n = \theta x_{n-1} + \varepsilon_n$ to be recurrent. Assume ε_n has a strictly positive density on $\mathbb{R}$. Define $V(x) = |x| + 1$. Then

$$\begin{aligned} \mathbb{E}_x[V(X_1)] &= 1 + \mathbb{E}[|\theta X + \varepsilon_1|] \\ &\leq 1 + |\theta| \, |x| + \mathbb{E}[|\varepsilon_1|] \\ &= |\theta| \, V(x) + \mathbb{E}[|\varepsilon_1|] + 1 - |\theta| \end{aligned}$$

and

Proof. If $C = \{x; V(x) \leq r\}$ and M is a bound on V, the conditions (6.38) are satisfied by

$$\tilde{V}(x) = \begin{cases} (M - V(x))/(M - r) & \text{if } x \in C^c \\ 1 & \text{if } x \in C. \end{cases}$$

Since $\tilde{V}(x) < 1$ for $x \in C^c$, $V^*(x) = P_x(\tau_C < \infty) < 1$ on C^c, and this implies the transience of C, therefore the transience of (X_n). The converse can be deduced from a (partial) converse to Proposition 6.31 (see Meyn and Tweedie 1993, p. 190). □

Condition (6.39) describes an average increase of $V(x_n)$ once a certain level has been attained, and therefore does not allow a sure return to 0 of V. The condition is thus incompatible with the stability associated with recurrence. On the other hand, if there exists a potential function V "attracted" to 0, the chain is recurrent.

Theorem 6.72. *Consider (X_n) a ψ-irreducible Markov chain. If there exist a small set C and a function V such that*

$$C_V(n) = \{x; V(x) \leq n\}$$

is a small set for every n, the chain is recurrent if

$$\Delta V(x) \leq 0 \text{ on } C^c.$$

The fact that $C_V(n)$ is small means that the function V is not bounded outside small sets. The attraction of the chain toward smaller values of V on the sets where V is large is thus a guarantee of stability for the chain. The proof of the above result is, again, quite involved, based on the fact that $P_x(\tau_C < \infty) = 1$ (see Meyn and Tweedie 1993, p. 191).

Example 6.73. (Continuation of Example 6.39) If the distribution of W_n has a finite support and zero expectation, (X_n) *is recurrent.* When considering $V(x) = |x|$ and r such that $\gamma_x = 0$ for $|x| > r$, we get

$$\Delta V(x) = \sum_{n=-r}^{r} \gamma_n (|x + n| - |x|) ,$$

which is equal to

$$\sum_{n=-r}^{r} \gamma_n n \text{ if } x \geq r \qquad \text{and} \qquad -\sum_{n=-r}^{r} \gamma_n n \text{ if } x \leq -r .$$

Therefore, $\Delta V(x) = 0$ for $x \notin \{-r+1, \ldots, r-1\}$, which is a small set. Conversely, if W_n has a nonzero mean, X_n is transient. ‖

For Harris recurrent chains, positivity can also be related to a drift condition and to a "regularity" condition on visits to small sets.

Theorem 6.74. *If (X_n) is Harris recurrent with invariant measure π, there is equivalence between*

(a) π is finite;

(a) Show that

$$\text{var}[\mathbb{E}(X_k|X_0)] = \mathbb{E}[\mathbb{E}(X_k|X_0)]^2,$$
$$\text{var}[\mathbb{E}(X_k|X_0)] \geq \text{var}[\mathbb{E}(X_{k+1}|X_0)]] \,.$$

(*Hint*: Write $f_{k+1}(y|x) = \int f_k(y|x')f(x'|x)dx'$ and use Fubini and Jensen.)

(b) Show that

$$\mathbb{E}[\text{var}(X_k|X_0)] \leq \mathbb{E}[\text{var}(X_{k+1}|X_0)]$$

and that

$$\lim_{k\to\infty} \mathbb{E}[\text{var}(X_k|X_0)] = \sigma^2.$$

6.9 Notes

6.9.1 Drift Conditions

Besides atoms and small sets, Meyn and Tweedie (1993) rely on another tool to check or establish various stability results, namely, *drift criteria*, which can be traced back to Lyapunov. Given a function V on $\mathcal{X}$, the *drift of* V is defined by

$$\Delta V(x) = \int V(y)\, P(x, dy) - V(x) \,.$$

(Functions V appearing in this setting are often referred to as *potentials*; see Norris 1997.) This notion is also used in the following chapters to verify the convergence properties of some MCMC algorithms (see, e.g., Theorem 7.15 or Mengersen and Tweedie 1996).

The following lemma is instrumental in deriving drift conditions for the transience or the recurrence of a chain (X_n).

Lemma 6.70. *If $C \in \mathcal{B}(\mathcal{X})$, the smallest positive function which satisfies the conditions*

$$\Delta V(x) \leq 0 \quad \text{if} \quad x \notin C, \quad V(x) \geq 1 \quad \text{if} \quad x \in C \tag{6.38}$$

is given by

$$V^*(x) = P_x(\sigma_C < \infty) \,,$$

where σ_C denotes

$$\sigma_C = \inf\{n \geq 0; x_n \in C\} \,.$$

Note that, if $x \notin C$, $\sigma_C = \tau_C$, while $\sigma_C = 0$ on C. We then have the following necessary and sufficient condition.

Theorem 6.71. *The ψ-irreducible chain (X_n) is transient if and only if there exist a bounded positive function V and a real number $r \geq 0$ such that for every x for which $V(x) > r$, we have*

$$\Delta V(x) > 0 \,. \tag{6.39}$$

6.61 (Kemeny and Snell 1960) Show that for an aperiodic irreducible Markov chain with finite state-space and with transition matrix $\mathbb{P}$, there always exists a stationary probability distribution which satisfies

$$\pi = \pi\mathbb{P}.$$

(a) Show that if $\beta < 0$, the random walk is recurrent. (*Hint:* Use the drift function $V(x) = x$ as in Theorem 6.71.)
(b) Show that if $\beta = 0$ and $\text{var}(\epsilon_n) < \infty$, (X_n) is recurrent. (*Hint:* Use $V(x) = \log(1+x)$ for $x > R$ and $V(x) = 0$, otherwise, for an adequate bound R.)
(c) Show that if $\beta > 0$, the random walk is transient.

6.62 Show that if there exist a finite potential function V and a small set C such that V is bounded on C and satisfies (6.40), the corresponding chain is Harris positive.

6.63 Show that the random walk on $\mathbb{Z}$ is transient when $\mathbb{E}[W_n] \neq 0$.

6.64 Show that the chains defined by the kernels (6.46) and (6.48) are either both recurrent or both transient.

6.65 Referring to Example 6.66, show that the AR(1) chain is reversible.

6.66 We saw in Section 6.6.2 that a stationary Markov chain is *geometrically ergodic* if there is a non-negative real-valued function M and a constant $r < 1$ such that for any $A \in \mathcal{X}$,

$$|P(X_n \in A | X_0 \in B) - P(X_n \in A)| \leq M(x)r^n.$$

Prove that the following Central Limit Theorem (due to Chan and Geyer 1994) can be considered a corollary to Theorem 6.82 (see Note 6.9.4):

Corollary 6.69. *Suppose that the stationary Markov chain $X_0, X_1, X_2, \ldots$ is geometrically ergodic with $M^* = \int |M(x)| f(x)dx < \infty$ and satisfies the moment conditions of Theorem 6.82. Then*

$$\sigma^2 = \lim_{n\to\infty} n \,\text{var}\bar{X}_n < \infty$$

and if $\sigma^2 > 0$, $\sqrt{n}\bar{X}_n/\sigma$ tends in law to $\mathcal{N}(0, \sigma^2)$.

(*Hint*: Integrate (with respect to f) both sides of the definition of geometric ergodicity to conclude that the chain has exponentially fast α-mixing, and apply Theorem 6.82.)

6.67 Suppose that $X_0, X_1, \ldots, X_n$ have a common mean ξ and variance σ^2 and that $\text{cov}(X_i, X_j) = \rho_{j-i}$. For estimating ξ, show that

(a) $\bar{X}$ may not be consistent if $\rho_{j-i} = \rho \neq 0$ for all $i \neq j$. (*Hint*: Note that $\text{var}(\bar{X}) > 0$ for all sufficiently large n requires $\rho \geq 0$ and determine the distribution of $\bar{X}$ in the multivariate normal case.)
(b) $\bar{X}$ is consistent if $|\rho_{j-i}| \leq M\gamma^{j-i}$ with $|\gamma| < 1$.

6.68 For the situation of Example 6.84:

(a) Prove that the sequence (X_n) is stationary provided $\sigma^2 = 1/(1-\beta^2)$.
(b) Show that $\mathbb{E}(X_k|x_0) = \beta^k x_0$. (*Hint*: Consider $\mathbb{E}[(X_k - \beta X_{k-1})|x_0]$.)
(c) Show that $\text{cov}(X_0, X_k) = \beta^k/(1-\beta^2)$.

6.69 Under the conditions of Theorem 6.85, it follows that $\mathbb{E}[\mathbb{E}(X_k|X_0)]^2 \to 0$. There are some other interesting properties of this sequence.

6.55 (Continuation of Problem 6.54) Show that an ergodic random walk on a finite state-space is reversible.

6.56 (Kemeny and Snell 1960) A Markov chain (X_n) is *lumpable* with respect to a nontrivial partition of the state-space, $(A_1, \ldots, A_k)$, if, for every initial distribution μ, the process

$$Z_n = \sum_{i=1}^{k} i \mathbb{I}_{A_i}(X_n)$$

is a Markov chain with transition probabilities independent of μ.

(a) Show that a necessary and sufficient condition for lumpability is that

$$p_{uA_j} = \sum_{v \in A_j} p_{uv}$$

is constant (in n) on A_i for every i.

(b) Examine whether

$$\mathbb{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$

is lumpable for $A_1 = \{1, 2\}$, $A_2 = \{3, 4\}$, and $A_3 = \{5\}$.

6.57 Consider the random walk on $\mathbb{R}^+$, $X_{n+1} = (X_n + \epsilon_n)^+$, with $\mathbb{E}[\epsilon_n] = \beta$.

(a) Establish Lemma 6.70. (*Hint:* Consider an alternative V to V^* and show by recurrence that

$$V(x) \geq \int_C K(x, y)V(y)dy + \int_{C^c} K(x, y)V(y)dy \geq \cdots \geq V^*(x)\,.)$$

(b) Establish Theorem 6.72 by assuming that there exists x^* such that $P_{x^*}(\tau_C < \infty) < 1$, choosing M such that $M \geq V(x^*)/[1 - P_{x^*}(\tau_C < \infty)]$ and establishing that $V(x^*) \geq M[1 - P_{x^*}(\tau_C < \infty)]$.

6.58 Show that

(a) a time-homogeneous Markov chain (X_n) is stationary if the initial distribution is the invariant distribution;

(b) the invariant distribution of a stationary Markov chain is also the marginal distribution of any X_n.

6.59 Referring to Section 6.7.1, let X_n be a Markov chain and $h(\cdot)$ a function with $\mathbb{E}h(X_n) = 0$, $\mathrm{Var}h(X_n) = \sigma^2 > 0$, and $\mathbb{E}h(X_{n+1}|x_n) = h(x_n)$, so $h(\cdot)$ is a nonconstant harmonic function.

(a) Show that $\mathbb{E}h(X_{n+1}|x_0) = h(x_0)$.

(b) Show that $\mathrm{Cov}(h(x_0), h(X_n)) = \sigma^2$.

(c) Use (6.52) to establish that $\mathrm{Var}\left(\frac{1}{n+1}\sum_{i=0}^{n} h(X_i)\right) \to \infty$ as $n \to \infty$, showing that the chain is not ergodic.

6.60 Show that if an irreducible Markov chain has a σ-finite invariant measure, this measure is unique up to a multiplicative factor. (*Hint:* Use Theorem 6.63.)

(a) Show that for every state i, $\mathbb{E}_i[f_j]$ is finite.
(b) Show that the matrix M with entries $m_{ij} = \mathbb{E}_i[f_j]$ can be written $M = \mathbb{P}(M - M_d) + E$, where M_d is the diagonal matrix with same diagonal as M and E is the matrix made of 1's.
(c) Deduce that $m_{ii} = 1/\pi_i$.
(d) Show that πM is the vector of the z_{ii}/π_i's.
(e) Show that for every pair of initial distributions, (μ, ν),

$$\mathbb{E}_\mu[f_i] - \mathbb{E}_\nu[f_i] = (\mu - \nu)(I - \mathbb{Z})D,$$

where D is the diagonal matrix $\text{diag}(1/\pi_i)$.

6.50 If h is a function taking values on a finite state-space $\{1, \ldots, r\}$, with $h(i) = h_i$, and if (X_n) is an irreducible Markov chain, show that

$$\lim_{n\to\infty} \frac{1}{n} \text{var}\left(\sum_{t=1}^{n} h(x_t)\right) = \sum_{i,j} h_i c_{ij} h_j,$$

where $c_{ij} = \pi_i z_{ij} + \pi_j z_{ji} - \pi_i \delta_{ij} - \pi_i \pi_j$ and δ_{ij} is Kroenecker's 0–1 function.

6.51 For the two-state transition matrix $\mathbb{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$, show that
(a) the stationary distribution is $\pi = (\beta/(\alpha+\beta), \alpha/(\alpha+\beta))$;
(b) the mean first passage matrix is

$$M = \begin{pmatrix} (\alpha+\beta)/\beta & 1/\alpha \\ 1/\beta & (\alpha+\beta)/\alpha \end{pmatrix};$$

(c) and the limiting variance for the number of times in state j is $\alpha\beta(2-\alpha-\beta)/(\alpha+\beta)^3$, for $j = 1, 2$.

6.52 Show that a finite state-space chain is always geometrically ergodic.

6.53 (Kemeny and Snell 1960) Given a finite state-space Markov chain, with transition matrix $\mathbb{P}$, define a second transition matrix by

$$p_{ij}(n) = \frac{P_\mu(X_{n-1} = j) P(X_n = i | X_{n-1} = j)}{P_\mu(X_n = j)}.$$

(a) Show that $p_{ij}(n)$ does not depend on n if the chain is stationary (i.e., if $\mu = \pi$).
(b) Explain why, in this case, the chain with transition matrix $\tilde{\mathbb{P}}$ made of the probabilities

$$\tilde{p}_{ij} = \frac{\pi_j p_{ji}}{\pi_i}$$

is called the *reverse* Markov chain.
(c) Show that the limiting variance C is the same for both chains.

6.54 (Continuation of Problem 6.53) A Markov chain is *reversible* if $\tilde{\mathbb{P}} = \mathbb{P}$. Show that every two-state ergodic chain is reversible and that an ergodic chain with symmetric transition matrix is reversible. Examine whether the matrix

$$\mathbb{P} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

is reversible. (*Hint:* Show that $\pi = (0.1, 0.2, 0.4, 0.2, 0.1)$.)

6.44 (Continuation of Problem 6.43) Consider two independent forward recurrence time processes (V_n^+) and (W_n^+) with the same generating probability distribution p.

(a) Give the transition probabilities of the joint process $V_n^* = (V_n^+, W_n^+)$.

(b) Show that (V_n^*) is irreducible when p is aperiodic. (*Hint:* Consider r and s such that g.c.d.$(r, s) = 1$ with $p(r) > 0$, $p(q) > 0$, and show that if $nr - ms = 1$ and $i \geq j$, then

$$P_{(i,j)}(V^*_{j+(i-j)nr} = (1,1)) > 0\ .)$$

(c) Show that $\pi^* = \pi \times \pi$, with π defined in Problem 6.43 is invariant and, therefore, that (V_n^*) is positive Harris recurrent when $m_p < \infty$.

6.45 (Continuation of Problem 6.44) Consider V_n^* defined in Problem 6.44 associated with (S_n, S'_n) and define $\tau_{1,1} = \min\{n; V_n^* = (1,1)\}$.

(a) Show that $T_{pq} = \tau_{1,1} + 1$.

(b) Use (c) in Problem 6.44 to show Lemma 6.49.

6.46 (Kemeny and Snell 1960) Establish (directly) the Law of Large Numbers for a finite irreducible state-space chain (X_n) and for $h(x_n) = \mathbb{I}_j(x_n)$, if j is a possible state of the chain; that is,

$$\frac{1}{N}\sum_{n=1}^{N} \mathbb{I}_j(x_n) \longrightarrow \pi_j\ ,$$

where $\pi = (\pi_1, \ldots, \pi_j, \ldots)$ is the stationary distribution.

6.47 (Kemeny and Snell 1960) Let $\mathbb{P}$ be a regular transition matrix, that is, $\mathbb{P}A = A\mathbb{P}$ (see Problem 6.9), with limiting (stationary) matrix A; that is, each column of A is equal to the stationary distribution.

(a) Show that the so-called *fundamental matrix* $\mathbb{Z} = (I - (\mathbb{P} - A))^{-1}$ exists.

(b) Show that $\mathbb{Z} = I + \sum_{n=1}^{\infty}(\mathbb{P}^n - A)$.

(c) Show that $\mathbb{Z}$ satisfies $\pi\mathbb{Z} = \pi$ and $\mathbb{P}\mathbb{Z} = \mathbb{Z}\mathbb{P}$, where π denotes a row of A (this is the stationary distribution).

6.48 (Continuation of Problem 6.47) Let $N_j(n)$ be the number of times the chain is in state j in the first n instants.

(a) Show that for every initial distribution μ,

$$\lim_{n\to\infty} \mathbb{E}_\mu[N_j(n)] - n\pi_j = \mu(\mathbb{Z} - A).$$

(*Note:* This convergence shows the strong stability of a recurrent chain since each term in the difference goes to infinity.)

(b) Show that for every pair of initial distributions, (μ, ν),

$$\lim_{n\to\infty} \mathbb{E}_\mu[N_j(n)] - \mathbb{E}_\nu[N_j(n)] = (\mu - \nu)\mathbb{Z}.$$

(c) Deduce that for every pair of states, (u, v),

$$\lim_{n\to\infty} \mathbb{E}_u[N_j(n)] - \mathbb{E}_v[N_j(n)] = z_{uj} - z_{vj},$$

which is called the *divergence* $\mathrm{div}_j(u, v)$.

6.49 (Continuation of Problem 6.47) Let f_j denote the number of steps before entering state j.

6.40 (a) Verify (6.26), namely, that $\|\mu\|_{TV} = (1/2)\sup_{|h|\le 1} \left|\int h(x)\mu(dx)\right|$.
(b) Show that (6.26) is compatible with the definition of the total variation norm. Establish the relation with the alternative definition

$$||\mu||_{TV} = \sup_A \mu(A) - \inf_A \mu(A).$$

6.41 Show that if (X_n) and (X'_n) are coupled at time N_0 and if $X_0 \sim \pi$, then $X'_n \sim \pi$ for $n > N_0$ for any initial distribution of X'_0.

