A short course in Markov chains

Randal Douc

Chapter 1

Atomic Markov chains

Time schedule (**Note** 1): Session 6

In this chapter, we will examine the properties of Markov kernels that admit atoms (to be formally defined later). Specifically, we will establish conditions under which the following holds:

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

Additionally, we will briefly discuss how the presence of atoms simplifies the treatment of other properties, such as the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT). The chapter will provide an overview of the approaches used to achieve these types of convergence.

We first recall the solidarity lemma, which will be used several times throughout the chapter:

LEMMA 1.1 (▶ *The solidarity lemma).* Let $A,B\in\mathscr{X}$ such that $\inf_{x\in A}\mathbb{P}_x(\sigma_B<\infty)>0$. Then, for any probability measure $\xi\in\mathscr{M}_1(\mathsf{X})$,

$$\{N_A=\infty\}\subset\{N_B=\infty\}\,,\quad \mathbb{P}_\xi-a.s.$$

Hence, if for some initial distribution $\xi\in\mathscr{M}_1(\mathsf{X})$, $\mathbb{P}_\xi(N_A=\infty)=1$, then we also have $\mathbb{P}_\xi(N_B=\infty)=1$.

▶**Q-1.1.** Sorry, but I don't remember the notation N_A...

It represents the number of visits to the set A. If you'd like, I can provide a brief refresher on the notation.

▶**Q-1.2.** It would be so nice of you!!

You're welcome. Let $A\in\mathscr{X}$. Recall that $\sigma_A^0=1$ (by convention) and for $n\geqslant 1$, $\sigma_A^n=\sigma_A^{n-1}+\sigma_A\circ\theta_{\sigma_A^{n-1}}$. In words, $\left\{\sigma_A^n\,:\,n\geqslant 1\right\}$ are the successive return times to the set A.

Then, $N_A=\sum_{k=0}^\infty \mathbf{1}_A(X_k)=\mathbf{1}_A(X_0)+\sum_{n=1}^\infty \mathbf{1}_{\{\sigma_A^n<\infty\}}$ and we have

$$U(x,A)=\mathbb{E}_x[N_A]=\sum_{k=0}^\infty P^k(x,A)=\mathbf{1}_A(x)+\sum_{n=1}^\infty \mathbb{P}_x(\sigma_A^n<\infty)$$

It consists of several equivalent expressions. However, it is important to keep all of them in mind, as certain expressions may be more suitable in specific contexts than others.

▶**Q-1.3.** What is U?

By definition, $U=\sum_{k=0}^\infty P^k$, is called the potential kernel. While it is indeed a kernel, note that this kernel is not finite, since $U(x,\mathsf{X})=\sum_{k=0}^\infty P^k(x,\mathsf{X})=\infty$ because $P(x,\mathsf{X})=1$. We will not use it frequently in this course, and in most situations, we will write $\mathbb{E}_x[N_A]$ (which is more intuitive) instead of the more abstract expression $U(x,A)$.

1.1 Recurrent, transient atoms

In all this chapter, $(\mathsf{X}, \mathscr{X})$ is a measurable space and P is a Markov kernel on $\mathsf{X} \times \mathscr{X}$.

DEFINITION 1.2 . We say that $\alpha \in \mathscr{X}$ is an atom for P if there exists a probability measure $\nu \in \mathscr{M}_1(\mathsf{X})$ such that

$$P(x, \cdot) = \nu(\cdot), \quad \forall x \in \alpha. \tag{1.1}$$

►**Q-1.4.** Does it happen often?

Usually, no. In most Markov chains on a general space, the condition (1.1) is too stringent and often, it is replaced by the following. There exist $n \in \mathbb{N}^*$, $\varepsilon > 0$, $\nu \in \mathscr{M}_1(\mathsf{X})$ such that

$$P^n(x, \cdot) \geqslant \varepsilon \nu(\cdot), \quad \forall x \in \alpha. \tag{1.2}$$

However, if the latter condition holds, we can show that the initial Markov chain can be extended by incorporating a carefully chosen additional component, such that the resulting Markov kernel admits an accessible atom. We then analyze the extended atomic Markov chain and subsequently transfer its properties to the original chain. This approach is known as the splitting method—but we are getting ahead of ourselves. For now, we will focus solely on properties of Markov kernels with accessible atoms.

►**Q-1.5.** You said accessible atoms? Why accessible?

While any singleton is an atom, in many cases—especially outside discrete state spaces—singletons are not accessible. Therefore, most of the properties discussed in this chapter rely on the assumption of the existence of accessible atoms. Now, let us return to the topic of atomic Markov kernels.

In what follows, if for some set $\alpha \in \mathscr{X}$, any of $\mathbb{P}_x(A)$, $\mathbb{E}_x[Z]$, $Pf(x)$, or $P^n(x,B)$ does not depend on $x \in \alpha$, we will simply write $\mathbb{P}_\alpha(A)$, $\mathbb{E}_\alpha[Z]$, $Pf(\alpha)$, or $P^n(\alpha,B)$. While this is technically an abuse of notation, it is often very convenient.

►**Q-1.6.** Before you continue, may I ask you another question?

Of course.

►**Q-1.7.** When you enter an atom, the next state of the Markov chain is drawn according to some fixed distribution ν. Does it mean that we forget everything from the past?

You've touched on a fundamental property of Markov kernels with atoms. Due to this fixed distribution, we will see that the excursions between successive visits to an atom are i.i.d. This allows us to leverage properties of these i.i.d. excursions to derive important results for our Markov chain. To this aim, we need to clearly understand the behavior of return times to the atom. Let's delve into that.

LEMMA 1.3 (► *Maximum principle*). For any atom α,

$$\mathbb{E}_\alpha[N_\alpha] = \sup_{x \in \mathsf{X}} \mathbb{E}_x[N_\alpha]$$

PROOF. For any $x \in \mathsf{X}$, applying the strong Markov property at the stopping time τ_α in the second equality below yields

$$\mathbb{E}_x[N_\alpha] = \mathbb{E}_x\left[\mathbf{1}_{\{\tau_\alpha < \infty\}} N_\alpha \circ \theta_{\tau_\alpha}\right] = \mathbb{P}_x(\tau_\alpha < \infty)\mathbb{E}_\alpha[N_\alpha] \leqslant \mathbb{E}_\alpha[N_\alpha]$$

which completes the proof. ■

►**Q-1.8.** So, you have again some property on the number of visits?

You seem surprised... That's everyday life in Markov chain theory. The solidarity lemma and the maximum principle are often combined to derive elegant results regarding the number of visits.

Now, let us move on to a useful result for verifying the accessibility of a set A when the Markov kernel P admits accessible atoms.

LEMMA 1.4 . Assume that P admits an accessible atom α.

(i) A is accessible $\Leftrightarrow \mathbb{P}_\alpha(\sigma_A < \infty) > 0$

(ii) A is not accessible $\Rightarrow A^c$ is accessible.

PROOF. We start with (i). The implication $\Rightarrow$ follows since A is accessible. We next show $\Leftarrow$. Assume $\mathbb{P}_\alpha(\sigma_A < \infty) > 0$. For any $x \in \mathsf{X}$, the strong Markov property applied to the stopping time σ_α yields:

$$\mathbb{P}_x(\sigma_A < \infty) \geqslant \mathbb{P}_x(\sigma_\alpha < \infty, \sigma_A \circ \theta_{\sigma_\alpha} < \infty) = \mathbb{P}_x(\sigma_\alpha < \infty)\mathbb{P}_\alpha(\sigma_A < \infty)$$

and the rhs is positive as a product of positive numbers. Hence, A is accessible and the proof of (i) is completed.

We now prove (ii) using (i). If A is not accessible, by (i) we deduce that $\mathbb{P}_\alpha(\sigma_A < \infty) = 0$ and hence $P(\alpha, A^c) = 1 > 0$. It implies $\mathbb{P}_\alpha(\sigma_{A^c} < \infty) > 0$. Applying again (i) with A replaced by A^c, we get that A^c is accessible. ■

DEFINITION 1.5 . Let α be an atom for P.