6.42 Using the notation of Section 6.6.1, set

$$u(n) = \sum_{j=0}^{\infty} p^{j\star}(n)$$

with $p^{j\star}$ the distribution of the sum $S_1 + \cdots + S_j$, $p^{0\star}$ the Dirac mass at 0, and

$$Z(n) = \mathbb{I}_{\exists j; S_j = n}.$$

(a) Show that $P_q(Z(n) = 1) = q \star u(n)$.
(b) Show that

$$|q \star u(n) - p \star u(n)| \le 2\, P(T_{pq} > n).$$

(This bound is often called *Orey's inequality*, from Orey 1971. See Problem 7.10 for a slightly different formulation.)
(c) Show that if m_p is finite,

$$e(n) = \frac{\sum_{n+1}^{\infty} p(j)}{m_p}$$

is the invariant distribution of the renewal process in the sense that $P_e(Z(n) = 1) = 1/m_p$ for every n.
(d) Deduce from Lemma 6.49 that

$$\lim_n \left| q \star u(n) - \frac{1}{m_p} \right| = 0$$

when the mean renewal time is finite.

6.43 Consider the so-called "forward recurrence time" process V_n^+, which is a Markov chain on $\mathbb{N}_+$ with transition probabilities

$$\begin{aligned} P(1, j) &= p(j), \qquad j \ge 1, \\ P(j, j-1) &= 1, \qquad j > 1, \end{aligned}$$

where p is an arbitrary probability distribution on $\mathbb{N}_+$. (See Problem 6.12.)
(a) Show that (V_n^+) is recurrent.
(b) Show that

$$P(V_n^+ = j) = p(j + n - 1).$$

(c) Deduce that the invariant measure satisfies

$$\pi(j) = \sum_{n \ge j} p(n)$$

and show it is finite if and only if

$$m_p = \sum_n np(n) < \infty.$$

6.32 Referring to (6.21):
(a) Show that $\mathbb{E}_\nu[\tau_C] < \infty$;
(b) show that $\sum_\nu P(\tau_C \geq t) = \mathbb{E}_\nu[\tau_C]$.

6.33 Let $\Gamma = \{Z_n : n = 0, 1, \ldots\}$ be a discrete time homogeneous Markov chain with state space $\mathcal{Z}$ and Markov transition kernel

$$M(z, \cdot) = \omega\nu(\cdot) + (1 - \omega)K(z, \cdot)\,, \tag{6.37}$$

where $\omega \in (0, 1)$ and ν is a probability measure.
(a) Show that the measure

$$\phi(\cdot) = \sum_{i=1}^{\infty} \omega(1 - \omega)^{i-1} K^{i-1}(\nu, \cdot)$$

is an invariant probability measure for Γ.
(b) Deduce that Γ is positive recurrent.
(c) Show that, when Φ satisfies a minorization condition with $C = \mathcal{X}$, (6.10) holds for all $x \in \mathcal{X}$ and is thus a mixture of the form (6.37).
(*Note:* Even if the Markov chain associated with K is badly behaved, e.g., transient, Γ is still positive recurrent. Breyer and Roberts (2000b) propose another derivation of this result, through the functional identity

$$\int \phi(x)M(x, z)dx = \pi(z)\,.\Big)$$

6.34 Establish the equality (6.14).

6.35 Consider the simple Markov chain (X_n), where each X_i takes on the values -1 and 1 with $P(X_{i+1} = 1|X_i = -1) = 1$, $P(X_{i+1} = -1|X_i = 1) = 1$, and $P(X_0 = 1) = 1/2$.
(a) Show that this is a stationary Markov chain.
(b) Show that $\text{cov}(X_0, X_k)$ does not go to zero.
(c) The Markov chain is not strictly positive. Verify this by exhibiting a set that has positive unconditional probability but zero conditional probability.
(*Note:* The phenomenon seen here is similar to what Seidenfeld and Wasserman 1993 call a *dilation*.)

6.36 In the setup of Example 6.5, find the stationary distribution associated with the proposed transition when $\pi_i = \pi_j$ and in general.

6.37 Show the decomposition of the "first entrance and last exit" equation (6.23).

6.38 If (a_n) is a sequence of real numbers converging to a, and if $b_n = (a_1 + \cdots + a_n)/n$, then show that

$$\lim_n b_n = a\,.$$

(*Note:* The sum $(1/n)\sum_{i=1}^n a_i$ is called a *Cesàro average*; see Billingsley 1995, Section A30.)

6.39 Consider a sequence (a_n) of positive numbers which is converging to a^* and a convergent series with running term b_n. Show that the convolution

$$\sum_{j=1}^{n-1} a_j b_{n-j} \overset{n\to\infty}{\longrightarrow} a^* \sum_{j=1}^{\infty} b_j.$$

(*Hint:* Use the Dominated Convergence Theorem.)

6.25 (Continuation of Problem 6.22) A Markov chain that is not positive recurrent may be either null recurrent or *transient*. In either of these latter two cases, the invariant distribution, if it exists, is not a probability distribution (it does not have a finite integral), and the difference is one of expected return times. For any integer j, the probability of returning to j in k steps is $p_{jj}^{(k)} = P(X_{i+k} = j|X_i = j)$, and the expected return time is thus $m_{jj} = \sum_{k=1}^{\infty} k p_{jj}^{(k)}$.

(a) Show that since the Markov chain is irreducible, $m_{jj} = \infty$ either for all j or for no j; that is, for any two states x and y, x is transient if and only if y is transient.

(b) An irreducible Markov chain is *transient* if $m_{jj} = \infty$; otherwise it is recurrent. Show that the random walk is positive recurrent if $p < 1/2$ and transient if $p > 1/2$.

(c) Show that the random walk is null recurrent if $p = 1/2$. This is the interesting case where each state will be visited infinitely often, but the expected return time is infinite.

6.26 Explain why the resolvant chain is necessarily strongly irreducible.

6.27 Consider a random walk on $\mathbb{R}_+$, defined as

$$X_{n+1} = (X_n + \epsilon)^+.$$

Show that the sets $(0, c)$ are small, provided $P(\epsilon < 0) > 0$.

6.28 Consider a random walk on $\mathbb{Z}$ with transition probabilities

$$P(Z_t = n+1|Z_{t-1} = n) = 1 - P(Z_t = n-1|Z_{t-1} = n) \propto n^{-\alpha}$$

and

$$P(Z_t = 1|Z_{t-1} = 0) = 1 - P(Z_t = -1|Z_{t-1} = 0) = 1/2\,.$$

Study the recurrence properties of the chain in terms of α.

6.29 Establish (i) and (ii) of Theorem 6.28.

(a) Use

$$K^n(x, A) \geq K^r(x, \alpha) K^s(\alpha, \alpha) K^t(\alpha, A)$$

for $r + s + t = n$ and r and s such that

$$K^r(x, \alpha) > 0 \qquad \text{and} \qquad K^s(\alpha, A) > 0$$

to derive from the Chapman–Kolmogorov equations that $\mathbb{E}_x[\eta_A] = \infty$ when $\mathbb{E}_\alpha[\eta_\alpha] = \infty$.

(b) To show (ii):

a) Establish that transience is equivalent to $P_\alpha(\tau_\alpha < \infty) < 1$.

b) Deduce that $\mathbb{E}_x[\eta_\alpha] < \infty$ by using a generating function as in the proof of Proposition 6.31.

c) Show that the covering of $\mathcal{X}$ is made of the

$$\overline{\alpha}_j = \{y; \sum_{n=1}^{j} K^n(y, \alpha) > j^{-1}\}.$$

6.30 Referring to Definition 6.32, show that if $P(\eta_A = \infty) \neq 0$ then $\mathbb{E}_x[\eta_A] = \infty$, but that $P(\eta_A = \infty) = 0$ does not imply $\mathbb{E}_x[\eta_A] < \infty$.

6.31 In connection with Example 6.42, show that the chain is null recurrent when $f'(1) = 1$.

The *random walk* (Examples 6.40 and 6.39) is a useful probability model and has been given many colorful interpretations. (A popular one is the description of an inebriated individual whose progress along a street is composed of independent steps in random directions, and a question of interest is to describe where the individual will end up.) Here, we look at a simple version to illustrate a number of the Markov chain concepts.

6.22 A random walk on the non-negative integers $I = \{0, 1, 2, \ldots\}$ can be constructed in the following way. For $0 < p < 1$, let $Y_0, Y_1, \ldots$ be iid random variables with $P(Y_i = 1) = p$ and $P(Y_i = -1) = 1 - p$, and $X_k = \sum_{i=0}^{k} Y_i$. Then, (X_n) is a Markov chain with transition probabilities

$$P(X_{i+1} = j + 1 | X_i = j) = p, \qquad P(X_{i+1} = j - 1 | X_i = j) = 1 - p,$$

but we make the exception that $P(X_{i+1} = 1 | X_i = 0) = p$ and $P(X_{i+1} = 0 | X_i = 0) = 1 - p$.

(a) Show that (X_n) is a Markov chain.

(b) Show that (X_n) is also irreducible.

(c) Show that the invariant distribution of the chain is given by

$$a_k = \left(\frac{p}{1-p}\right)^k a_0, \quad k = 1, 2, \ldots,$$

where a_k is the probability that the chain is at k and a_0 is arbitrary. For what values of p and a_0 is this a probability distribution?

(d) If $\sum a_k < \infty$, show that the invariant distribution is also the stationary distribution of the chain; that is, the chain is ergodic.

6.23 If (X_t) is a random walk, $X_{t+1} = X_t + \epsilon_t$, such that ϵ_t has a moment generating function f, defined in a neighborhood of 0, give the moment generating function of X_{t+1}, g_{t+1} in terms of g_t and f, when $X_0 = 0$. Deduce that there is no invariant distribution with a moment generating function in this case.

Although the property of aperiodicity is important, it is probably less important than properties such as recurrence and irreducibility. It is interesting that Feller (1971, Section XV.5) notes that the classification into periodic and aperiodic states "represents a nuisance." However, this is less true when the random variables are continuous.

6.24 (Continuation of Problem 6.22)

(a) Using the definition of periodic given here, show that the random walk of Problem 6.22 is periodic with period 2.

(b) Suppose that we modify the random walk of Problem 6.22 by letting $0 < p + q < 1$ and redefining

$$Y_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p - q \\ -1 & \text{with probability } q. \end{cases}$$

Show that this random walk is irreducible and aperiodic. Find the invariant distribution, and the conditions on p and q for which the Markov chain is positive recurrent.

6.14 Show that the multiplicative random walk

$$X_{t+1} = X_t \epsilon_t$$

is not irreducible when $\epsilon_t \sim \mathcal{E}xp(1)$ and $x_0 \in \mathbb{R}$. (*Hint:* Show that it produces two irreducible components.)

6.15 Show that in the setup of Example 6.17, the chain is not irreducible when ϵ_n is uniform on $[-1, 1]$ and $|\theta| > 1$.

6.16 In the spirit of Definition 6.25, we can define a *uniformly transient set* as a set A for which there exists $M < \infty$ with

$$\mathbb{E}_x[\eta_A] \leq M, \qquad \forall x \in A\,.$$

Show that *transient* sets are denumerable unions of *uniformly transient* sets.

6.17 Show that the split chain defined on $\mathcal{X} \times \{0, 1\}$ by the following transition kernel:

$$\begin{aligned}
&P(\check{X}_{n+1} \in A \times \{0\} | (x_n, 0)) \\
&= \mathbb{I}_C(x_n) \left\{ \frac{P(X_n, A \cap C) - \epsilon\nu(A \cap C)}{1 - \epsilon}(1 - \epsilon) \right. \\
&\quad \left. + \frac{P(X_n, A \cap C^c) - \epsilon\nu(A \cap C^c)}{1 - \epsilon} \right\} \\
&\quad + \mathbb{I}_{C^c}(X_n) \; \{P(X_n, A \cap C)(1 - \epsilon) + P(X_n, A \cap C^c)\}\,, \\
&P(\check{X}_{n+1} \in A \times \{1\} | (x_n, 0)) \\
&= \mathbb{I}_C(X_n) \; \frac{P(X_n, A \cap C) - \epsilon\nu(A \cap C)}{1 - \epsilon}\epsilon + \mathbb{I}_{C^c}(X_n) \; P(X_n, A \cap C)\epsilon,
\end{aligned}$$

$$P(\check{X}_{n+1} \in A \times \{0\} | (x_n, 1)) = \nu(A \cap C)(1 - \epsilon) + \nu(A \cap C^c),$$
$$P(\check{X}_{n+1} \in A \times \{1\} | (x_n, 1)) = \nu(A \cap C)\epsilon\,,$$

satisfies

$$P(\check{X}_{n+1} \in A \times \{1\} | \check{x}_n) = \varepsilon\nu(A \cap C),$$
$$P(\check{X}_{n+1} \in A \times \{0\} | \check{x}_n) = \nu(A \cap C^c) + (1 - \varepsilon)\nu(A \cap C)$$

for every $\check{x}_n \in C \times \{1\}$. Deduce that $C \times \{1\}$ is an atom of the split chain $(\check{X}_n)$.

6.18 If C is a small set and $B \subset C$, under which conditions on B is B a small set?

6.19 If C is a small set and $D = \{x; P^m(x, D) > \delta\}$, show that D is a small set for δ small enough. (*Hint:* Use the Chapman–Kolmogorov equations.)

6.20 Show that the period d given in Definition 6.23 is independent of the selected small set C and that this number characterizes the chain (X_n).

6.21 Given the transition matrix

$$\mathbb{P} = \begin{pmatrix}
0.0 & 0.4 & 0.6 & 0.0 & 0.0 \\
0.6 & 0.0 & .35 & 0.0 & 0.05 \\
0.32 & .68 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.12 & 0.0 & 0.88 \\
0.14 & 0.3 & 0.0 & 0.56 & 0.0
\end{pmatrix},$$

show that the corresponding chain is aperiodic, despite the null diagonal.

6.7 Given the transition matrix

$$\mathbb{P} = \begin{pmatrix} 0.0 & 0.4 & 0.6 & 0.0 & 0.0 \\ 0.65 & 0.0 & 0.35 & 0.0 & 0.0 \\ 0.32 & 0.68 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.12 & 0.88 \\ 0.0 & 0.0 & 0.0 & 0.56 & 0.44 \end{pmatrix},$$

examine whether the corresponding chain is irreducible and aperiodic.

6.8 Show that irreducibility in the sense of Definition 6.13 coincides with the more intuitive notion that two arbitrary states are connected when the Markov chain has a discrete support.

6.9 Show that an aperiodic Markov chain on a finite state-space with transition matrix $\mathbb{P}$ is irreducible if and only if there exists $N \in \mathbb{N}$ such that $\mathbb{P}^N$ has no zero entries. (The matrix is then called *regular*.)

6.10 (Kemeny and Snell 1960) Show that for a regular matrix $\mathbb{P}$:

(a) The sequence $(\mathbb{P}^n)$ converges to a stochastic matrix A.
(b) Each row of A is the same probability vector π.
(c) All components of π are positive.
(d) For every probability vector μ, $\mu\mathbb{P}^n$ converges to π.
(e) π satisfies $\pi = \pi\mathbb{P}$.

(*Note:* See Kemeny and Snell 1960, p. 71 for a full proof.)

6.11 Show that for the measure ψ given by (6.9), the chain (X_n) is irreducible in the sense of Definition 6.13. Show that for two measures φ_1 and φ_2, such that (X_n) is φ_i-irreducible, the corresponding ψ_i's given by (6.9) are equivalent.

6.12 Let $Y_1, Y_2, \ldots$ be iid rv's concentrated on $\mathbb{N}_+$ and Y_0 be another rv also concentrated on $\mathbb{N}_+$. Define

$$Z_n = \sum_{i=0}^{n} Y_i.$$

(a) Show that (Z_n) is a Markov chain. Is it irreducible?
(b) Define the forward recurrence time as

$$V_n^+ = \inf\{Z_m - n; Z_m > n\}.$$

Show that (V_n^+) is also a Markov chain.
(c) If $V_n^+ = k > 1$, show that $V_{n+1}^+ = k - 1$. If $V_n^+ = 1$, show that a renewal occurs at $n+1$. (*Hint:* Show that $V_{n+1}^+ \sim Y_i$ in the latter case.)

6.13 Detail the proof of Theorem 6.15. In particular, show that the fact that K_ϵ includes a Dirac mass does not invalidate the irreducibility. (*Hint*: Establish that

$$\mathbb{E}_x[\eta_A] = \sum_n P_x^n(A) > P_x(\tau_A < \infty),$$

$$\lim_{\epsilon \to 1} K_\epsilon(x, A) > P_x(\tau_A < \infty),$$

$$K_\epsilon(x, A) = (1-\epsilon) \sum_{i=1}^{\infty} \epsilon^i P^i(x, A) > 0$$

imply that there exists n such that $K^n(x, A) > 0$. See Meyn and Tweedie 1993, p. 87.)

(6.36) $$\sqrt{R}\left(\bar{h}_{\tau_R} - \mathbb{E}^\pi[h(X)]\right) \overset{\mathcal{L}}{\rightsquigarrow} \mathcal{N}(0, \sigma_h^2),$$

where

$$\sigma_h^2 = \frac{\mathbb{E}^Q\left[(\tilde{S}_1 - N_1\mathbb{E}^\pi[h(X)])^2\right]}{\{\mathbb{E}^Q[N_1]\}^2}.$$

While it seems that (6.33) and (6.36) are very similar, the advantage in using this approach is that σ_h^2 can be estimated much more easily due to the underlying independent structure. For instance,

$$\hat{\sigma}_h^2 = \frac{\sum_{t=1}^R (\tilde{S}_t - \bar{h}_{\tau_R} N_t)^2}{R\bar{N}^2}$$

is a consistent estimator of σ_h^2.

In addition, the conditions on $\mathbb{E}^Q[\tilde{S}_1^2]$ and $\mathbb{E}^Q[N_1^2]$ appearing in Theorem 6.68 are minimal in that they hold when the conditions of Theorem 6.67 hold (see Hobert et al. 2002, for a proof).

6.8 Problems

6.1 Examine whether a Markov chain (X_t) may always be represented by the deterministic transform $X_{t+1} = \psi(X_t, \epsilon_t)$, where (ϵ_t) is a sequence of iid rv's. (*Hint:* Consider that ϵ_t can be of infinite dimension.)

6.2 Show that if (X_n) is a time-homogeneous Markov chain, the transition kernel does not depend on n. In particular, if the Markov chain has a finite state-space, the transition matrix is constant.