(a) α is *recurrent* if $\mathbb{E}_\alpha[N_\alpha] = \infty$.

(b) α is *transient* if $\mathbb{E}_\alpha[N_\alpha] < \infty$.

►**Q-1.9.** So, an atom is either recurrent or transient?

Yes! Next, I will provide some tools for checking recurrence or transience for atoms. Before that, we will establish a connection between the expected number of visits starting from α and $\mathbb{P}_\alpha(\sigma_\alpha < \infty)$. More specifically, for any $n \geqslant 1$, by applying the strong Markov property at the stopping time σ_α^{n-1},

$$\begin{aligned}\mathbb{P}_\alpha(\sigma_\alpha^n < \infty) &= \mathbb{P}_\alpha(\sigma_\alpha^{n-1} + \sigma_\alpha \circ \theta_{\sigma_\alpha^{n-1}} < \infty) = \mathbb{P}_\alpha(\sigma_\alpha^{n-1} < \infty,\ \sigma_\alpha \circ \theta_{\sigma_\alpha^{n-1}} < \infty) \\ &= \mathbb{P}_\alpha(\sigma_\alpha^{n-1} < \infty)\mathbb{P}_\alpha(\sigma_\alpha < \infty) = \mathbb{P}_\alpha(\sigma_\alpha < \infty)^n \end{aligned} \tag{1.3}$$

Hence, using $\sum_{k=1}^\infty \mathbf{1}_\alpha(X_k) = \sum_{n=1}^\infty \mathbf{1}_{\{\sigma_\alpha^n < \infty\}}$, we have the equivalent expressions:

$$\mathbb{E}_\alpha[N_\alpha] = \sum_{k=0}^\infty P^k(\alpha, \alpha) = 1 + \sum_{n=1}^\infty \mathbb{P}_\alpha(\sigma_\alpha^n < \infty) = \sum_{n=0}^\infty \mathbb{P}_\alpha(\sigma_\alpha < \infty)^n \tag{1.4}$$

where we have used (1.3) in the last equality.

►**Q-1.10.** The last identity is stunning!

I'm glad to see your enthusiasm. However, I must emphasize that the last identity holds under the assumption that α is an atom. The following theorem provides several equivalent expressions, any of which can be used to determine whether a given atom is recurrent or transient.

THEOREM 1.6 . Let α be an atom for P.

(i) $\boxed{\alpha \text{ recurrent} \Leftrightarrow \mathbb{E}_\alpha[N_\alpha] = \infty \Leftrightarrow \mathbb{P}_\alpha(N_\alpha = \infty) = 1 \Leftrightarrow \mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1.}$

Moreover, if any of the above conditions holds, for all $x \in \mathsf{X}$, $\mathbb{P}_x(\sigma_\alpha < \infty) = \mathbb{P}_x(N_\alpha = \infty)$.

(ii) $\boxed{\alpha \text{ transient} \Leftrightarrow \mathbb{E}_\alpha[N_\alpha] < \infty \Leftrightarrow \mathbb{P}_\alpha(N_\alpha < \infty) = 1 \Leftrightarrow \mathbb{P}_\alpha(\sigma_\alpha < \infty) < 1.}$

Moreover, if any of the above conditions holds, $\mathbb{E}_\alpha[N_\alpha] = \frac{1}{1-\mathbb{P}(\sigma_\alpha<\infty)}$.

PROOF. We start with (i). For convenience, we write

$$\underbrace{\mathbb{E}_\alpha[N_\alpha] = \infty}_{a}, \quad \underbrace{\mathbb{P}_\alpha(N_\alpha = \infty) = 1}_{b}, \quad \underbrace{\mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1}_{c}.$$

Applying (1.4), $a \Leftrightarrow c$. Obviously, $b \Rightarrow a$. We next show $c \Rightarrow b$. Assuming c, (1.3) shows that $\mathbb{P}_\alpha(\sigma_\alpha^n < \infty) = 1$. And hence, $1 = \mathbb{P}_\alpha\left(\cap_{n\in\mathbb{N}}\{\sigma_\alpha^n < \infty\}\right) = \mathbb{P}_\alpha(N_\alpha = \infty)$, showing b. Finally, $a \Leftrightarrow b \Leftrightarrow c$. Now, if any of these conditions holds, the strong Markov property applied at σ_α yields

$$\mathbb{P}_x(N_\alpha = \infty) = \mathbb{P}_x(\sigma_\alpha < \infty, N_\alpha = \infty) = \mathbb{P}_x(\sigma_\alpha < \infty, N_\alpha \circ \theta_{\sigma_\alpha} = \infty) = \mathbb{P}_x(\sigma_\alpha < \infty)\underbrace{\mathbb{P}_\alpha(N_\alpha = \infty)}_{1}$$

We now turn to (ii). For convenience, we write

$$\underbrace{\mathbb{E}_\alpha[N_\alpha] < \infty}_{d}, \quad \underbrace{\mathbb{P}_\alpha(N_\alpha < \infty) = 1}_{e}, \quad \underbrace{\mathbb{P}_\alpha(\sigma_\alpha < \infty) < 1}_{f}$$

Obviously, $d \Rightarrow e$. Moreover, $e \Rightarrow \bar{b} \Rightarrow \bar{c} = f$ (where we use the convention: for $\gamma \in \{a,b,c,d,e,f\}$, $\bar{\gamma}$ means that γ is not true). We have $f \Rightarrow \bar{c} \Rightarrow \bar{a} = d$. Finally $d \Rightarrow e \Rightarrow f \Rightarrow d$ and all the equivalences are thus established. If any of these condtions holds, (1.4) shows $U(\alpha,\alpha) = \frac{1}{1-\mathbb{P}(\sigma_\alpha<\infty)}$. ■

▶**Q-1.11.** Thanks for these great results! I think I'm almost done learning everything about transience and recurrence.

Not so fast! We've only just defined recurrent and transient atoms. The definition for general sets is provided below.

DEFINITION 1.7 (▶ *Recurrent set, recurrent kernel).*

- A set $A \in \mathscr{X}$ is *recurrent* if $\mathbb{E}_x[N_A] = \infty$ for all $x \in A$.
- The kernel P is *recurrent* if any accessible set is recurrent.

DEFINITION 1.8 (▶ *Uniformly transient set, transient set, transient kernel).*

- A set $A \in \mathscr{X}$ is *uniformly transient* if $\sup_{x\in A} \mathbb{E}_x[N_A] < \infty$.
- A set A is *transient* if it is a countable union of uniformly transient sets.
- The kernel P is *transient* if X is transient.

▶**Q-1.12.** So, a set cannot be recurrent and uniformly transient at the same time?

Of course not. As for Markov kernels P with accessible atoms, they are always either recurrent or transient as shown below.

THEOREM 1.9 . If P admits an accessible atom α, then

(i) P is recurrent if and only if α is recurrent.

(ii) P is transient if and only if α is transient.

Proof. We start with (i). For convenience, we set

$$\underbrace{P \text{ is recurrent}}_{a}, \quad \underbrace{\alpha \text{ is recurrent}}_{b}$$

Since α is accessible, the implication $a \Rightarrow b$ is obvious. Now assume b. Let A be an accessible set. We must show that A is recurrent, that is, for any $x \in \mathsf{X}$, $\mathbb{E}_x[N_A] = \infty$. Since α is an accessible atom and A accessible, we deduce from Lemma 1.4-(i) that $\mathbb{P}_\alpha(\sigma_A < \infty) > 0$. This allows to apply the solidarity lemma –Lemma 1.1– and since α is recurrent, we get $1 = \mathbb{P}_\alpha(N_\alpha = \infty) \leqslant \mathbb{P}_\alpha(N_A = \infty)$. Now, for any $x \in \mathsf{X}$, using the Markov property with the stopping time σ_α in the equality below

$$\mathbb{E}_x[N_A] \geqslant \mathbb{E}_x[\mathbf{1}_{\{\sigma_\alpha<\infty\}} N_A] \geqslant \mathbb{E}_x[\mathbf{1}_{\{\sigma_\alpha<\infty\}} N_A \circ \theta_{\sigma_\alpha+1}] = \mathbb{P}_x(\sigma_\alpha < \infty)\mathbb{E}_\alpha[N_A \circ \theta]$$

The rhs is infinite since the first term is positive (because α is accessible) and the second term infinite (because $\mathbb{P}_\alpha(N_A = \infty) = 1$).