6.3 Show that an ARMA(p, q) model, defined by

$$X_n = \sum_{i=1}^p \alpha_i X_{n-i} + \sum_{j=1}^q \beta_j \varepsilon_{n-j} + \varepsilon_n,$$

does not produce a Markov chain. (*Hint:* Examine the relation with an AR(q) process through the decomposition

$$Z_n = \sum_{i=1}^p \alpha_i Z_{n-i} + \varepsilon_n, \qquad Y_n = \sum_{j=1}^q \beta_j Z_{n-j} + Z_n,$$

since (Y_n) and (X_n) are then identically distributed.)

6.4 Show that the resolvant kernel of Definition 6.8 is truly a kernel.

6.5 Show that the properties of the resolvant kernel are preserved if the geometric distribution $\mathcal{G}eo(\epsilon)$ is replaced by a Poisson distribution $\mathcal{P}(\lambda)$ with arbitrary parameter λ.

6.6 Derive the strong Markov property from the decomposition

$$\begin{aligned}&\mathbb{E}_\mu[h(X_{\zeta+1}, X_{\zeta+2}, \ldots)|x_\zeta, x_{\zeta-1}, \ldots] \\ &= \sum_{n=1}^\infty \mathbb{E}_\mu[h(X_{n+1}, X_{n+2}, \ldots)|x_n, x_{n-1}, \ldots, \zeta = n] P(\zeta = n|x_n, x_{n-1}, \ldots)\end{aligned}$$

and from the weak Markov property.

(6.32) $$\mathbb{E}^\pi[|h(X)|^{2+\varepsilon}] < \infty$$

for some $\varepsilon > 0$, *then*

(6.33) $$\sqrt{n}\,(S_n(h)/n - \mathbb{E}^\pi[h(X)]) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma_h^2)\,,$$

where γ_h^2 *is defined as in Theorem 6.65.*

They first discuss the difficulty in finding such estimates, as fixed batch mean approximations are *not* consistent when the batch size is fixed. We can, however, use regeneration (Mykland et al. 1995) when available; that is, when a minorization condition as in Sections 6.3.2 and 6.5.2 holds: there exists a function $0 < s(x) < 1$ and a probability measure Q such that, for all $x \in \mathcal{X}$ and all measurable sets A,

(6.34) $$P(x, A) \geq s(x)\, Q(A)\,.$$

Following an idea first developed in Robert (1995a) for MCMC algorithms, Hobert et al. (2002) then construct legitimate asymptotic standard errors bypassing the estimation of γ_g^2.

The approach is to introduce the regeneration times $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$ associated with the Markov chain (X_t) and to write $S_n(h)$ in terms of the regeneration times, namely, if the chain is started as $X_0 \sim Q$ and stopped after the T-th regeneration,

$$S_{\tau_T}(h) = \sum_{t=1}^{T} \sum_{j=\tau_{t-1}}^{\tau_t - 1} h(X_j) = \sum_{t=1}^{T} \tilde{S}_t\,,$$

where the $\tilde{S}_t$'s are the partial sums appearing in Theorem 6.63, which are iid. If we define the inter-regeneration lengths $N_t = \tau_t - \tau_{t-1}$, then

(6.35) $$\bar{h}_{\tau_T} = \frac{\sum_{t=1}^{T} \tilde{S}_t}{\sum_{t=1}^{T} N_t} = \frac{\bar{S}_T}{\bar{N}_T} = \frac{1}{\tau_T} \sum_{j=0}^{\tau_T - 1} g(X_j)$$

converges almost surely to $\mathbb{E}^\pi[h(X)]$ when T goes to infinity, by virtue of the Ergodic Theorem (Theorem 6.63), since τ_T converges almost surely to ∞.

By Theorem 6.37, $\mathbb{E}^Q[N_1] = 1/\mathbb{E}^\pi[s(X)]$ (which is assumed to be finite). It follows from the Strong Law of Large Numbers that $\bar{N}$ converges almost surely to $\mathbb{E}^Q[N_1]$, which together with (6.35) implies that $\bar{S}_T$ converges almost surely to $\mathbb{E}^Q[N_1]\,\mathbb{E}^\pi[h(X)]$. This implies in particular that $\mathbb{E}^Q[|\tilde{S}_1|] < \infty$ and $\mathbb{E}^Q[\tilde{S}_1] = \mathbb{E}^Q[N_1]\mathbb{E}^\pi[h(X)]$. Hence, the random variables $\tilde{S}_t - N_t\,\mathbb{E}^\pi[h(X)]$, are iid and centered. Thus

Theorem 6.68. *If* $\mathbb{E}^Q[\tilde{S}_1^2]$ *and* $\mathbb{E}^Q[N_1^2]$ *are both finite, the Central Limit Theorem applies:*

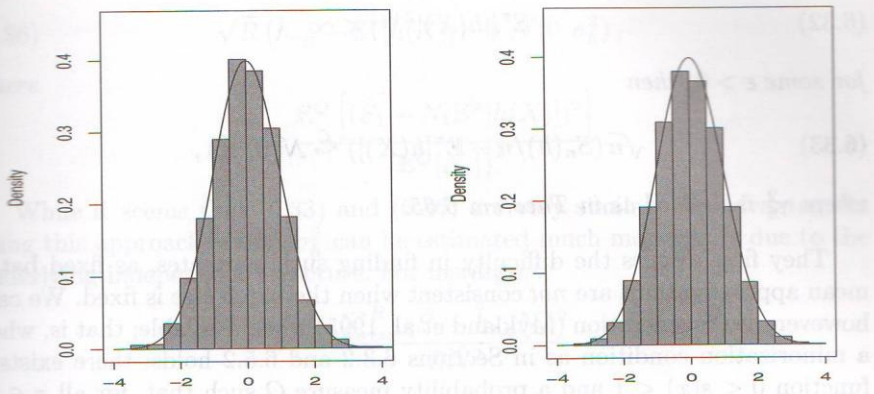

Fig. 6.2. Histogram of 2500 means (each based on 50 observations) from an AR(1) chain. The left panel corresponds to $\theta = .5$, which results in an ergodic chain. The right panel corresponds to $\theta = 2$, which corresponds to a transient chain.

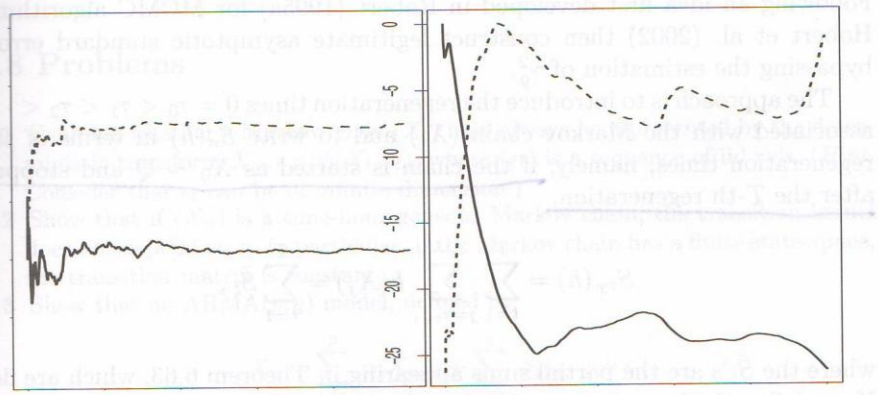

Fig. 6.3. Trajectories of mean (solid line) and standard deviation (dashed line) from the AR(1) process of Example 6.66. The left panel has $\theta = .5$, resulting in an ergodic Markov chain, and displays convergence of the mean and standard deviation. The right panel has $\theta = 1.0001$, resulting in a transient Markov chain and no convergence.

6.7.2.3 Geometric Ergodicity and Regeneration

There is yet another approach to the Central Limit Theorem for Markov chains. It relies on geometric ergodicity, a Liapounov-type moment condition on the function h, and a regeneration argument. Hobert et al. (2002), extending work of Chan and Geyer (1994) (see Problem 6.66), give specific conditions for Theorem 6.67 to apply, namely for Liapounov condition to apply and a consistent estimate of γ_h^2 to be found.

Theorem 6.67. *If (X_n) is aperiodic, irreducible, positive Harris recurrent with invariant distribution π and geometrically ergodic, and if, in addition,*

6.7.2.2 Reversibility

The following theorem avoids the verification of a drift condition, but rather requires the Markov chain to be reversible (see Definition 6.44).

With the assumption of reversibility, this Central Limit Theorem directly follows from the strict positivity of γ_g. This was established by Kipnis and Varadhan (1986) using a proof that is beyond our reach.

Theorem 6.65. *If (X_n) is aperiodic, irreducible, and reversible with invariant distribution π, the Central Limit Theorem applies when*

$$0 < \gamma_g^2 = \mathbb{E}_\pi[\overline{g}^2(X_0)] + 2 \sum_{k=1}^{\infty} \mathbb{E}_\pi[\overline{g}(X_0)\overline{g}(X_k)] < +\infty.$$

The main point here is that even though reversibility is a very restrictive assumption in general, it is often easy to impose in Markov chain Monte Carlo algorithms by introducing additional simulation steps (see Geyer 1992, Tierney 1994, Green 1995). See also Theorem 6.77 for another version of the Central Limit Theorem, which relies on a "drift condition" (see Note 6.9.1) similar to geometric ergodicity.

Example 6.66 (Continuation of Example 6.43). For the AR(1) chain, the transition kernel corresponds to the $\mathcal{N}(\theta x_{n-1}, \sigma^2)$ distribution, and the stationary distribution is $\mathcal{N}(0, \sigma^2/(1-\theta^2))$. It is straightforward to verify that the chain is reversible by showing that (Problem 6.65)

$$X_{n+1}|X_n \sim \mathcal{N}(\theta x_n, \sigma^2) \text{ and } X_n|X_{n+1} \sim \mathcal{N}(\theta x_{n+1}, \sigma^2).$$

Thus the chain satisfies the conditions for the CLT.

Figure 6.2 shows histograms of means for the cases of $\theta = .5$ and $\theta = 2$. In the first case (left panel) we have a positive recurrent chain that satisfies the conditions of the CLT. The right panel is most interesting, however, because $\theta = 2$ and the chain is transient. However, the histogram of the means "looks" quite well behaved, giving no sign that the chain is not converging.

It can happen that null recurrent and transient chains can often look well behaved when examined graphically through some output. However, another picture shows a different story. In Figure 6.3 we look at the trajectories of the cumulative mean and standard deviation from one chain of length 1000. There, the left panel corresponds to the ergodic case with $\theta = .5$, and the right panel corresponds to the (barely) transient case of $\theta = 1.0001$. However, it is clear that there is no convergence. See Section 10.4.3 for the manifestation of this in MCMC algorithms. ‖

6.7.2.1 The Discrete Case

The discrete case can be solved directly, as shown by Problems 6.50 and 6.51.

Theorem 6.64. *If (X_n) is Harris positive with an atom α such that*

$$\mathbb{E}_\alpha[\tau_\alpha^2] < \infty, \qquad \mathbb{E}_\alpha\left[\left(\sum_{n=1}^{\tau_\alpha} |h(X_n)|\right)^2\right] < \infty \tag{6.31}$$

and

$$\gamma_h^2 = \pi(\alpha)\,\mathbb{E}_\alpha\left[\left(\sum_{n=1}^{\tau_\alpha} \{h(X_n) - \mathbb{E}^\pi[h]\}\right)^2\right] > 0,$$

the Central Limit Theorem applies; that is,

$$\frac{1}{\sqrt{N}}\left(\sum_{n=1}^{N} (h(X_n) - \mathbb{E}^\pi[h])\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma_h^2)\,.$$

Proof. Using the same notation as in the proof of Theorem 6.63, if $\overline{h}$ denotes $h - \mathbb{E}^\pi[h]$, we get

$$\frac{1}{\sqrt{\ell_N}} \sum_{i=1}^{\ell_N} S_i(\overline{h}) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathbb{E}_\alpha\left[\sum_{n=1}^{\tau_\alpha} \overline{h}(X_n)\right]^2\right),$$

following from the Central Limit Theorem for the independent variables $S_i(\overline{f})$, while N/ℓ_N converges a.s. to $\mathbb{E}_\alpha[S_0(1)] = 1/\pi(\alpha)$. Since

$$\left|\sum_{i=1}^{\ell_N - 1} S_i(\overline{h}) - \sum_{k=1}^{N} \overline{h}(X_k)\right| \le S_{\ell_N}(|\overline{h}|)$$

and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} S_i(|\overline{h}|)^2 = \mathbb{E}_\alpha[S_0(|\overline{h}|)^2]\,,$$

we get

$$\limsup_{N\to\infty} \frac{S_{\ell_N}(|\overline{h}|)}{\sqrt{N}} = 0,$$

and the remainder goes to 0 almost surely. □

This result indicates that an extension of the Central Limit Theorem to the nonatomic case will be more delicate than for the Ergodic Theorem: Conditions (6.31) are indeed expressed in terms of the split chain $(\check{X}_n)$. (See Section 12.2.3 for an extension to cases when there exists a small set.) In Note 6.9.1, we present some alternative versions of the Central Limit Theorem involving a drift condition.

Proof. If (i) holds, take f to be the indicator function of a set A with finite measure and g an arbitrary function with finite and positive integral. If $\pi(A) > 0$,

$$P_x(X \in A \text{ infinitely often}) = 1$$

for every $x \in \mathcal{X}$, which establishes Harris recurrence.

If (ii) holds, we need only to consider the atomic case by a splitting argument. Let α be an atom and $\tau_\alpha(k)$ be the time of the $(k+1)$th visit to α. If ℓ_N is the number of visits to α at time N, we get the bounds

$$\begin{aligned}\sum_{j=0}^{\ell_N-1} \sum_{n=\tau_\alpha(j)+1}^{\tau_\alpha(j+1)} f(x_n) &\le \sum_{k=1}^{N} f(x_k) \\ &\le \sum_{j=0}^{\ell_N} \sum_{n=\tau_\alpha(j)+1}^{\tau_\alpha(j+1)} f(x_n) + \sum_{k=1}^{\tau_\alpha(0)} f(x_k)\,.\end{aligned}$$

The blocks

$$S_j(f) = \sum_{n=\tau_\alpha(j)+1}^{\tau_\alpha(j+1)} f(x_n)$$

are independent and identically distributed. Therefore,

$$\frac{\sum_{i=1}^{n} f(x_i)}{\sum_{i=1}^{n} g(x_i)} \le \frac{\ell_N}{\ell_N - 1} \frac{(\sum_{j=0}^{\ell_N} S_j(f) + \sum_{k=1}^{\tau_\alpha} f(x_k))/\ell_N}{\sum_{j=0}^{\ell_N-1} s_j(g)/(\ell_N - 1)}\,.$$

The theorem then follows by an application of the strong Law of Large Numbers for iid rv's. □

An important aspect of Theorem 6.63 is that π does not need to be a probability measure and, therefore, that there can be some type of strong stability even if the chain is null recurrent. In the setup of a Markov chain Monte Carlo algorithm, this result is sometimes invoked to justify the use of improper posterior measures, although we fail to see the relevance of this kind of argument (see Section 10.4.3).

6.7.2 Central Limit Theorems

There is a natural progression from the Law of Large Numbers to the Central Limit Theorem. Moreover, the proof of Theorem 6.63 suggests that there is a direct extension of the Central Limit Theorem for iid variables. Unfortunately this is not the case, as conditions on the finiteness of the variance explicitly involve the atom α of the split chain. Therefore, we provide alternative conditions for the Central Limit Theorem to apply in different settings.

Proposition 6.61 can be interpreted as a continuity property of the transition functional $Kh(x) = \mathbb{E}_x[h(X_1)]$ in the following sense. By induction, a harmonic function h satisfies $h(x) = \mathbb{E}_x[h(X_n)]$ and by virtue of Theorem 6.53, $h(x)$ is almost surely equal to $\mathbb{E}^\pi[h(X)]$; that is, it is constant *almost everywhere*. For Harris recurrent chains, Proposition 6.61 states that this implies $h(x)$ is constant *everywhere*. (Feller 1971, pp. 265–267, develops a related approach to ergodicity, where Harris recurrence is replaced by a regularity constraint on the kernel.)

Proposition 6.61 will be most useful in establishing Harris recurrence of some Markov chain Monte Carlo algorithms. Interestingly, the behavior of bounded harmonic functions characterizes Harris recurrence, as the converse of Proposition 6.61 is true. We state it without its rather difficult proof (see Meyn and Tweedie 1993, p. 415).

Lemma 6.62. *For Harris recurrent Markov chains, the constants are the only bounded harmonic functions.*

A consequence of Lemma 6.62 is that if (X_n) is Harris positive with stationary distribution π and if $S_n(h)$ converges μ_0-almost surely (μ_0 a.s.) to

$$\int_{\mathcal{X}} h(x)\, \pi(dx)\ ,$$

for an initial distribution μ_0, this convergence occurs for every initial distribution μ. Indeed, the convergence probability

$$P_x\left(S_N(h) \to \mathbb{E}^\pi[h]\right)$$

is then harmonic. Once again, this shows that Harris recurrence is a superior type of stability in the sense that *almost sure convergence* is replaced by convergence at every point.

Of course, we now know that if functions other than bounded functions are harmonic, the chain is not Harris recurrent. This is looked at in detail in Problem 6.59.

The main result of this section, namely the Law of Large Numbers for Markov chains (which is customarily called the *Ergodic Theorem*), guarantees the convergence of $S_n(h)$.

Theorem 6.63. Ergodic Theorem *If (X_n) has a σ-finite invariant measure π, the following two statements are equivalent:*

(i) If $f, g \in L^1(\pi)$ with $\int g(x)d\pi(x) \neq 0$, then

$$\lim_{n\to\infty} \frac{S_n(f)}{S_n(g)} = \frac{\int f(x)d\pi(x)}{\int g(x)d\pi(x)}\ .$$

(ii) The Markov chain (X_n) is Harris recurrent.

6.7.1 Ergodic Theorems

Given observations $X_1, \ldots, X_n$ of a Markov chain, we now examine the limiting behavior of the partial sums

$$S_n(h) = \frac{1}{n} \sum_{i=1}^{n} h(X_i)$$

when n goes to infinity, getting back to the iid case through renewal when (X_n) has an atom. Consider first the notion of *harmonic functions*, which is related to ergodicity for Harris recurrent Markov chains.

Definition 6.60. A measurable function h is *harmonic* for the chain (X_n) if

$$\mathbb{E}[h(X_{n+1})|x_n] = h(x_n).$$

These functions are *invariant* for the transition kernel (in the functional sense) and they characterize Harris recurrence as follows.

Proposition 6.61. *For a positive Markov chain, if the only bounded harmonic functions are the constant functions, the chain is Harris recurrent.*

Proof. First, the probability of an infinite number of returns, $Q(x, A) = P_x(\eta_A = \infty)$, as a function of x, $h(x)$, is clearly a harmonic function. This is because

$$\mathbb{E}_y[h(X_1)] = \mathbb{E}_y[P_{X_1}(\eta_A = \infty)] = P_y(\eta_A = \infty),$$

and thus, $Q(x, A)$ is constant (in x).