We now prove (ii). Assume first that α is transient. Define $\mathsf{X}_m = \{x \in \mathsf{X} : \mathbb{P}_x(\sigma_\alpha < \infty) \geqslant 1/m\}$. Since α is accessible, we have that $\mathsf{X} = \cup_{m\in\mathbb{N}^*}\mathsf{X}_m$. We now show that X_m is uniformly transient. For any $x \in \mathsf{X}$,

$$\mathbb{E}_x[N_{\mathsf{X}_m}] \leqslant \mathbb{E}_x[N_\alpha] \leqslant \mathbb{E}_\alpha[N_\alpha] < \infty$$

where the first inequality is a consequence of the solidarity lemma (Lemma 1.1), the second follows from the maximum principle (Lemma 1.3), and the last term is finite because α is transient.

Assume now that P is transient. We will show that α is transient by contradiction. Assume indeed that α is recurrent. Then, by (i), every accessible set is recurrent. But since P is transient, there exists uniformly transient sets $\{\mathsf{X}_m : m \in \mathbb{N}\}$ such that $\mathsf{X} = \cup_{m\in\mathbb{N}}\mathsf{X}_m$. Since $\mathbb{P}(\alpha, \mathsf{X}) = 1$, we deduce that there exists some m such that $P(\alpha, \mathsf{X}_m) > 0$. Hence, Lemma 1.4 applies and X_m is accessible and hence recurrent. This contradicts that X_m is uniformly transient. Hence, by contradiction, α is transient and the proof is completed. ■

1.2 Period

Let us now turn to periodic properties of Markov kernels.

Definition 1.10 . Let P be a Markov kernel with an atom α. We say that $d(\alpha)$ is the period of α if

$$d(\alpha) = \text{g.c.d}\{n > 0 : P^n(\alpha, \alpha) > 0\}$$

with the convention that $\text{g.c.d}(\emptyset) = \infty$.

If $d(\alpha) = 1$, we say that α is *aperiodic*.

Define $E_\alpha = \{n > 0 : P^n(\alpha, \alpha) > 0\}$.

Lemma 1.11 . Let α be an atom for P. Then, $d(\alpha) = \text{g.c.d}\{n > 0 : \mathbb{P}_\alpha(\sigma_\alpha = n) > 0\}$.

Proof. Define $E'_\alpha = \{n > 0 : \mathbb{P}_\alpha(\sigma_\alpha = n) > 0\}$ and let $d'(\alpha) = \text{g.c.d}\, E'_\alpha$. We must show that $d(\alpha) = d'(\alpha)$. For any $n \in E'_\alpha$,

$$0 < \mathbb{P}_\alpha(\sigma_\alpha = n) \leqslant P^n(\alpha, \alpha)$$

and thus $n \in E_\alpha$. Consequently, $d(\alpha)$ divides any $n \in E'_\alpha$, and therefore $d(\alpha)$ divides $d'(\alpha)$.

Next, let $n \in E_\alpha$. Then,

$$0 < P^n(\alpha,\alpha) = \mathbb{P}_\alpha(X_n \in \alpha) = \sum_{k=1}^{n} \left(\sum_{0=s_0<s_1<s_2<\cdots<s_k=n} \mathbb{P}_\alpha\left(X_{s_\ell} \in \alpha \text{ for } \ell \in [0:k], X_j \notin \alpha \text{ for } j \in [0:n] \setminus \{s_0,\ldots,s_k\}\right) \right)$$
$$= \sum_{k=1}^{n} \left(\sum_{0=s_0<s_1<s_2<\cdots<s_k=n} \prod_{\ell=1}^{k} \mathbb{P}_\alpha(\sigma_\alpha = s_\ell - s_{\ell-1}) \right)$$

Hence, there exist integers $n_1,\ldots,n_k \in \mathbb{N}^*$ such that $\mathbb{P}_\alpha(\sigma_\alpha = n_\ell) > 0$ for any $\ell \in [1:k]$ and $n = n_1 + \ldots + n_k$. Thus, $d'(\alpha)$ divides $n_1,\ldots,n_k$ and also divides n. Since n is arbitrary in E_α, we deduce that $d'(\alpha)$ divides $d(\alpha)$. This completes the proof. ■

Time schedule (Note 2): Session 7

LEMMA 1.12 . There exists $m \in \mathbb{N}$ such that for any $n \geqslant m$, $nd(\alpha) \in E_\alpha$.