The function $Q(x, A)$ describes a *tail event*, an event whose occurrence does not depend on $X_1, X_2, \ldots, X_m$, for any finite m. Such events generally obey a $0 - 1$ law, that is, their probabilities of occurrence are either 0 or 1. However, $0 - 1$ laws are typically established in the independence case, and, unfortunately, extensions to cover Markov chains are beyond our scope. (For example, see the *Hewitt–Savage* $0 - 1$ *Law*, in Billingsley 1995, Section 36.) For the sake of our proof, we will just state that $Q(x, A)$ obeys a $0 - 1$ Law and proceed.

If π is the invariant measure and $\pi(A) > 0$, the case $Q(x, A) = 0$ is impossible. To see this, suppose that $Q(x, A) = 0$. It then follows that the chain almost surely visits A only a finite number of times and the average

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_A(X_i)$$

will not converge to $\pi(A)$, contradicting the Law of Large Numbers (see Theorem 6.63). Thus, for any x, $Q(x, A) = 1$, establishing that the chain is a Harris chain. □

chapters we will see that this last section is essential to the processing of these algorithms. In fact, the different convergence results (ergodicity) obtained in Section 6.6 deal only with the probability measure P_x^n (through different norms), which is somewhat of a "snapshot" of the chain (X_n) at time n. So, it determines the probabilistic properties of *average* behavior of the chain at a fixed instant. Such properties, even though they provide justification for the simulation methods, are of lesser importance for the control of convergence of a given simulation, where the properties of the *realization* (x_n) of the chain are the only characteristics that truly matter. (Meyn and Tweedie 1993 call this type of properties "sample path" properties.)

We are thus led back to some basic ideas, previously discussed in a statistical setup by Robert (2001, Chapters 1 and 11); that is, we must consider the difference between *probabilistic analysis*, which describes the average behavior of samples, and *statistical inference*, which must reason by induction from the observed sample. While probabilistic properties can justify or refute some statistical approaches, this does not contradict the fact that statistical analysis must be done *conditional on the observed sample*. Such a consideration can lead to the Bayesian approach in a statistical setup (or at least to consideration of the *Likelihood Principle*; see, e.g., Berger and Wolpert 1988, or Robert 2001, Section 1.3). In the setup of Markov chains, a conditional analysis can take advantage of convergence properties of P_x^n to π only to verify the convergence, to a quantity of interest, of functions of the observed path of the chain. Indeed, the fact that $\|P_x^n - \pi\|$ is close to 0, or even converges geometrically fast to 0 with speed ρ^n $(0 < \rho < 1)$, does not bring direct information about the unique available observation from P_x^n, namely X_n.

The problems in directly applying the classical convergence theorems (Law of Large Numbers, Law of the Iterated Logarithm, Central Limit Theorem, etc.) to the sample $(X_1, \ldots, X_n)$ are due both to the Markovian dependence structure between the observations X_i and to the non-stationarity of the sequence. (Only if $X_0 \sim \pi$, the stationary distribution of the chain, will the chain be stationary. Since this is equivalent to integrating over the initial conditions, it eliminates the need for a conditional analysis. Such an occurrence, especially in Markov chain Monte Carlo, is somewhat rare.[4])

We therefore assume that the chain is started from a point X_0 whose distribution is not the stationary distribution of the chain, and thus we deal with non-stationary chains directly. We begin with a detailed presentation of convergence results equivalent to the Law of Large Numbers, which are often called *ergodic theorems*. We then mention in Section 6.7.2 various versions of the Central Limit Theorem whose assumptions are usually (and unfortunately) difficult to check.

[4] Nonetheless, there is considerable research in MCMC theory about *perfect simulation*; that is, ways of starting the algorithm with $X_0 \sim \pi$. See Chapter 13.

Definition 6.58. The chain (X_n) is *uniformly ergodic* if

$$\lim_{n\to\infty} \sup_{x\in\mathcal{X}} \|K^n(x,\cdot) - \pi\|_{TV} = 0\,.$$

Uniform ergodicity can be established through one of the following equivalent properties:

Theorem 6.59. *The following conditions are equivalent:*

(a) (X_n) *is uniformly ergodic;*
(b) there exist $R < \infty$ *and* $r > 1$ *such that*

$$\|K^n(x,\cdot) - \pi\|_{TV} < Rr^{-n}\,, \qquad \text{for all } x \in \mathcal{X}\,;$$

(c) (X_n) *is aperiodic and* $\mathcal{X}$ *is a small set;*
(d) (X_n) *is aperiodic and there exist a small set* C *and a real* $\kappa > 1$ *such that*

$$\sup_{x\in\mathcal{X}} \mathbb{E}_x[\kappa^{\tau_C}] < \infty\,.$$

If the whole space $\mathcal{X}$ is small, there exist a probability distribution, φ, on $\mathcal{X}$, and constants $\varepsilon < 1$, $\delta > 0$, and n such that, if $\varphi(A) > \varepsilon$ then

$$\inf_{x\in\mathcal{X}} K^n(x,A) > \delta\,.$$

This property is sometimes called *Doeblin's condition*. This requirement shows the strength of the uniform ergodicity and suggests difficulties about the verification. We will still see examples of Markov chain Monte Carlo algorithms which achieve this superior form of ergodicity (see Example 10.17). Note, moreover, that in the finite case, uniform ergodicity can be derived from the smallness of $\mathcal{X}$ since the condition

$$P(X_{n+1} = y | X_n = x) \geq \inf_z p_{zy} = \rho_y \qquad \text{for every } x, y \in \mathcal{X},$$

leads to the choice of the minorizing measure ν as

$$\nu(y) = \frac{\rho_y}{\sum_{z\in\mathcal{X}} \rho_z}$$

as long as $\rho_y > 0$ for some $y \in \mathcal{X}$. (If (X_n) is recurrent and aperiodic, this positivity condition can be attained by a subchain $(Y_m) = (X_{nd})$ for d large enough. See Meyn and Tweedie 1993, Chapter 16, for more details.)

6.7 Limit Theorems

Although the notions and results introduced in the previous sections are important in justifying Markov chain Monte Carlo algorithms, in the following

Definition 6.55. An accessible atom α is *geometrically ergodic* if there exists $r > 1$ such that

$$\sum_{n=1}^{\infty} |K^n(\alpha, \alpha) - \pi(\alpha)|\, r^n < \infty$$

and α is a *Kendall atom* if there exists $\kappa > 1$ such that

$$\mathbb{E}_\alpha[\kappa^{\tau_\alpha}] < \infty\,.$$

If α is a Kendall atom, it is thus geometrically ergodic and ensures geometric ergodicity for (X_n):

Theorem 6.56. *If (X_n) is ψ-irreducible, with invariant distribution π, and if there exists a geometrically ergodic atom α, then there exist $r > 1$, $\kappa > 1$, and $R < \infty$ such that, for almost every $x \in \mathcal{X}$,*

$$\sum_{n=1}^{\infty} r^n \|K^n(x, \cdot) - \pi\|_{TV} < R\, \mathbb{E}_x\left[\kappa^{\tau_\alpha}\right] < \infty\,.$$

Example 6.57. Nongeometric returns to 0. For a chain on $\mathbb{Z}_+$ with transition matrix $\mathbb{P} = (p_{ij})$ such that

$$p_{0j} = \gamma_j, \quad p_{jj} = \beta_j, \quad p_{j0} = 1 - \beta_j, \quad \sum_j \gamma_j = 1,$$

Meyn and Tweedie (1993, p. 361) consider the return time to 0, τ_0, with mean

$$\begin{aligned}\mathbb{E}_0[\tau_0] &= \sum_j \gamma_j \left\{(1-\beta_j) + 2\beta_j(1-\beta_j) + \cdots\right\} \\ &= \sum_j \gamma_j \left\{1 + (1-\beta_j)^{-1}\right\}.\end{aligned}$$

The state 0 is thus an ergodic atom when all the γ_j's are positive (yielding irreducibility) and $\sum_j \gamma_j(1-\beta_j)^{-1} < \infty$. Now, for $r > 0$,

$$\mathbb{E}_0[r^{\tau_0}] = r \sum_j \gamma_j \mathbb{E}_j[r^{\tau_0 - 1}] = r \sum_j \gamma_j \sum_{k=0}^{\infty} r^k \beta_j^k (1-\beta_j)\,.$$

For $r > 1$, if $\beta_j \to 1$ as $j \to \infty$, the series in the above expectation always diverges for j large enough. Thus, the chain is not geometrically ergodic. ‖

6.6.3 Uniform Ergodicity

The property of uniform ergodicity is stronger than geometric ergodicity in the sense that the rate of geometric convergence must be uniform over the whole space. It is used in the Central Limit Theorem given in Section 6.7.

6.6.2 Geometric Convergence

The convergence (6.28) of the expectation of $h(x)$ at time n to the expectation of $h(x)$ under the stationary distribution π somehow ensures the proper behavior of the chain (X_n) whatever the initial value X_0 (or its distribution). A more precise description of convergence properties involves the study of *the speed of convergence* of K^n to π. An evaluation of this speed is important for Markov chain Monte Carlo algorithms in the sense that it relates to stopping rules for these algorithms; minimal convergence speed is also a requirement for the application of the Central Limit Theorem.

To study the speed of convergence more closely, we first introduce an extension of the total variation norm, denoted by $\|\cdot\|_h$, which allows for an upper bound other than 1 on the functions. The generalization is defined by

$$\|\mu\|_h = \sup_{|g|\le h} \left| \int g(x)\mu(dx) \right| .$$

Definition 6.54. A chain (X_n) is *geometrically h-ergodic*, with $h \ge 1$ on $\mathcal{X}$, if (X_n) is Harris positive, with stationary distribution π, if (X_n) satisfies $\mathbb{E}^\pi[h] < \infty$, and if there exists $r_h > 1$ such that

$$\sum_{n=1}^{\infty} r_h^n \|K^n(x,\cdot) - \pi\|_h < \infty \tag{6.30}$$

for every $x \in \mathcal{X}$. The case $h = 1$ corresponds to the *geometric ergodicity* of (X_n).

Geometric h-ergodicity means that $\|K^n(x,\cdot) - \pi\|_h$ is decreasing at least at a geometric speed, since (6.30) implies

$$\|K^n(x,\cdot) - \pi\|_h \le M r_h^{-n}$$

with

$$M = \sum_{n=1}^{\infty} r_h^n \|K^n(x,\cdot) - \pi\|_h .$$

If (X_n) has an atom α, (6.30) implies that for a real number $r > 1$,

$$\mathbb{E}_x\left[\sum_{n=1}^{\tau_\alpha} h(X_n) r^n\right] < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} |P_\alpha(X_n \in \alpha) - \pi(\alpha)| r^n < \infty.$$

The series associated with $|P_\alpha(X_n \in \alpha) - \pi(\alpha)|\ r^n$ converges outside of the unit circle if the power series associated with $P_\alpha(\tau_\alpha = n)$ converges for values of $|r|$ strictly larger than 1. (The proof of this result, called *Kendall's Theorem*, is based on the renewal equations established in the proof of Proposition 6.31.) This equivalence justifies the following definition.

since, by definition, $K^{n+1}(x,dy) = \int K^n(x,dw)K(w,dy)$ and by the invariance of π, $\pi(dy) = \int K(w,dy)\pi(dw)$. Regrouping terms, we can write,

$$\begin{aligned} 2\left\| \int K^{n+1}(x,\cdot)\mu(dx) - \pi \right\|_{TV} &= \sup_{|h|\le 1} \left| \int \left[\int h(y)K(w,dy) \right] K^n(x,dw)\mu(dx) \right. \\ &\quad \left. - \int \left[\int h(y)K(w,dy) \right] \pi(dw) \right| \\ &\le \sup_{|h|\le 1} \left| \int h(w)K^n(x,dw)\mu(dx) - \int h(w)\pi(dw) \right|, \end{aligned}$$

where the inequality follows from the fact that the quantity in square brackets is a function with norm less than 1. Hence, monotonicity of the total variation norm is established. □

Note that the equivalence (6.26) also implies the convergence

$$\lim_{n\to\infty} |\mathbb{E}_\mu[h(X_n)] - \mathbb{E}^\pi[h(X)]| = 0 \tag{6.27}$$

for every *bounded function* h. This equivalence is, in fact, often taken as the defining condition for convergence of distributions (see, for example, Billingsley 1995, Theorem 25.8). We can, however, conclude (6.27) from a slightly weaker set of assumptions, where we do not need the full force of Harris recurrence (see Theorem 6.80 for an example).

The extension of (6.27) to more general functions h is called *h-ergodicity* by Meyn and Tweedie (1993, pp. 342–344).

Theorem 6.53. *Let (X_n) be positive, recurrent, and aperiodic.*

(a) If $\mathbb{E}^\pi[|h(X)|] = \infty$, $\mathbb{E}_x[|h(X_n)|] \to \infty$ for every x.
(b) If $\int |h(x)|\pi(dx) < \infty$, then

$$\lim_{n\to\infty} \sup_{|m(x)|\le|h(x)|} |\mathbb{E}_y[m(X_n)] - \mathbb{E}^\pi[m(X)]| = 0 \tag{6.28}$$

on all small sets C such that

$$\sup_{y\in C} \mathbb{E}_y\left[\sum_{t=0}^{\tau_C - 1} h(X_t) \right] < \infty\,. \tag{6.29}$$

Similar conditions appear as necessary conditions for the Central Limit Theorem (see (6.31) in Theorem 6.64). Condition (6.29) relates to a coupling argument, in the sense that the influence of the initial condition vanishes "fast enough," as in the proof of Theorem 6.63.

Theorem 6.50. *For a positive recurrent aperiodic Markov chain on a countable space, for every initial state x,*

$$\lim_{n\to\infty} \|K^n(x,\cdot) - \pi\|_{TV} = 0\,.$$

Proof. Since (X_n) is positive recurrent, $\mathbb{E}_\alpha[\tau_\alpha]$ is finite by Theorem 6.37. Therefore, m_p is finite, (6.25) holds, and every atom is ergodic. The result follows from Proposition 6.48. □

For general state-spaces $\mathcal{X}$, Harris recurrence is nonetheless necessary in the derivation of the convergence of K^n to π. (Note that another characterization of Harris recurrence is the convergence of $\|K_x^n - \pi\|_{TV}$ to 0 for *every* value x, instead of almost every value.)

Theorem 6.51. *If (X_n) is Harris positive and aperiodic, then*

$$\lim_{n\to\infty} \left\| \int K^n(x,\cdot)\mu(dx) - \pi \right\|_{TV} = 0$$

for every initial distribution μ.

This result follows from an extension of the denumerable case to strongly aperiodic Harris positive chains by splitting, since these chains always allow for small sets (see Section 6.3.3), based on an equivalent to the "first entrance and last exit" formula (6.23). It is then possible to move to arbitrary chains by the following result.

Proposition 6.52. *If π is an invariant distribution for P, then*

$$\left\| \int K^n(x,\cdot)\mu(dx) - \pi \right\|_{TV}$$

is decreasing in n.

Proof. First, note the equivalent definition of the norm (Problem 6.40)

$$\|\mu\|_{TV} = \frac{1}{2} \sup_{|h|\le 1} \left| \int h(x)\mu(dx) \right|. \tag{6.26}$$

We then have

$$\begin{aligned} 2 &\left\| \int K^{n+1}(x,\cdot)\mu(dx) - \pi \right\|_{TV} \\ &= \sup_{|h|\le 1} \left| \int h(y) K^{n+1}(x,dy)\mu(dx) - \int h(y)\pi(dy) \right| \\ &= \sup_{|h|\le 1} \left| \int h(y) \int K^n(x,dw) K(w,dy)\mu(dx) \right. \\ &\qquad \left. - \int h(y) \int K(w,dy)\pi(dw) \right|, \end{aligned}$$

time it takes for the two chains to meet), is finite for almost every starting point, the ergodicity of the chain follows.

For a recurrent atom α on a denumerable space $\mathcal{X}$, let $\tau_\alpha(k)$ denote the kth visit to α ($k = 1, 2, \ldots$), and let $p = (p(1), p(2), \cdots)$ be the distribution of the *excursion time*,

$$S_k = \tau_\alpha(k+1) - \tau_\alpha(k),$$

between two visits to α. If $q = (q(0), q(1), \ldots)$ represents the distribution of $\tau_\alpha(1)$ (which depends on the initial condition, x_0 or μ), then the distribution of $\tau_\alpha(n+1)$ is given by the convolution product $q \star p^{n\star}$ (that is, the distribution of the sum of n iid rv's distributed from p and of a variable distributed from q), since

$$\tau_\alpha(n+1) = S_n + \cdots + S_1 + \tau_\alpha(1).$$

Thus, consider two sequences (S_i) and (S_i') such that $S_1, S_2, \ldots$ and $S_1', S_2', \ldots$ are iid from p with $S_0 \sim q$ and $S_0' \sim r$. We introduce the indicator functions

$$Z_q(n) = \sum_{j=0}^{n} \mathbb{I}_{S_1+\cdots+S_j=n} \quad \text{and} \quad Z_r(n) = \sum_{j=0}^{n} \mathbb{I}_{S_1'+\cdots+S_j'=n},$$

which correspond to the events that the chains (X_n) and (X_n') visit α at time n. The *coupling time* is then given by

$$T_{qr} = \min \{j; Z_q(j) = Z_r(j) = 1\},$$

which satisfies the following lemma, whose proof can be found in Problem 6.45.

Lemma 6.49. *If the mean excursion time satisfies*

$$m_p = \sum_{n=0}^{\infty} np(n) < \infty$$

and if p is aperiodic (the g.c.d. of the support of p is 1), then the coupling time T_{pq} is almost surely finite, that is,

$$P(T_{pq} < \infty) = 1,$$

for every q.