PROOF. First note that E_α is closed under addition. Specifically, if $m,n \in E_\alpha$, then

$$P^{m+n}(\alpha,\alpha) \geqslant \underbrace{P^m(\alpha,\alpha)}_{>0}\underbrace{P^n(\alpha,\alpha)}_{>0}$$

and hence, $m+n \in E_\alpha$. Next, define $d = \text{g.c.d}\{n > 0 : P^n(\alpha,\alpha) > 0\}$. There exist $n_1,\ldots,n_s \in E_\alpha$ such that $d = \text{g.c.d}(n_1,\ldots,n_s)$. Hence, by Bezout's theorem, there exist $a_1,\ldots,a_s \in \mathbb{Z}$ such that

$$d = \sum_{i=1}^{s} a_i n_i = \underbrace{\sum_{i=1}^{s} a_i^+ n_i}_{q} - \underbrace{\sum_{i=1}^{s} a_i^- n_i}_{p} = q - p$$

where $q,p \in E_\alpha$ because E_α is closed under addition. Since $p \in E_\alpha$, there exists $k \in \mathbb{N}$ such that $p = kd$ and hence $q = d + p = (k+1)p$. Now, let $n \geqslant k^2$, then there exists $r \in [0:(k-1)]$ and $m \geqslant k > r$ such that $n = mk + r$. Hence,

$$nd = (mk+r)(q-p) = m\underbrace{kq}_{(k+1)p} + rq - p(mk+r) = \underbrace{(m-r)p}_{\in E_\alpha} + \underbrace{rq}_{\in E_\alpha}$$

and the proof is concluded. ■

PROPOSITION 1.13 . If α,β are two accessible atoms, then $d(\alpha) = d(\beta)$

PROOF. We will prove that $d(\alpha)$ divides $d(\beta)$, which, by symmetry upon interchanging α and β, will complete the proof. For any $n \in E_\beta$, $P^n(\beta,\beta) > 0$. Moreover, since α,β are accessible, there exist ℓ,m such that $P^\ell(\alpha,\beta) > 0$ and $P^m(\beta,\alpha) > 0$. Therefore, the following inequalities holds:

$$P^{\ell+m}(\alpha,\alpha) \geqslant P^\ell(\alpha,\beta)P^m(\beta,\alpha) > 0$$
$$P^{\ell+n+m}(\alpha,\alpha) \geqslant P^\ell(\alpha,\beta)P^n(\beta,\beta)P^m(\beta,\alpha) > 0$$

These two inequalities imply that $d(\alpha)$ divides both $\ell+m$ and $\ell+n+m$. Consequently, $d(\alpha)$ must divide n. Since n is arbitrary in E_β, we conclude that $d(\alpha)$ divides $d(\beta)$.

■

DEFINITION 1.14 . We say that P has a period d if the period of any accessible atom is d. In the particular case where $d = 1$, we say that the Markov kernel P is *aperiodic*.

DEFINITION 1.15 . An atom α is said to be:

(i) *positive* if $\mathbb{E}_\alpha[\sigma_\alpha] < \infty$.

(ii) *null-recurrent* if it is recurrent and $\mathbb{E}_\alpha[\sigma_\alpha] = \infty$.

►**Q-1.13.** Don't we call it a positive-recurrent atom? It seems more consistent with the "null-recurrent" terminology, right?

Actually, we do use that term sometimes. However, $\mathbb{E}_\alpha[\sigma_\alpha] < \infty$ implies $\mathbb{P}_\alpha(\sigma_\alpha < 1) = 1$. Therefore, a positive atom is always recurrent. So, saying "positive-recurrent atom" would be somewhat redundant — "positive atom" suffices.