If p is aperiodic with finite mean m_p, this implies that Z_p satisfies

$$\lim_{n\to\infty} |P(Z_q(n) = 1) - m_q^{-1}| = 0, \tag{6.25}$$

as shown in Problem 6.42. The probability of visiting α at time n is thus asymptotically independent of the initial distribution and this result implies that Proposition 6.48 holds without imposing constraints in the discrete case.

to 0 with n. The expression (6.17) of the invariant measure implies, in addition, that

$$\pi(y) = \pi(\alpha) \sum_{j=1}^{\infty} P_\alpha(X_j = y, \tau_\alpha \geq j).$$

These two expressions then lead to

$$\begin{aligned} ||K^n(x,\cdot) - \pi||_{TV} &= \sum_y |K^n(x,y) - \pi(y)| \\ &\leq \sum_y P_x(X_n = y, \tau_\alpha \geq n) \\ &\quad + \sum_y \sum_{j=1}^{n-1} \left| \sum_{k=1}^{j} P_x(X_k \in \alpha, \tau_\alpha = k) K^{j-k}(\alpha,\alpha) - \pi(\alpha) \right| \\ &\qquad \times P_\alpha(X_{n-j} = y, \tau_\alpha \geq n-j) \\ &\quad + \sum_y \pi(\alpha) \sum_{j=n-1}^{\infty} P_\alpha(X_j = y, \tau_\alpha \geq j) \,. \end{aligned}$$

The second step in the proof is to show that each term in the above decomposition goes to 0 as n goes to infinity. The first term is actually $P_x(\tau_\alpha \geq n)$ and goes to 0 since the chain is Harris recurrent. The third term is the remainder of the convergent series

$$\sum_y \pi(\alpha) \sum_{j=1}^{\infty} P_\alpha(X_j = y, \tau_\alpha \geq j) = \sum_y \pi(y) \,. \tag{6.24}$$

The middle term is the sum over the y's of the convolution of the two sequences $a_n = |\sum_{k=1}^{n} P_x(X_k \in \alpha, \tau_\alpha = k) K^{n-k}(\alpha,\alpha) - \pi(\alpha)|$ and $b_n = P_\alpha(X_n = y, \tau_\alpha \geq n)$. The sequence (a_n) is converging to 0 since the atom α is ergodic and the series of the b_n's is convergent, as mentioned. An algebraic argument (see Problem 6.39) then implies that (6.24) goes to 0 as n goes to ∞. □

The decomposition (6.23) is quite revealing in that it shows the role of the atom α as the generator of a renewal process. Below, we develop an extension which allows us to deal with the general case using *coupling* techniques. (These techniques are also useful in the assessment of convergence for Markov chain Monte Carlo algorithms.) Lindvall (1992) provides an introduction to coupling.

The *coupling* principle uses two chains (X_n) and (X'_n) associated with the same kernel, the "coupling" event taking place when they meet in α; that is, at the first time n_0 such that $X_{n_0} \in \alpha$ and $X'_{n_0} \in \alpha$. After this instant, the probabilistic properties of (X_n) and (X'_n) are identical and if one of the two chains is stationary, there is no longer any dependence on initial conditions for either chain. Therefore, if we can show that the *coupling time* (that is, the

6.6 Ergodicity and Convergence

6.6.1 Ergodicity

Considering the Markov chain (X_n) from a temporal perspective, it is natural (and important) to establish the limiting behavior of X_n; that is, *To what is the chain converging*? The existence (and uniqueness) of an invariant distribution π makes that distribution a natural candidate for the limiting distribution, and we now turn to finding sufficient conditions on (X_n) for X_n to be asymptotically distributed according to π. The following theorems are fundamental convergence results for Markov chains and they are at the core of the motivation for Markov chain Monte Carlo algorithms. They are, unfortunately, if not surprisingly, quite difficult to establish and we restrict the proof to the countable case, the extension to the general case being detailed in Meyn and Tweedie (1993, pp. 322–323).

There are many conditions that can be placed on the convergence of P^n, the distribution of X_n, to π. Perhaps, the most fundamental and important is that of *ergodicity*, that is, independence of initial conditions.

Definition 6.47. For a Harris positive chain (X_n), with invariant distribution π, an atom α is *ergodic* if

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

In the countable case, the existence of an ergodic atom is, in fact, sufficient to establish convergence according to the *total variation norm*,

$$\|\mu_1 - \mu_2\|_{TV} = \sup_A |\mu_1(A) - \mu_2(A)|.$$

Proposition 6.48. *If (X_n) is Harris positive on $\mathcal{X}$ and denumerable, and if there exists an ergodic atom $\alpha \subset \mathcal{X}$, then, for every $x \in \mathcal{X}$,*

$$\lim_{n\to\infty} \|K^n(x,\cdot) - \pi\|_{TV} = 0\,.$$

Proof. The first step follows from a decomposition formula called *"first entrance and last exit"*:

$$\begin{aligned} K^n(x,y) &= P_x(X_n = y, \tau_\alpha \geq n) \\ &\quad + \sum_{j=1}^{n-1} \left[\sum_{k=1}^{j} P_x(X_k \in \alpha, \tau_\alpha \geq k) K^{j-k}(\alpha,\alpha)\right] \\ &\quad \times P_\alpha(X_{n-j} = y, \tau_\alpha \geq n-j), \end{aligned} \tag{6.23}$$

which relates $K^n(x,y)$ to the last visit to α. (See Problem 6.37.) This shows the reduced influence of the initial value x, since $P_x(X_n = y, \tau_\alpha \geq n)$ converges

does not matter in the dynamics of the chain (see also Problems 6.53 and 6.54).

Definition 6.44. A stationary Markov chain (X_n) is *reversible* if the distribution of X_{n+1} conditionally on $X_{n+2} = x$ is the same as the distribution of X_{n+1} conditionally on $X_n = x$.

In fact, reversibility can be linked with the existence of a stationary measure π if a condition stronger than in Definition 6.35 holds.

Definition 6.45. A Markov chain with transition kernel K satisfies the *detailed balance condition* if there exists a function f satisfying

$$K(y,x)f(y) = K(x,y)f(x) \tag{6.22}$$

for every (x, y).

While this condition is not *necessary* for f to be a stationary measure associated with the transition kernel K, it provides a sufficient condition that is often easy to check and that can be used for most MCMC algorithms. The balance condition (6.22) express an equilibrium in the flow of the Markov chain, namely that the probability of being in x and moving to y is the same as the probability of being in y and moving back to x. When f is a density, it also implies that the chain is reversible.[3] More generally,

Theorem 6.46. *Suppose that a Markov chain with transition function K satisfies the detailed balance condition with π a probability density function. Then*

(i) The density π is the invariant density of the chain.
(ii) The chain is reversible.

Proof. Part (i) follows by noting that, by the detailed balance condition, for any measurable set B,

$$\begin{aligned}\int_{\mathcal{Y}} K(y,B)\pi(y)dy &= \int_{\mathcal{Y}}\int_B K(y,x)\pi(y)dxdy \\ &= \int_{\mathcal{Y}}\int_B K(x,y)\pi(x)dxdy = \int_B \pi(x)dx,\end{aligned}$$

since $\int K(x,y)dy = 1$. With the existence of the kernel K and invariant density π, it is clear that detailed balance and reversibility are the same property. □

If $f(x,y)$ is a joint density, then we can write (with obvious notation)

$$\begin{aligned}f(x,y) &= f_{X|Y}(x|y)f_Y(y) \\ f(x,y) &= f_{Y|X}(y|x)f_X(x),\end{aligned}$$

and thus detailed balance requires that $f_X = f_Y$ and $f_{X|Y} = f_{Y|X}$, that is, there is symmetry in the conditionals and the marginals are the same.

[3] If there are no measure-theoretic difficulties with the definition of the kernel K, both notions are equivalent.

a representation of π that is associated with Kac's Theorem, under the form

$$\pi(A) = \sum_{t=1}^{\infty} \Pr(N_t \in A) \Pr(T^* = t)\,, \tag{6.20}$$

where T^* is an integer valued random variable and, for each $t \in \mathbb{N}$, N_t is a random variable whose distribution depends on t. This representation is very closely related to Kac's (6.17), but offers a wider range of applicability since it does not require the chain to have an atom. We simply assume that the minorizing condition (6.10) is satisfied (and for simplicity's sake we take $m = 1$ in (6.10)). The random variables N_t and T^* can then be defined in terms of the split chain of Section 6.3.2. If $\tau_C \geq 1$ denotes the renewal time associated with the small set C in (6.10), then $\mathbb{E}_\nu[\tau_C] < \infty$ by recurrence (Problem 6.32), and T^* is given by the tail probabilities of τ_C as

$$P(T^* = t) = \frac{\Pr_\nu(\tau_C \geq t)}{\mathbb{E}_\nu(\tau_C)}\,. \tag{6.21}$$

The random variable N_t is then logically distributed from ν if $t = 1$ and as K conditional on no renewal before time t otherwise, following from (6.10). Breyer and Roberts (2000b) derive the representation (6.20) by the mean of a functional equation (see Problem 6.33).

Simulating from π thus amounts to simulating T^* according to (6.21) and then, for $T^* = t$, to simulating N_t. Simulating the latter starts from the minorizing measure ν and then runs $t-1$ steps of the residual distribution

$$\tilde{K}(x,\cdot) = \frac{K(x,\cdot) - \epsilon \mathbb{I}_C(x)\nu(\cdot)}{1 - \epsilon \mathbb{I}_C(x)}\,.$$

In cases when simulating from the residual is too complex, a brute force Accept–Reject implementation is to run the split chain t iterations until $\tau_C \geq t$, but this may be too time-consuming in many situations. Hobert and Robert (2004) also propose more advanced approaches to the simulation of T^*.

Note that, when the state space $\mathcal{X}$ is small, the chain is said to be *uniformly ergodic* (see Definition 6.58 below), $\tilde{K}(x,y) = (K(x,y) - \epsilon\nu(y))/(1-\epsilon)$ and the mixture representation (6.20) translates into the following algorithm.

Algorithm A.23 –Kac's Mixture Implementation–

1. Simulate $X_0 \sim \nu$, $\omega \sim \mathcal{G}eo(\varepsilon)$.
2. Run the transition $X_{t+1} \sim \tilde{K}(x_t,\cdot)$ $t = 0, \cdots, \omega - 1$, and take X_ω.

6.5.3 Reversibility and the Detailed Balance Condition

The stability property inherent to stationary chains can be related to another stability property called *reversibility*, which states that the direction of time

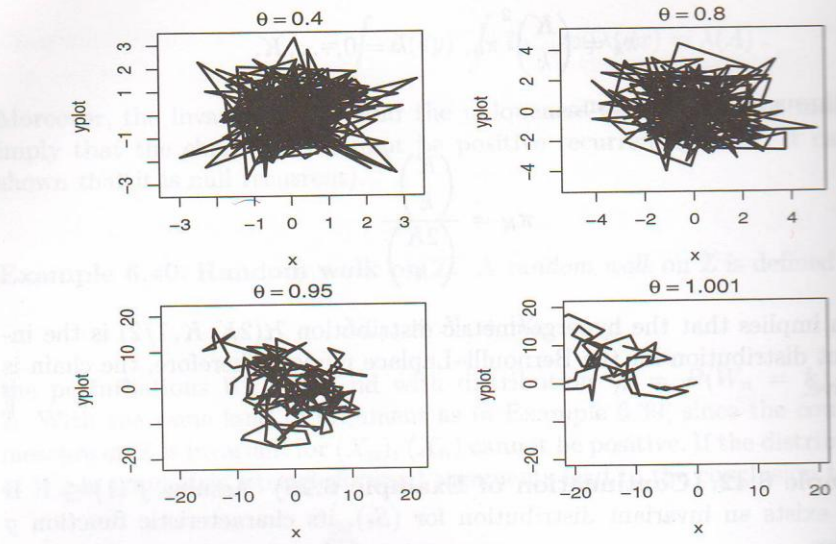

Fig. 6.1. Trajectories of four AR(1) chains with $\sigma = 1$. The first three panels show positive recurrent chains, and a θ increases the chain gets closer to transience. The fourth chain with $\theta = 1.0001$ is transient, and the trajectory never returns.

These conditions imply that $\mu = 0$ and that $\tau^2 = \sigma^2/(1-\theta^2)$, which can only occur for $|\theta| < 1$. In this case, $\mathcal{N}(0, \sigma^2/(1-\theta^2))$ is indeed the unique stationary distribution of the AR(1) chain.

So if $|\theta| < 1$, the marginal distribution of the chain is a proper density independent of n, and the chain is positive (hence recurrent). Figure 6.1 shows the two-dimensional trajectories of an AR(1) chain, where each coordinate is a univariate AR(1) chain. (We use two dimensions to better graphically illustrate the behavior of the chain.)

In the first three panels of Figure 6.1 we see increasing θ, but all three are positive recurrent. This results in the chain "filling" the space; and we can see as θ increases there is less dense filling in. Finally, the fourth chain, with $\theta = 1.001$, is transient, and not only it does not fill the space, but it escapes and never returns. Note the scale on that panel.

When we use a Markov chain to explore a space, we want it to fill the space. Thus, we want our MCMC chains to be positive recurrent.

Note that the converse to Proposition 6.38 does not hold: there exist transient Markov chains with stationary measures. For instance, the random walks in $\mathbb{R}^3$ and $\mathbb{Z}^3$, corresponding to Examples 6.39 and 6.40, respectively, are both transient *and* have the Lebesgue and the counting measures as stationary measures (see Problem 6.25).

In the case of a general Harris recurrent irreducible and aperiodic Markov chain (X_n) with stationary distribution π, Hobert and Robert (2004) propose

$$\pi_k = \binom{K}{k}^2 \pi_0, \qquad k = 0, \ldots, K,$$

and through normalization,

$$\pi_K = \frac{\binom{K}{k}^2}{\binom{2K}{k}},$$

which implies that the hypergeometric distribution $\mathcal{H}(2K, K, 1/2)$ is the invariant distribution for the Bernoulli–Laplace model. Therefore, the chain is positive. ‖

Example 6.42. (Continuation of Example 6.26) Assume $f'(1) \leq 1$. If there exists an invariant distribution for (S_t), its characteristic function g satisfies

$$g(s) = f(s)g(0) + g[f(s)] - g(0)\,. \tag{6.18}$$

In the simplest case, that is to say, when the number of siblings of a given individual is distributed according to a Bernoulli distribution $\mathcal{B}(p)$, $f(s) = q + ps$, where $q = 1 - p$, and $g(s)$ is solution of

$$g(s) = g(q + ps) + p(s-1)g(0)\,. \tag{6.19}$$

Iteratively substituting (6.18) into (6.19), we obtain

$$\begin{aligned} g(s) &= g[q + p(q + ps)] + p(q + ps - 1)g(0) + p(s-1)g(0) \\ &= g(q + pq + \cdots + p^{k-1}q + p^k s) + (p + \cdots + p^k)(s-1)g(0)\,, \end{aligned}$$

for every $k \in \mathbb{N}$. Letting k go to infinity, we have

$$\begin{aligned} g(s) &= g[q/(1-p)] + [p/(1-p)](s-1)g(0) \\ &= 1 + \frac{p}{q}\,(s-1)g(0)\,, \end{aligned}$$

since $q/(1-p) = 1$ and $g(1) = 1$. Substituting $s = 0$ implies $g(0) = q$ and, hence, $g(s) = 1 + p(s-1) = q + ps$. The Bernoulli distribution is thus the invariant distribution and the chain is positive. ‖

Example 6.43. (Continuation of Example 6.20) Given that the transition kernel corresponds to the $\mathcal{N}(\theta x_{n-1}, \sigma^2)$ distribution, a normal distribution $\mathcal{N}(\mu, \tau^2)$ is stationary for the AR(1) chain only if

$$\mu = \theta\mu \qquad \text{and} \qquad \tau^2 = \tau^2\theta^2 + \sigma^2\,.$$

$$= \int \Gamma(dy) \int \mathbb{I}_{A-y}(x)\lambda(dx) = \lambda(A)\,.$$

Moreover, the invariance of λ and the uniqueness of the invariant measure imply that the chain (X_n) cannot be positive recurrent (in fact, it can be shown that it is null recurrent).

Example 6.40. Random walk on $\mathbb{Z}$. A *random walk* on $\mathbb{Z}$ is defined by

$$X_{n+1} = X_n + W_n,$$

the perturbations W_n being iid with distribution $\gamma_k = P(W_n = k)$, $k \in \mathbb{Z}$. With the same kind of argument as in Example 6.39, since the counting measure on $\mathbb{Z}$ is invariant for (X_n), (X_n) cannot be positive. If the distribution of W_n is symmetric, straightforward arguments lead to the conclusion that

$$\sum_{n=1}^{\infty} P_0(X_n = 0) = \infty\,,$$

from which we derive the (null) recurrence of (X_n) (see Feller 1970, Durrett 1991, or Problem 6.25).

Example 6.41. (Continuation of Example 6.24) Given the quasi-diagonal shape of the transition matrix, it is possible to directly determine the invariant distribution, $\pi = (\pi_0, \ldots, \pi_K)$. In fact, it follows from the equation $P^t\pi = \pi$ that

$$\begin{aligned}
\pi_0 &= P_{00}\pi_0 + P_{10}\pi_1,\\
\pi_1 &= P_{01}\pi_0 + P_{11}\pi_1 + P_{21}\pi_2,\\
&\vdots\\
\pi_K &= P_{(K-1)K}\pi_{K-1} + P_{KK}\pi_K\,.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\pi_1 &= \frac{P_{01}}{P_{10}}\,\pi_0,\\
\pi_2 &= \frac{P_{01}P_{12}}{P_{21}P_{10}}\,\pi_0,\\
&\vdots\\
\pi_k &= \frac{P_{01}\cdots P_{(k-1)k}}{P_{k(k-1)}\cdots P_{10}}\,\pi_0\,.
\end{aligned}$$

Hence,

indicates how positivity is a stability property stronger than recurrence. In fact, the latter corresponds to

$$P_\alpha(\tau_\alpha = \infty) = 0,$$

which is a necessary condition for $\mathbb{E}_\alpha[\tau_\alpha] < \infty$.

Proof. If $\mathbb{E}_\alpha[\tau_\alpha] < \infty$, $P_\alpha(\tau_\alpha < \infty) = 1$; thus, (X_n) is recurrent from Proposition 6.31. Consider a measure π given by

$$\pi(A) = \sum_{n=1}^{\infty} P_\alpha(X_n \in A, \tau_\alpha \geq n) \tag{6.17}$$

as in (6.16). This measure is invariant since $\pi(\alpha) = P_\alpha(\tau_\alpha < \infty) = 1$ and

$$\begin{aligned}\int K(x,A)\pi(dx) &= \pi(\alpha)K(\alpha,A) + \int_{\alpha^c} \sum_{n=1}^{\infty} K(x_n,A)\, P_\alpha(\tau_\alpha \geq n, dx_n) \\ &= K(\alpha,A) + \sum_{n=2}^{\infty} P_\alpha(X_n \in A, \tau_\alpha \geq n) = \pi(A).\end{aligned}$$

It is also finite as

$$\begin{aligned}\pi(\mathcal{X}) &= \sum_{n=1}^{\infty} P_\alpha(\tau_\alpha \geq n) = \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} P_\alpha(\tau_\alpha = m) \\ &= \sum_{m=1}^{\infty} m\, P_\alpha(\tau_\alpha = m) = \mathbb{E}_\alpha[\tau_\alpha] < \infty\,.\end{aligned}$$

Since π is invariant when (X_n) is positive, the uniqueness of the invariant distribution implies finiteness of $\pi(\mathcal{X})$, thus of $\mathbb{E}_\alpha[\tau_\alpha]$. Renormalizing π to $\pi/\pi(\mathcal{X})$ implies $\pi(\alpha) = (\mathbb{E}_\alpha[\tau_\alpha])^{-1}$. □

Following a now "classical" approach, the general case can be treated by splitting (X_n) to $(\check{X}_n)$ (which has an atom) and the invariant measure of $(\check{X}_n)$ induces an invariant measure for (X_n) by marginalization. A converse of Proposition 6.31 establishes the generality of invariance for Markov chains (see Meyn and Tweedie 1993, pp. 240–245, for a proof).

Theorem 6.38. *If (X_n) is a recurrent chain, there exists an invariant σ-finite measure which is unique up to a multiplicative factor.*

Example 6.39. Random walk on $\mathbb{R}$. Consider the random walk on $\mathbb{R}$, $X_{n+1} = X_n + W_n$, where W_n has a cdf Γ. Since $K(x,\cdot)$ is the distribution with cdf $\Gamma(y-x)$, the distribution of X_{n+1} is invariant by translation, and this implies that the Lebesgue measure is an invariant measure:

$$\int K(x,A)\lambda(dx) = \int\int_{A-x} \Gamma(dy)\lambda(dx)$$

Proof. If (X_n) is transient, there exists a covering of $\mathcal{X}$ by uniformly transient sets, A_j, with corresponding bounds

$$\mathbb{E}_x[\eta_{A_j}] \leq M_j, \qquad \forall x \in A_j,\ \forall j \in \mathbb{N}\,.$$

Therefore, by the invariance of π,

$$\pi(A_j) = \int K(x, A_j)\,\pi(dx) = \int K^n(x, A_j)\,\pi(dx).$$

Therefore, for every $k \in \mathbb{N}$,

$$k\,\pi(A_j) = \sum_{n=0}^{k} \int K^n(x, A_j)\,\pi(dx) \leq \int \mathbb{E}_x[\eta_{A_j}]\,\pi(dx) \leq M_j,$$

since, from (6.8) it follows that $\sum_{n=0}^{k} K^n(x, A_j) \leq \mathbb{E}_x[\eta_{A_j}]$. Letting k go to ∞ shows that $\pi(A_j) = 0$, for every $j \in \mathbb{N}$, and hence the impossibility of obtaining an invariant probability measure. □

We may, therefore, talk of *positive chains* and of *Harris positive chains*, without the superfluous denomination *recurrent* and *Harris recurrent*. Proposition 6.36 is useful only when the positivity of (X_n) can be proved, but, again, the chains produced by Markov chain Monte Carlo methods are, by nature, guaranteed to possess an invariant distribution.

6.5.2 Kac's Theorem

A classical result (see Feller 1970) on irreducible Markov chains with discrete state-space is that the stationary distribution, when it exists, is given by

$$\pi_x = (\mathbb{E}_x[\tau_x])^{-1}\,, \qquad x \in \mathcal{X},$$

where, from (6.7), we can interpret $\mathbb{E}_x[\tau_x]$ as the average number of excursions between two passages in x. (It is sometimes called *Kac's Theorem*.) It also follows that $(\mathbb{E}_x[\tau_x]^{-1})$ is the eigenvector associated with the eigenvalue 1 for the transition matrix $\mathbb{P}$ (see Problems 6.10 and 6.61). We now establish this result in the more general case when (X_n) has an atom, α.

Theorem 6.37. *Let (X_n) be ψ-irreducible with an atom α. The chain is positive if and only if $\mathbb{E}_\alpha[\tau_\alpha] < \infty$. In this case, the invariant distribution π for (X_n) satisfies*

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

The notation $\mathbb{E}_\alpha[\,\cdot\,]$ is legitimate in this case since the transition kernel is the same for every $x \in \alpha$ (see Definition 6.18). Moreover, Theorem 6.37

a discussion of the "almost" Harris recurrence of recurrent chains.) Tierney (1994) and Chan and Geyer (1994) analyze the role of Harris recurrence in the setup of Markov chain Monte Carlo algorithms and note that Harris recurrence holds for most of these algorithms (see Chapters 7 and 10).[2]

6.5 Invariant Measures

6.5.1 Stationary Chains

An increased level of stability for the chain (X_n) is attained if the marginal distribution of X_n is independent of n. More formally, this is a requirement for the existence of a probability distribution π such that $X_{n+1} \sim \pi$ if $X_n \sim \pi$, and Markov chain Monte Carlo methods are based on the fact that this requirement, which defines a particular kind of recurrence called *positive recurrence*, can be met. The Markov chains constructed from Markov chain Monte Carlo algorithms enjoy this greater stability property (except in very pathological cases; see Section 10.4.3). We therefore provide an abridged description of invariant measures and positive recurrence.

Definition 6.35. A σ-finite measure π is *invariant* for the transition kernel $K(\cdot,\cdot)$ (and for the associated chain) if

$$\pi(B) = \int_{\mathcal{X}} K(x, B)\, \pi(dx)\,, \qquad \forall B \in \mathcal{B}(\mathcal{X})\,.$$

When there exists an *invariant probability measure* for a ψ-irreducible (hence recurrent by Theorem 6.30) chain, the chain is *positive*. Recurrent chains that do not allow for a finite invariant measure are called *null recurrent*.

The invariant distribution is also referred to as *stationary* if π is a probability measure, since $X_0 \sim \pi$ implies that $X_n \sim \pi$ for every n; thus, the chain is stationary in distribution. (Note that the alternative case when π is not finite is more difficult to interpret in terms of behavior of the chain.) It is easy to show that if the chain is irreducible and allows for an σ-finite invariant measure, this measure is unique, up to a multiplicative factor (see Problem 6.60). The link between positivity and recurrence is given by the following result, which formalizes the intuition that the existence of a invariant measure prevents the probability mass from "escaping to infinity."

Proposition 6.36. *If the chain (X_n) is positive, it is recurrent.*

[2] Chan and Geyer (1994) particularly stress that *"Harris recurrence essentially says that there is no measure-theoretic pathology (...) The main point about Harris recurrence is that asymptotics do not depend on the starting distribution because of the 'split' chain construction."*

Definition 6.32. A set A is *Harris recurrent* if $P_x(\eta_A = \infty) = 1$ for all $x \in A$. The chain (X_n) is *Harris recurrent* if there exists a measure ψ such that (X_n) is ψ-irreducible and for every set A with $\psi(A) > 0$, A is Harris recurrent.

Recall that recurrence corresponds to $\mathbb{E}_x[\eta_\alpha] = \infty$, a weaker condition than $P_x(\eta_A = \infty) = 1$ (see Problem 6.30). The following proposition expresses Harris recurrence as a condition on $P_x(\tau_A < \infty)$ defined in (6.8).

Proposition 6.33. *If for every $A \in \mathcal{B}(\mathcal{X})$, $P_x(\tau_A < \infty) = 1$ for every $x \in A$, then $P_x(\eta_A = \infty) = 1$, for all $x \in \mathcal{X}$, and (X_n) is Harris recurrent.*

Proof. The average number of visits to B before a first visit to A is

$$U_A(x, B) = \sum_{n=1}^{\infty} P_x(X_n \in B, \tau_A \geq n)\,. \tag{6.16}$$

Then, $U_A(x, A) = P_x(\tau_A < \infty)$, since, if $B \subset A$, $P_x(X_n \in B, \tau_A \geq n) = P_x(X_n \in B, \tau = n) = P_x(\tau_B = n)$. Similarly, if $\tau_A(k), k > 1$, denotes the time of the kth visit to A, $\tau_A(k)$ satisfies

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

for every $x \in A$ and, by induction,

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

Since $P_x(\eta_A \geq k) = P_x(\tau_A(k) < \infty)$ and

$$P_x(\eta_A = \infty) = \lim_{k \to \infty} P_x(\eta_A \geq k),$$

we deduce that $P_x(\eta_A = \infty) = 1$ for $x \in A$. □

Note that the property of Harris recurrence is needed only when $\mathcal{X}$ is not denumerable. If $\mathcal{X}$ is finite or denumerable, we can indeed show that $\mathbb{E}_x[\eta_x] = \infty$ if and only if $P_x(\tau_x < \infty) = 1$ for every $x \in \mathcal{X}$, through an argument similar to the proof of Proposition 6.31. In the general case, it is possible to prove that if (X_n) is Harris recurrent, then $P_x(\eta_B = \infty) = 1$ for every $x \in \mathcal{X}$ and $B \in \mathcal{B}(\mathcal{X})$ such that $\psi(B) > 0$. This property then provides a sufficient condition for Harris recurrence which generalizes Proposition 6.31.

Theorem 6.34. *If (X_n) is a ψ-irreducible Markov chain with a small set C such that $P_x(\tau_C < \infty) = 1$ for all $x \in \mathcal{X}$, then (X_n) is Harris recurrent.*

Contrast this theorem with Proposition 6.31, where $P_x(\tau_C < \infty) = 1$ only for $x \in C$. This theorem also allows us to replace recurrence by Harris recurrence in Theorem 6.72. (See Meyn and Tweedie 1993, pp. 204–205 for

6.4.2 Criteria for Recurrence

The previous results establish a clear dichotomy between transience and recurrence for irreducible Markov chains. Nevertheless, given the requirement of Definition 6.29, it is useful to examine simpler criteria for recurrence. By analogy with discrete state-space Markov chains, a first approach is based on small sets.

Proposition 6.31. *A ψ-irreducible chain (X_n) is recurrent if there exists a small set C with $\psi(C) > 0$ such that $P_x(\tau_C < \infty) = 1$ for every $x \in C$.*

Proof. First, we show that the set C is recurrent. Given $x \in C$, consider $u_n = K^n(x, C)$ and $f_n = P_x(X_n \in C, X_{n-1} \notin C, \ldots, X_1 \notin C)$, which is the probability of first visit to C at the nth instant, and define

$$\tilde{U}(s) = 1 + \sum_{n=1}^{\infty} u_n s^n \qquad \text{and} \qquad Q(s) = \sum_{n=1}^{\infty} f_n s^n .$$

The equation

$$u_n = f_n + f_{n-1}u_1 + \cdots + f_1 u_{n-1} \tag{6.15}$$

describes the relation between the probability of a visit of C at time n and the probabilities of first visit of C. This implies

$$\tilde{U}(s) = \frac{1}{1 - Q(s)} ,$$

which connects $\tilde{U}(1) = \mathbb{E}_x[\eta_C] = \infty$ with $Q(1) = P_x(\tau_C < \infty) = 1$. Equation (6.15) is, in fact, valid for the split chain $(\check{X}_n)$ (see Problem 6.17), since a visit to $C \times \{0\}$ ensures independence by renewal. Since $\mathbb{E}_x[\eta_C]$, associated with (X_n), is larger than $\mathbb{E}_{\check{x}}[\eta_{C\times\{0\}}]$, associated with $(\check{x}_n)$, and $P_x(\tau_C < \infty)$ for (X_n) is equal to $P_{\check{X}}(\tau_{C\times\{0\}} < \infty)$ for $(\check{X}_n)$, the recurrence can be extended from $(\check{x}_n)$ to (X_n). The recurrence of $(\check{X}_n)$ follows from Theorem 6.28, since $C \times \{0\}$ is a recurrent atom for $(\check{X}_n)$. □□

A second method of checking recurrence is based on a generalization of the notions of small sets and minorizing conditions. This generalization involves a *potential function* V and a *drift condition* like (6.38) and uses the transition kernel $K(\cdot, \cdot)$ rather than the sequence K^n. Note 6.9.1 details this approach, as well as its bearing on the following stability and convergence results.

6.4.3 Harris Recurrence

It is actually possible to strengthen the stability properties of a chain (X_n) by requiring not only an infinite average number of visits to every small set but also an infinite number of visits for every path of the Markov chain. Recall that η_A is the number of passages of (X_n) in A, and we consider $P_x(\eta_A = \infty)$, the probability of visiting A an infinite number of times. The following notion of recurrence was introduced by Harris (1956).

Definition 6.27. A set A is called *recurrent* if $\mathbb{E}_x[\eta_A] = +\infty$ for every $x \in A$. The set A is *uniformly transient* if there exists a constant M such that $\mathbb{E}_x[\eta_A] < M$ for every $x \in A$. It is *transient* if there exists a covering of $\mathcal{X}$ by uniformly transient sets; that is, a countable collection of uniformly transient sets B_i such that

$$A = \bigcup_i B_i.$$

Theorem 6.28. *Let (X_n) be ψ-irreducible Markov chain with an accessible atom α.*

(i) If α is recurrent, every set A of $\mathcal{B}(\mathcal{X})$ such that $\psi(A) > 0$ is recurrent.
(ii) If α is transient, $\mathcal{X}$ is transient.

Property (i) is the most relevant in the Markov chain Monte Carlo setup and can be derived from the Chapman–Kolmogorov equations. Property (ii) is more difficult to establish and uses the fact that $P_\alpha(\tau_\alpha < \infty) < 1$ for a transient set when $\mathbb{E}_x[\eta_A]$ is decomposed conditionally on the last visit to α (see Meyn and Tweedie 1993, p. 181, and Problem 6.29).

Definition 6.29. A Markov chain (X_n) is *recurrent* if

(i) there exists a measure ψ such that (X_n) is ψ-irreducible, and
(ii) for every $A \in \mathcal{B}(\mathcal{X})$ such that $\psi(A) > 0$, $\mathbb{E}_x[\eta_A] = \infty$ for every $x \in A$.

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

The classification result of Theorem 6.28 can be easily extended to *strongly aperiodic* chains since they satisfy a minorizing condition (6.11), thus can be split as in (6.3.2), while the chain (X_n) and its split version $(\check{X}_n)$ (see Problem 6.17) are either both recurrent or both transient. The generalization to an arbitrary irreducible chain follows from the properties of the corresponding K_ε-chain which is strongly aperiodic, through the relation

$$\sum_{n=0}^{\infty} K_\varepsilon^n = \frac{1-\varepsilon}{\varepsilon} \sum_{n=0}^{\infty} K^n , \tag{6.14}$$

since

$$\mathbb{E}_x[\eta_A] = \sum_{n=0}^{\infty} K^n(x, A) = \frac{\epsilon}{1-\epsilon} \sum_{n=0}^{\infty} K_\epsilon^n(x, A).$$

This provides us with the following classification result:

Theorem 6.30. *A ψ-irreducible chain is either recurrent or transient.*

chain (X_n), but this property is too weak to ensure that the trajectory of (X_n) will enter A often enough. Consider, for instance, a maximization problem using a random walk on the surface of the function to maximize (see Chapter 5). The convergence to the global maximum cannot be guaranteed without a systematic sweep of this surface. Formalizing this stability of the Markov chain leads to different notions of *recurrence*. In a discrete setup, the *recurrence of a state* is equivalent to a guarantee of a sure return. This notion is thus necessarily satisfied for irreducible chains on a finite space.

Definition 6.25. In a finite state-space $\mathcal{X}$, a state $\omega \in \mathcal{X}$ is *transient* if the average number of visits to ω, $\mathbb{E}_\omega[\eta_\omega]$, is finite, and *recurrent* if $\mathbb{E}_\omega[\eta_\omega] = \infty$.

For irreducible chains, the properties of recurrence and transience are properties of the chain, not of a particular state. This fact is easily deduced from the Chapman–Kolmogorov equations. Therefore, if η_A denotes the number of visits defined in (6.8), for every $(x, y) \in \mathcal{X}^2$ either $\mathbb{E}_x[\eta_y] < \infty$ in the transient case or $\mathbb{E}_x[\eta_y] = \infty$ in the recurrent case. The chain is then said to be *transient* or *recurrent*, one of the two properties being necessarily satisfied in the irreducible case.

Example 6.26. Branching process. Consider a population whose individuals reproduce independently of one another. Each individual has X sibling(s), $X \in \mathbb{N}$, distributed according to the distribution with generating function $\phi(s) = \mathbb{E}[s^X]$. If individuals reproduce at fixed instants (thus defining *generations*), the size of the tth generation S_t $(t > 1)$ is given by

$$S_t = X_1 + \cdots + X_{S_{t-1}} ,$$

where the $X_i \sim \phi$ are independent. Starting with a single individual at time 0, $S_1 = X_1$, the generating function of S_t is $g_t(s) = \phi^t(s)$, with $\phi^t = \phi \circ \phi^{t-1}$ $(t > 1)$. The chain (S_t) is an example of a *branching process* (see Feller 1971, Chapter XII).

If ϕ does not have a constant term (i.e., if $P(X_1 = 0) = 0$), the chain (S_t) is necessarily transient since it is increasing. If $P(X_1 = 0) > 0$, the probability of a return to 0 at time t is $\rho_t = P(S_t = 0) = g_t(0)$, which thus satisfies the recurrence equation $\rho_t = \phi(\rho_{t-1})$. Therefore, there exists a limit ρ different from 1, such that $\rho = \phi(\rho)$, if and only if $\phi'(1) > 1$; namely if $\mathbb{E}[X] > 1$. The chain is thus transient when the average number of siblings per individual is larger than 1. If there exists a restarting mechanism in 0, $S_{t+1}|S_t = 0 \sim \phi$, it is easily shown that when $\phi'(1) > 1$, the number of returns to 0 follows a geometric distribution with parameter ρ. If $\phi'(1) \leq 1$, one can show that the chain is recurrent (see Example 6.42). ‖

The treatment of the general (that is to say, non-discrete) case is based on chains with atoms, the extension to general chains (with small sets) following from Athreya and Ney's (1978) splitting. We begin by extending the notions of recurrence and transience.

where the blocks D_i are stochastic matrices. This representation clearly illustrates the forced passage from one group of states to another, with a return to the initial group occurring every dth step. If the chain is irreducible (so all states communicate), there is only one value for the period. An irreducible chain is *aperiodic* if it has period 1. The extension to the general case requires the existence of a small set.

Definition 6.23. A ψ-irreducible chain (X_n) has a *cycle of length* d if there exists a small set C, an associated integer M, and a probability distribution ν_M such that d is the g.c.d. of

$$\{m \geq 1;\ \exists\, \delta_m > 0 \text{ such that } C \text{ is small for } \nu_m \geq \delta_m \nu_M\}.$$

A decomposition like (6.13) can be established in general. It is easily shown that the number d is independent of the small set C and that this number intrinsically characterizes the chain (X_n). The *period* of (X_n) is then defined as the largest integer d satisfying Definition 6.23 and (X_n) is *aperiodic* if $d = 1$. If there exists a small set A and a minorizing measure ν_1 such that $\nu_1(A) > 0$ (so it is possible to go from A to A in a single step), the chain is said to be *strongly aperiodic*). Note that the K_ε-chain can be used to transform an aperiodic chain into a strongly aperiodic chain.

In discrete setups, if one state $x \in \mathcal{X}$ satisfies $P_{xx} > 0$, the chain (X_n) is aperiodic, although this is not a necessary condition (see Problem 6.35).

Example 6.24. (Continuation of Example 6.14) The Bernoulli-Laplace chain is aperiodic and even strongly aperiodic since the diagonal terms satisfy $P_{xx} > 0$ for every $x \in \{0, \ldots, K\}$.

When the chain is continuous and the transition kernel has a component which is absolutely continuous with respect to the Lebesgue measure, with density $f(\cdot|x_n)$, a sufficient condition for aperiodicity is that $f(\cdot|x_n)$ is positive in a neighborhood of x_n. The chain can then remain in this neighborhood for an arbitrary number of instants before visiting any set A. For instance, in Example 6.3, (X_n) is strongly aperiodic when ε_n is distributed according to $\mathcal{U}_{[-1,1]}$ and $|\theta| < 1$ (in order to guarantee irreducibility). The next chapters will demonstrate that Markov chain Monte Carlo algorithms lead to aperiodic chains, possibly via the introduction of additional steps.

6.4 Transience and Recurrence

6.4.1 Classification of Irreducible Chains

From an algorithmic point of view, a Markov chain must enjoy good *stability* properties to guarantee an acceptable approximation of the simulated model. Indeed, *irreducibility* ensures that every set A will be visited by the Markov

$$\epsilon\nu(A) + (1-\epsilon)\frac{K(x_n, A) - \epsilon\nu(A)}{1-\epsilon} = K(x_n, A), \qquad \forall A \in \mathcal{B}(\mathcal{X}),$$

but it produces *renewal times* for each time j such that $X_j \in C$ and $X_{j+1} \sim \nu$.

Now we clearly see how the renewal times result in independent chains. When $X_{j+1} \sim \nu$, this event is totally independent of any past history, as the current state of the chain has no effect on the measure ν. Note also the key role that is played by the minorization condition. It allows us to create the split chain with the same marginal distribution as the original chain. We denote by $(j > 0)$

$$\tau_j = \inf\{n > \tau_{j-1};\ X_n \in C \text{ and } X_{n+1} \sim \nu\}$$

the sequence of *renewal times* with $\tau_0 = 0$. Athreya and Ney (1978) introduce the *augmented chain*, also called the *split chain* $\check{X}_n = (X_n, \check{\omega}_n)$, with $\check{\omega}_n = 1$ when $X_n \in C$ and X_{n+1} is generated from ν. It is then easy to show that the set $\check{\alpha} = C \times \{1\}$ *is an atom of the chain* $(\check{X}_n)$, the resulting subchain (X_n) being still a Markov chain with transition kernel $K(x_n, \cdot)$ (see Problem 6.17).

The notion of small set is useful only in finite and discrete settings when the individual probabilities of states are too small to allow for a reasonable rate of renewal. In these cases, small sets are made of collections of states with ν defined as a minimum. Otherwise, small sets reduced to a single value are also atoms.

6.3.3 Cycles and Aperiodicity

The behavior of (X_n) may sometimes be restricted by deterministic constraints on the moves from X_n to X_{n+1}. We formalize these constraints here and show in the following chapters that the chains produced by Markov chain Monte Carlo algorithms do not display this behavior and, hence, do not suffer from the associated drawbacks.

In the discrete case, the *period* of a state $\omega \in \mathcal{X}$ is defined as

$$d(\omega) = \text{g.c.d.}\ \{m \geq 1; K^m(\omega, \omega) > 0\}\ ,$$

where we recall that g.c.d. is the greatest common denominator. The value of the period is constant on all states that communicate with ω. In the case of an irreducible chain on a finite space $\mathcal{X}$, the transition matrix can be written (with a possible reordering of the states) as a block matrix

$$P = \begin{pmatrix} 0 & D_1 & 0 & \cdots & 0 \\ 0 & 0 & D_2 & & 0 \\ & & \ddots & & \\ D_d & 0 & 0 & & 0 \end{pmatrix}\ , \tag{6.13}$$

The proof of this characterization result is rather involved (see Meyn and Tweedie 1993, pp. 107–109). The decomposition of $\mathcal{X}$ as a denumerable union of small sets is based on an arbitrary small set C and the sequence

$$C_{nm} = \{y; K^n(y, C) > 1/m\}$$

(see Problem 6.19).

Small sets are obviously easier to exhibit than atoms, given the freedom allowed by the minorizing condition (6.11). Moreover, they are, in fact, very common since, in addition to Theorem 6.21, Meyn and Tweedie (1993, p. 134) show that for sufficiently regular (in a topological sense) Markov chains, every *compact* set is small. Atoms, although a special case of small sets, enjoy stronger stability properties since the transition probability is invariant on α. However, *splitting methods* (see below) offer the possibility of extending most of these properties to the general case and it will be used as a technique of proof in the remainder of the chapter.

If the minorizing condition holds for (X_n), there are two ways of deriving a companion Markov chain $(\check{X}_n)$ sharing many properties with (X_n) and possessing an atom α. The first method is called *Nummelin's splitting* and constructs a chain made of two copies of (X_n) (see Nummelin 1978 and Meyn and Tweedie 1993, Section 5.1).

A second method, discovered at approximately the same time, is due to Athreya and Ney (1978) and uses a stopping time to create an atom. We prefer to focus on this latter method because it is related to notions of *renewal time*, which are also useful in the control of Markov chain Monte Carlo algorithms (see Section 12.2.3).

Definition 6.22. A *renewal time* (or *regeneration time*) is a stopping rule τ with the property that $(X_\tau, X_{\tau+1}, \ldots)$ is independent of $(X_{\tau-1}, X_{\tau-2}, \ldots)$.

For instance, in Example 6.12, the returns to zero gain are renewal times. The excursions between two returns to zero are independent and identically distributed (see Feller 1970, Chapter III). More generally, visits to atoms are renewal times, whose features are quite appealing in convergence control for Markov chain Monte Carlo algorithms (see Chapter 12).

If (6.10) holds and if the probability $P_x(\tau_C < \infty)$ of a return to C in a finite time is identically equal to 1 on $\mathcal{X}$, Athreya and Ney (1978) modify the transition kernel when $X_n \in C$, by simulating X_{n+1} as

$$X_{n+1} \sim \begin{cases} \nu & \text{with probability } \varepsilon \\ \dfrac{K(X_n, \cdot) - \varepsilon\nu(\cdot)}{1-\varepsilon} & \text{with probability } 1-\varepsilon; \end{cases} \tag{6.12}$$

that is, by simulating X_{n+1} from ν with probability ε every time X_n is in C. This modification does not change the marginal distribution of X_{n+1} conditionally on x_n, since

While it trivially applies to every possible value of X_n in the discrete case, this notion is often too strong to be of use in the continuous case since it implies that the transition kernel is *constant* on a set of positive measure. A more powerful generalization is the so-called *minorizing condition*, namely that there exists a set $C \in \mathcal{B}(\mathcal{X})$, a constant $\varepsilon > 0$, and a probability measure ν such that

$$K(x, A) \geq \varepsilon \nu(A), \qquad \forall x \in C,\ \forall A \in \mathcal{B}(\mathcal{X})\ . \tag{6.10}$$

The probability measure ν thus appears as a constant component of the transition kernel on C. The minorizing condition (6.10) leads to the following notion, which is essential in this chapter and in Chapters 7 and 12 as a technique of proof and as the basis of *renewal theory*.

Definition 6.19. A set C is *small* if there exist $m \in \mathbb{N}^*$ and a nonzero measure ν_m such that

$$K^m(x, A) \geq \nu_m(A), \qquad \forall x \in C,\ \forall A \in \mathcal{B}(\mathcal{X})\ . \tag{6.11}$$

Example 6.20. (Continuation of Example 6.17) Since $X_n | x_{n-1} \sim \mathcal{N}(\theta x_{n-1}, \sigma^2)$, the transition kernel is bounded from below by

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{(-x_n^2 + 2\theta x_n \underline{w} - \theta^2 \underline{w}^2 \wedge \overline{w}^2)/2\sigma^2\right\} \text{ if } x_n > 0,$$
$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{(-x_n^2 + \theta x_n \overline{w}\sigma^{-2} - \theta^2 \underline{w}^2 \wedge \overline{w}^2)/2\sigma^2\right\} \text{ if } x_n < 0,$$

when $x_{n-1} \in [\underline{w}, \overline{w}]$. The set $C = [\underline{w}, \overline{w}]$ is indeed a small set, as the measure ν_1, with density

$$\frac{\exp\{(-x^2 + 2\theta x \underline{w})/2\sigma^2\}\, \mathbb{I}_{x>0} + \exp\{(-x^2 + 2\theta x \overline{w})/2\sigma^2\}\, \mathbb{I}_{x<0}}{\sqrt{2\pi}\,\sigma[\Phi(-\theta\underline{w}/\sigma^2)\, \exp\{\theta^2\underline{w}^2/2\sigma^2\} + [1 - \Phi(-\theta\underline{w}/\sigma^2)]\, \exp\{\theta^2\overline{w}^2/2\sigma^2\}]}\ ,$$

and

$$\varepsilon = \exp\{-\theta^2\underline{w}^2/2\sigma^2\}\, [\Phi(-\theta\underline{w}/\sigma^2)\, \exp\{\theta^2\underline{w}^2/2\sigma^2\} + [1 - \Phi(-\theta\underline{w}/\sigma^2)]\, \exp\{\theta^2\overline{w}^2/2\sigma^2\}]\ ,$$

satisfy (6.11) with $m = 1$. ‖

A sufficient condition for C to be small is that (6.11) is satisfied by the K_ε-chain in the special case $m = 1$. The following result indicates the connection between small sets and irreducibility.

Theorem 6.21. *Let (X_n) be a ψ-irreducible chain. For every set $A \in \mathcal{B}(\mathcal{X})$ such that $\psi(A) > 0$, there exist $m \in \mathbb{N}^*$ and a small set $C \subset A$ such that the associated minorizing measure satisfies $\nu_m(C) > 0$. Moreover, $\mathcal{X}$ can be decomposed in a denumerable partition of small sets.*

Theorem 6.16. *If (X_n) is φ-irreducible, there exists a probability measure ψ such that:*

(i) the Markov chain (X_n) is ψ-irreducible;
(ii) if there exists a measure ξ such that (X_n) is ξ-irreducible, then ξ is dominated by ψ; that is, $\xi \ll \psi$;
(iii) if $\psi(A) = 0$, then $\psi(\{y; P_y(\tau_A < \infty) > 0\}) = 0$;
(iv) the measure ψ is equivalent to

$$\psi_0(A) = \int_{\mathcal{X}} K_{1/2}(x, A)\, \varphi(dx), \qquad \forall A \in \mathcal{B}(\mathcal{X})\,; \tag{6.9}$$

that is, $\psi \ll \psi_0$ and $\psi_0 \ll \psi$.

This result provides a *constructive* method of determining the maximal irreducibility measure ψ through a candidate measure φ, which still needs to be defined.

Example 6.17. (Continuation of Example 6.6) When $X_{n+1} = \theta X_n + \varepsilon_{n+1}$ and ε_n are independent normal variables, the chain is irreducible, the reference measure being the *Lebesgue measure*, λ. (In fact, $K(x, A) > 0$ for every $x \in \mathbb{R}$ and every A such that $\lambda(A) > 0$.) On the other hand, if ε_n is uniform on $[-1, 1]$ and $|\theta| > 1$, the chain is not irreducible anymore. For instance, if $\theta > 1$, then

$$X_{n+1} - X_n \geq (\theta - 1)X_n - 1 \geq 0$$

for $X_n \geq 1/(\theta - 1)$. The chain is thus monotonically increasing and obviously cannot visit previous values. ‖

6.3.2 Atoms and Small Sets

In the discrete case, the transition kernel is necessarily atomic in the usual sense; that is, there exist points in the state-space with positive mass. The extension of this notion to the general case by Nummelin (1978) is powerful enough to allow for a control of the chain which is almost as "precise" as in the discrete case.

Definition 6.18. The Markov chain (X_n) has an *atom* $\alpha \in \mathcal{B}(\mathcal{X})$ if there exists an associated nonzero measure ν such that

$$K(x, A) = \nu(A), \qquad \forall x \in \alpha,\ \forall A \in \mathcal{B}(\mathcal{X})\,.$$

If (X_n) is ψ-irreducible, the atom is *accessible* when $\psi(\alpha) > 0$.

6.3 Irreducibility, Atoms, and Small Sets

6.3.1 Irreducibility

The property of irreducibility is a first measure of the sensitivity of the Markov chain to the initial conditions, x_0 or μ. It is crucial in the setup of Markov chain Monte Carlo algorithms, because it leads to a guarantee of convergence, thus avoiding a detailed study of the transition operator, which would otherwise be necessary to specify "acceptable" initial conditions.

In the discrete case, the chain is *irreducible* if all states communicate, namely if

$$P_x(\tau_y < \infty) > 0\,, \qquad \forall x, y \in \mathcal{X}\,,$$

τ_y being the first time y is visited, defined in (6.7). In many cases, $P_x(\tau_y < \infty)$ is uniformly equal to zero, and it is necessary to introduce an auxiliary measure φ on $\mathcal{B}(\mathcal{X})$ to correctly define the notion of irreducibility.

Definition 6.13. Given a measure φ, the Markov chain (X_n) with transition kernel $K(x, y)$ is *φ-irreducible* if, for every $A \in \mathcal{B}(\mathcal{X})$ with $\varphi(A) > 0$, there exists n such that $K^n(x, A) > 0$ for all $x \in \mathcal{X}$ (equivalently, $P_x(\tau_A < \infty) > 0$). The chain is *strongly φ-irreducible* if $n = 1$ for all measurable A.

Example 6.14. (Continuation of Example 6.3) In the case of the Bernoulli–Laplace model, the (finite) chain is indeed irreducible since it is possible to connect the states x and y in $|x - y|$ steps with probability

$$\prod_{i=x\wedge y}^{x\vee y-1} \left(\frac{K-i}{K}\right)^2 . \qquad \|$$

The following result provides equivalent definitions of irreducibility. The proof is left to Problem 6.13, and follows from (6.9) and the Chapman–Kolmogorov equations.

Theorem 6.15. *The chain (X_n) is φ-irreducible if and only if for every $x \in \mathcal{X}$ and every $A \in \mathcal{B}(\mathcal{X})$ such that $\varphi(A) > 0$, one of the following properties holds:*

(i) there exists $n \in \mathbb{N}^*$ such that $K^n(x, A) > 0$;
(ii) $\mathbb{E}_x[\eta_A] > 0$;
(iii) $K_\varepsilon(x, A) > 0$ for an $0 < \epsilon < 1$.

The introduction of the K_ε-chain then allows for the creation of a strictly positive kernel in the case of a φ-irreducible chain and this property is used in the following to simplify the proofs. Moreover, the measure φ in Definition 6.13 plays no crucial role in the sense that irreducibility is an intrinsic property of (X_n) and does not rely on φ.

The following theorem details the properties of the *maximal irreducibility measure* ψ.

Definition 6.10. Consider $A \in \mathcal{B}(\mathcal{X})$. The first n for which the chain enters the set A is denoted by

$$\tau_A = \inf\{n \geq 1; X_n \in A\} \tag{6.7}$$

and is called the *stopping time* at A with, by convention, $\tau_A = +\infty$ if $X_n \notin A$ for every n. More generally, a function $\zeta(x_1, x_2, \ldots)$ is called a *stopping rule* if the set $\{\zeta = n\}$ is measurable for the σ-algebra induced by $(X_0, \ldots, X_n)$. Associated with the set A, we also define

$$\eta_A = \sum_{n=1}^{\infty} \mathbb{I}_A(X_n), \tag{6.8}$$

the *number of passages* of (X_n) in A.

Of particular importance are the related quantities $\mathbb{E}_x[\eta_A]$ and $P_x(\tau_A < \infty)$, which are the *average number of passages in* A and the *probability of return to* A *in a finite number of steps*.

We will be mostly concerned with stopping rules of the form given in (6.7), which express the fact that τ_A takes the value n when none of the values of $X_0, X_1, \ldots, X_{n-1}$ are in the given state (or set) A, but the nth value is. The *strong Markov property* corresponds to the following result, whose proof follows from the weak Markov property and conditioning on $\{\zeta = n\}$:

Proposition 6.11. Strong Markov property *For every initial distribution μ and every stopping time ζ which is almost surely finite,*

$$\mathbb{E}_\mu[h(X_{\zeta+1}, X_{\zeta+2}, \cdots)|x_1, \ldots, x_\zeta] = \mathbb{E}_{x_\zeta}[h(X_1, X_2, \cdots)]\,,$$

provided the expectations exist.

We can thus condition on a random number of instants while keeping the fundamental properties of a Markov chain.

Example 6.12. Coin tossing. In a coin tossing game, player b has a gain of $+1$ if a head appears and player c has a gain of $+1$ if a tail appears (so player b has a "gain" of -1 (a loss) if a tail appears). If X_n is the sum of the gains of player b after n rounds of this coin tossing game, the transition matrix P is an infinite dimensional matrix with upper and lower subdiagonals equal to $1/2$. Assume that player b has B dollars and player c has C dollars, and consider the following return times:

$$\tau_1 = \inf\{n; X_n = 0\}, \quad \tau_2 = \inf\{n; X_n < -B\}, \quad \tau_3 = \inf\{n; X_n > C\},$$

which represent respectively the return to null and the ruins of the first and second players, that is to say, the first times the fortunes of both players, respectively B and C, are spent. The probability of ruin (bankruptcy) for the first player is then $P_0(\tau_2 > \tau_3)$. (Feller 1970, Chapter III, has a detailed analysis of this coin tossing game.) ‖

Lemma 6.7. Chapman–Kolmogorov equations *For every* $(m, n) \in \mathbb{N}^2$, $x \in \mathcal{X}$, $A \in \mathcal{B}(\mathcal{X})$,

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

(In a very informal sense, the Chapman–Kolmogorov equations state that to get from x to A in $m+n$ steps, you must pass through some y on the nth step.) In the discrete case, Lemma 6.7 is simply interpreted as a matrix product and follows directly from (6.2). In the general case, we need to consider K as an operator on the space of integrable functions; that is, we define

$$Kh(x) = \int h(y)\, K(x, dy)\,, \qquad h \in \mathcal{L}_1(\lambda)\,,$$

λ being the dominating measure of the model. K^n is then the nth composition of P, namely $K^n = K \circ K^{n-1}$.

Definition 6.8. A *resolvant* associated with the kernel P is a kernel of the form

$$K_\varepsilon(x, A) = (1-\varepsilon) \sum_{i=0}^{\infty} \varepsilon^i K^i(x, A), \qquad 0 < \epsilon < 1,$$

and the chain with kernel K_ε is a *K_ε-chain.*

Given an initial distribution μ, we can associate with the kernel K_ε a chain (X_n^ϵ) which formally corresponds to a subchain of the original chain (X_n), where the indices in the subchain are generated from a geometric distribution with parameter $1-\varepsilon$. Thus, K_ϵ is indeed a kernel, and we will see that the resulting Markov chain (X_n^ϵ) enjoys much stronger regularity. This will be used later to establish many properties of the original chain.

If $\mathbb{E}_\mu[\,\cdot\,]$ denotes the expectation associated with the distribution P_μ, the *(weak) Markov property* can be written as the following result, which just rephrases the limited memory properties of a Markov chain:

Proposition 6.9. Weak Markov property *For every initial distribution μ and every $(n+1)$ sample $(X_0, \ldots, X_n)$,*

$$\mathbb{E}_\mu[h(X_{n+1}, X_{n+2}, \ldots)|x_0, \ldots, x_n] = \mathbb{E}_{x_n}[h(X_1, X_2, \ldots)], \tag{6.6}$$

provided that the expectations exist.

Note that if h is the indicator function, then this definition is exactly the same as Definition 6.4. However, (6.6) can be generalized to other classes of functions—hence the terminology "weak"— and it becomes particularly useful with the notion of *stopping time* in the convergence assessment of Markov chain Monte Carlo algorithms in Chapter 12.

$$X_{n+1} = \begin{cases} Y & \text{with probability } \exp\{(E(Y) - E(X_n))/T\} \wedge 1 \\ X_n & \text{otherwise.} \end{cases}$$

If the temperature T depends on n, the chain is time heterogeneous.

Example 6.6. AR(1) Models. AR(1) models provide a simple illustration of Markov chains on continuous state-space. If

$$X_n = \theta X_{n-1} + \varepsilon_n\,, \qquad \theta \in \mathbb{R},$$

with $\varepsilon_n \sim N(0, \sigma^2)$, and if the ε_n's are independent, X_n is indeed independent from $X_{n-2}, X_{n-3}, \ldots$ conditionally on X_{n-1}. The Markovian properties of an AR(q) process can be derived by considering the vector $(X_n, \ldots, X_{n-q+1})$. On the other hand, ARMA(p, q) models do not fit in the Markovian framework (see Problem 6.3).

In the general case, the fact that the kernel K determines the properties of the chain (X_n) can be inferred from the relations

$$\begin{aligned} P_x(X_1 \in A_1) &= K(x, A_1)\,, \\ P_x((X_1, X_2) \in A_1 \times A_2) &= \int_{A_1} K(y_1, A_2)\, K(x, dy_1) \\ &\cdots \\ P_x((X_1, \ldots, X_n) \in A_1 \times \cdots \times A_n) &= \int_{A_1} \cdots \int_{A_{n-1}} K(y_{n-1}, A_n) \\ &\quad \times K(x, dy_1) \cdots K(y_{n-2}, dy_{n-1})\,. \end{aligned}$$

In particular, the relation $P_x(X_1 \in A_1) = K(x, A_1)$ indicates that $K(x_n, dx_{n+1})$ is a *version* of the conditional distribution of X_{n+1} given X_n. However, as we have defined a Markov chain by first specifying this kernel, we do not need to be concerned with different versions of the conditional probabilities. This is why we noted that constructing the Markov chain through the transition kernel was mathematically "cleaner." (Moreover, in the following chapters, we will see that the objects of interest are often these conditional distributions, and it is important that we need not worry about different versions. Nonetheless, the properties of a Markov chain considered in this chapter are independent of the version of the conditional probability chosen.)

If we denote $K^1(x, A) = K(x, A)$, the kernel for n transitions is given by ($n > 1$)

$$K^n(x, A) = \int_{\mathcal{X}} K^{n-1}(y, A)\, K(x, dy). \tag{6.5}$$

The following result provides convolution formulas of the type $K^{m+n} = K^m \star K^n$, which are called *Chapman–Kolmogorov equations*.

initial distribution $\mu = (\omega_1, \omega_2, \ldots)$, the marginal probability distribution of X_1 is obtained from the matrix multiplication

$$\mu_1 = \mu K \tag{6.2}$$

and, by repeated multiplication, $X_n \sim \mu_n = \mu K^n$. Similarly, in the continuous case, if μ denotes the initial distribution of the chain, namely if

$$X_0 \sim \mu\,, \tag{6.3}$$

then we let P_μ denote the probability distribution of (X_n) under condition (6.3). When X_0 is fixed, in particular for μ equal to the Dirac mass δ_{x_0}, we use the alternative notation P_{x_0}.

Definition 6.4. Given a transition kernel K, a sequence $X_0, X_1, \ldots, X_n, \ldots$ of random variables is a *Markov chain*, denoted by (X_n), if, for any t, the conditional distribution of X_t given $x_{t-1}, x_{t-2}, \ldots, x_0$ is the same as the distribution of X_t given x_{t-1}; that is,

$$\begin{aligned} P(X_{k+1} \in A | x_0, x_1, x_2, \ldots, x_k) &= P(X_{k+1} \in A | x_k) \\ &= \int_A K(x_k, dx)\,. \end{aligned} \tag{6.4}$$

The chain is *time homogeneous*, or simply *homogeneous*, if the distribution of $(X_{t_1}, \ldots, X_{t_k})$ given x_{t_0} is the same as the distribution of $(X_{t_1-t_0}, X_{t_2-t_0}, \ldots, X_{t_k-t_0})$ given x_0 for every k and every $(k+1)$-uplet $t_0 \le t_1 \le \cdots \le t_k$.

So, in the case of a Markov chain, if the initial distribution or the initial state is known, the construction of the Markov chain (X_n) is entirely determined by its *transition*, namely by the distribution of X_n conditionally on x_{n-1}.

The study of Markov chains is almost always restricted to the time-homogeneous case and we omit this designation in the following. It is, however, important to note here that an incorrect implementation of Markov chain Monte Carlo algorithms can easily produce nonhomogeneous Markov chains for which the standard convergence properties do not apply. (See also the case of the ARMS algorithm in Section 7.4.2.)

Example 6.5. Simulated Annealing. The *simulated annealing* algorithm (see Section 5.2.3 for details) is often implemented in a nonhomogeneous form and studied in time-homogeneous form. Given a finite state-space with size K, $\Omega = \{1, 2, \ldots, K\}$, an energy function $E(\cdot)$, and a temperature T, the simulated annealing Markov chain $X_0, X_1, \ldots$ is represented by the following transition operator: Conditionally on X_n, Y is generated from a fixed probability distribution $(\pi_1, \ldots, \pi_K)$ on Ω and the new value of the chain is given by

Thus, if the marginals are proper, for convergence we only need our chain to be aperiodic. This is easy to satisfy; a sufficient condition is that $K(x_n, \cdot) > 0$ (or, equivalently, $f(\cdot|x_n) > 0$) in a neighborhood of x_n. If the marginals are not proper, or if they do not exist, then the chain is not positive recurrent. It is either null recurrent or transient, and both cases are bad.

6.2 Basic Notions

A Markov chain is a sequence of random variables that can be thought of as evolving over time, with probability of a transition depending on the particular set in which the chain is. It therefore seems natural and, in fact, is mathematically somewhat cleaner to define the chain in terms of its *transition kernel*, the function that determines these transitions.

Definition 6.2. A *transition kernel* is a function K defined on $\mathcal{X} \times \mathcal{B}(\mathcal{X})$ such that

(i) $\forall x \in \mathcal{X}$, $K(x, \cdot)$ is a probability measure;
(ii) $\forall A \in \mathcal{B}(\mathcal{X})$, $K(\cdot, A)$ is measurable.

When $\mathcal{X}$ is *discrete*, the transition kernel simply is a (transition) matrix K with elements

$$P_{xy} = P(X_n = y | X_{n-1} = x) , \qquad x, y \in \mathcal{X}.$$

In the continuous case, the *kernel* also denotes the conditional density $K(x, x')$ of the transition $K(x, \cdot)$; that is, $P(X \in A|x) = \int_A K(x, x')dx'$.

Example 6.3. Bernoulli–Laplace Model. Consider $\mathcal{X} = \{0, 1, \ldots, M\}$ and a chain (X_n) such that X_n represents the state, at time n, of a tank which contains exactly M particles and is connected to another identical tank. Two types of particles are introduced in the system, and there are M of each type. If X_n denotes the number of particles of the first kind in the first tank at time n and the moves are restricted to a single exchange of particles between the two tanks at each instant, the transition matrix is given by (for $0 < x, y < M$) $P_{xy} = 0$ if $|x - y| > 1$,

$$P_{xx} = 2\,\frac{x(M-x)}{M^2}\,, \quad P_{x(x-1)} = \left(\frac{x}{M}\right)^2\,, \quad P_{x(x+1)} = \left(\frac{M-x}{M}\right)^2$$

and $P_{01} = P_{M(M-1)} = 1$. (This model is the *Bernoulli–Laplace* model; see Feller 1970, Chapter XV.) ‖

The chain (X_n) is usually defined for $n \in \mathbb{N}$ rather than for $n \in \mathbb{Z}$. Therefore, the distribution of X_0, the initial state of the chain, plays an important role. In the discrete case, where the kernel K is a transition matrix, given an

Definition 6.13 as the existence of $n \in \mathbb{N}$ such that $P(X_n \in A|X_0) > 0$ for every A such that $\pi(A) > 0$.) This property also ensures that most of the chains involved in MCMC algorithms are *recurrent* (that is, that the average number of visits to an arbitrary set A is infinite (Definition 6.29)), or even Harris recurrent (that is, such that the probability of an infinite number of returns to A is 1 (Definition 6.32)). Harris recurrence ensures that the chain has the same limiting behavior for *every* starting value instead of *almost* every starting value. (Therefore, this is the Markov chain equivalent of the notion of continuity for functions.)

This latter point is quite important in the context of MCMC algorithms. Since most algorithms are started from some arbitrary point x_0, we are in effect starting the algorithm from a set of measure zero (under a continuous dominating measure). Thus, insuring that the chain converges for almost every starting point is not enough, and we need Harris recurrence to guarantee convergence from every starting point.

The *stationary distribution* is also a *limiting distribution* in the sense that the limiting distribution of X_{n+1} is π under the total variation norm (see Proposition 6.48), notwithstanding the initial value of X_0. Stronger forms of convergence are also encountered in MCMC settings, like *geometric* and *uniform* convergences (see Definitions 6.54 and 6.58). In a simulation setup, a most interesting consequence of this convergence property is that the average

$$\frac{1}{N}\sum_{n=1}^{N} h(X_n) \tag{6.1}$$

converges to the expectation $\mathbb{E}_\pi[h(X)]$ almost surely. When the chain is *reversible* (Definition 6.44) (that is, when the transition kernel is symmetric), a Central Limit Theorem also holds for this average.

In Chapter 12, diagnostics will be based on a minorization condition; that is, the existence of a set C such that there also exists $m \in \mathbb{N}$, $\epsilon_m > 0$, and a probability measure ν_m such that

$$P(X_m \in A|X_0) \geq \epsilon_m \nu_m(A)$$

when $X_0 \in C$. The set C is then called a *small set* (Definition 6.19) and visits of the chain to this set can be exploited to create independent batches in the sum (6.1), since, with probability ϵ_m, the next value of the *m-skeleton Markov chain* $(X_{mn})_n$ is generated from the minorizing measure ν_m, which is independent of X_0.

As a final essential, it is sometimes helpful to associate the probabilistic language of Markov chains with the statistical language of data analysis.

Statistics		Markov Chain
marginal distribution	$\Leftrightarrow$	invariant distribution
proper marginals	$\Leftrightarrow$	positive recurrent

called *Markov processes*) since the very nature of simulation leads[1] us to consider only discrete-time stochastic processes, $(X_n)_{n\in\mathbb{N}}$. Indeed, Hastings (1970) notes that the use of pseudo-random generators and the representation of numbers in a computer imply that the Markov chains related with Markov chain Monte Carlo methods are, in fact, finite state-space Markov chains. However, we also consider arbitrary state-space Markov chains to allow for continuous support distributions and to avoid addressing the problem of approximation of these distributions with discrete support distributions, since such an approximation depends on both material and algorithmic specifics of a given technique (see Roberts et al. 1995, for a study of the influence of discretization on the convergence of Markov chains associated with Markov chain Monte Carlo algorithms).

6.1 Essentials for MCMC

For those familiar with the properties of Markov chains, this first section provides a brief survey of the properties of Markov chains that are contained in the chapter and are essential for the study of MCMC methods. Starting with Section 6.2, the theory of Markov chains is developed from first principles.

In the setup of MCMC algorithms, Markov chains are constructed from a *transition kernel* K (Definition 6.2), a conditional probability density such that $X_{n+1} \sim K(X_n, X_{n+1})$. A typical example is provided by the *random walk* process, formally defined as follows.

Definition 6.1. A sequence of random variables (X_n) is a *random walk* if it satisfies

$$X_{n+1} = X_n + \epsilon_n \,,$$

where ϵ_n is generated independently of $X_n, X_{n-1}, \ldots$. If the distribution of the ϵ_n is symmetric about zero, the sequence is called a *symmetric random walk*.

There are many examples of random walks (see Examples 6.39, 6.40, and 6.73), and random walks play a key role in many MCMC algorithms, particularly those based on the Metropolis–Hastings algorithm (see Chapter 7).

The chains encountered in MCMC settings enjoy a very strong stability property, namely a *stationary probability distribution* exists by construction (Definition 6.35); that is, a distribution π such that if $X_n \sim \pi$, then $X_{n+1} \sim \pi$, if the kernel K allows for free moves all over the state space. (This freedom is called *irreducibility* in the theory of Markov chains and is formalized in

[1] Some Markov chain Monte Carlo algorithms still employ a diffusion representation to speed up convergence to the stationary distribution (see, for instance, Section 7.8.5, Roberts and Tweedie 1995, or Phillips and Smith 1996).

6

Markov Chains

> Leaphorn never counted on luck. Instead, he expected order—the natural sequence of behavior, the cause producing the natural effect, the human behaving in the way it was natural for him to behave. He counted on that and on his own ability to sort out the chaos of observed facts and find in them this natural order.
> —Tony Hillerman, *The Blessing Way*

In this chapter we introduce fundamental notions of Markov chains and state the results that are needed to establish the convergence of various MCMC algorithms and, more generally, to understand the literature on this topic. Thus, this chapter, along with basic notions of probability theory, will provide enough foundation for the understanding of the following chapters. It is, unfortunately, a necessarily brief and, therefore, incomplete introduction to Markov chains, and we refer the reader to Meyn and Tweedie (1993), on which this chapter is based, for a thorough introduction to Markov chains. Other perspectives can be found in Doob (1953), Chung (1960), Feller (1970, 1971), and Billingsley (1995) for general treatments, and Norris (1997), Nummelin (1984), Revuz (1984), and Resnick (1994) for books entirely dedicated to Markov chains. Given the purely utilitarian goal of this chapter, its style and presentation differ from those of other chapters, especially with regard to the plethora of definitions and theorems and to the rarity of examples and proofs. In order to make the book accessible to those who are more interested in the implementation aspects of MCMC algorithms than in their theoretical foundations, we include a preliminary section that contains the essential facts about Markov chains.

Before formally introducing the notion of a Markov chain, note that we do not deal in this chapter with Markov models in *continuous time* (also

6.9.5 Covariance in Markov Chains

An application of Chebychev's inequality shows that the convergence of an average of random variables from a Markov chain can be connected to the behavior of the covariances, with a sufficient condition for convergence in probability being that the covariances go to zero.

We assume that the Markov chain is Harris positive and aperiodic, and is stationary. We also assume that the random variables of the chain have finite variance. Thus, let (X_n) be a stationary ergodic Markov chain with mean 0 and finite variance. The variance of the average of the X_i's is

$$\text{var}\left(\frac{\sum_{i=0}^n X_i}{(n+1)}\right) = \frac{1}{n+1}\text{var}(X_0) + \frac{2}{n+1}\sum_{k=1}^n \frac{n-k+1}{n+1}\text{cov}(X_0, X_k)\,, \tag{6.52}$$

so the covariance term in (6.52) will go to zero if $\sum_{k=1}^n \text{cov}(X_0, X_k)/n$ goes to zero, and a sufficient condition for this is that $\text{cov}(X_0, X_k)$ converges to 0 (Problem 6.38).

To see when $\text{cov}(X_0, X_k)$ converges to 0, write

$$\begin{aligned} |\text{cov}(X_0, X_k)| &= |\mathbb{E}[X_0 X_k]| \\ &= |\mathbb{E}[X_0 \mathbb{E}(X_k|X_0)]| \\ &\le [\mathbb{E}(X_0^2)]^{1/2}\{\mathbb{E}[\mathbb{E}(X_k|X_0)]^2\}^{1/2}, \end{aligned} \tag{6.53}$$

where we used the Cauchy-Schwarz inequality. Since $\mathbb{E}(X_0^2) = \sigma^2$, $\text{cov}(X_0, X_k)$ will go to zero if $\mathbb{E}[\mathbb{E}(X_k|X_0)]^2$ goes to 0.

Example 6.84. (Continuation of Example 6.6) Consider the AR(1) model

$$X_k = \theta X_{k-1} + \epsilon_k, \quad k = 0, \ldots, n, \tag{6.54}$$

when the ϵ_k's are iid $\mathcal{N}(0,1)$, θ is an unknown parameter satisfying $|\theta| < 1$, and $X_0 \sim \mathcal{N}(0, \sigma^2)$. The X_k's all have marginal normal distributions with mean zero. The variance of X_k satisfies $\text{var}(X_k) = \theta^2 \text{var}(X_{k-1}) + 1$ and, $\text{var}(X_k) = \sigma^2$ for all k, provided $\sigma^2 = 1/(1-\theta^2)$. This is the stationary case in which it can be shown that

$$\mathbb{E}(X_k|X_0) = \theta^k X_0 \tag{6.55}$$

and, hence, $\mathbb{E}[\mathbb{E}(X_k|X_0)]^2 = \theta^{2k}\sigma^2$, which goes to zero as long as $|\theta| < 1$. Thus, $\text{var}(\bar{X})$ converges to 0. (See Problem 6.68.) ‖

Returning to (6.53), let M be a positive constant and write

$$\begin{aligned} \mathbb{E}[\mathbb{E}(X_k|X_0)]^2 &= \mathbb{E}[\mathbb{E}(X_k \mathbb{I}_{X_k > M}|X_0) + \mathbb{E}(X_k \mathbb{I}_{X_k \le M}|X_0)]^2 \\ &\le 2\mathbb{E}[\mathbb{E}(X_k \mathbb{I}_{X_k > M}|X_0)]^2 + 2\mathbb{E}[\mathbb{E}(X_k \mathbb{I}_{X_k \le M}|X_0)]^2. \end{aligned} \tag{6.56}$$

Examining the two terms on the right side of (6.56), the first term can be made arbitrarily small using the fact that X_k has finite variance, while the second term converges to zero as a consequence of Theorem 6.51. We formalize this in the following theorem.

Theorem 6.85. *If the Markov chain (X_n) is positive and aperiodic, with $\text{var}(X_n) < \infty$, then $\text{cov}(X_0, X_k)$ converges to 0.*