PROPOSITION 1.16 . Let P be a Markov kernel with an accessible atom α.

- α is positive $\Leftrightarrow$ P admits an invariant probability measure π.

Moreover, in that case, the invariant probability measure is unique and, denoting it by π, we have:

$$\pi(f) = \frac{\mathbb{E}_\alpha[\sum_{k=1}^{\sigma_\alpha} f(X_k)]}{\mathbb{E}_\alpha[\sigma_\alpha]}, \quad f \in \mathscr{F}_+(\mathsf{X})$$

PROOF.

$\Rightarrow$ Assuming that α is positive, we get $\mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1$. Define $\mu \in \mathscr{M}_+(\mathsf{X})$ by $\mu(f) = \mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} f(X_k)\right]$ for any $f \in \mathsf{F}_\mathsf{b}^+(\mathsf{X})$. Then,

$$\mu(Pf) = \mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} Pf(X_k)\right] = \sum_{k=1}^{\infty} \mathbb{E}_\alpha[Pf(X_k)\mathbf{1}_{\{k\leqslant\sigma_\alpha\}}] \stackrel{(1)}{=} \sum_{k=1}^{\infty} \mathbb{E}_\alpha[f(X_{k+1})\mathbf{1}_{\{k\leqslant\sigma_\alpha\}}] = \mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} f(X_{k+1})\right]$$
$$= \mathbb{E}_\alpha\left[\sum_{\ell=2}^{\sigma_\alpha} f(X_\ell)\right] + \underbrace{\mathbb{E}_\alpha\left[f(X_{\sigma_\alpha+1})\right]}_{\mathbb{E}_\alpha\left[\mathbb{E}_{X_{\sigma_\alpha}}[f(X_1)]\right]} = \mathbb{E}_\alpha\left[\sum_{\ell=2}^{\sigma_\alpha} f(X_\ell)\right] + \mathbb{E}_\alpha[f(X_1)] = \mu(f)$$

Here, $\stackrel{(1)}{=}$ follows from the Markov property, and for the underbraced term, the strong Markov property is applied at the stopping time σ_α, which is valid since $\mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1$. Finally, μ is an invariant measure for P, and since $\mu(\mathbf{1}_\mathsf{X}) = \mathbb{E}_\alpha[\sigma_\alpha] < \infty$, we can define $\pi = \mu/\mu(\mathbf{1}_\mathsf{X}) \in \mathscr{M}_1(\mathsf{X})$, which is therefore an invariant probability measure for P.

$\Leftarrow$ Next, assume that $\pi P = \pi$ for some $\pi \in \mathscr{M}_1(\mathsf{X})$. Since α is accessible and hence π-accessible, Kac's theorem –Theorem 1.17– shows that for any $f \in \mathsf{F}_\mathsf{b}^+(\mathsf{X})$,

$$\pi(f) = \int_\alpha \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_\alpha} f(X_k)\right] = \pi(\alpha)\mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} f(X_k)\right]$$

Taking $f = 1$, we obtain $1 = \pi(\alpha)\mathbb{E}_\alpha[\sigma_\alpha]$, which implies $\mathbb{E}_\alpha[\sigma_\alpha] < \infty$ (and hence, α is positive), provided that $\pi(\alpha) > 0$. However, this follows directly from the fact that α is accessible. Indeed, α being accessible, we have for all $x \in \mathsf{X}$, $0 < \sum_{k=0}^{\infty} P^k(x,\alpha)$. Integrating with respect to the invariant probability measure π gives

$$0 < \int \pi(\mathrm{d}x) \sum_{k=0}^{\infty} P^k(x,\alpha) = \sum_{k=0}^{\infty} \pi P^k(\alpha) = \sum_{k=0}^{\infty} \pi(\alpha)$$

proving that $\pi(\alpha) > 0$.

It remains to show the last part of the theorem. From the inspection of the proof of $\Leftarrow$, we observe that any invariant probability distribution π for P satisfies:

$$\pi(f) = \pi(\alpha)\mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} f(X_k)\right], \quad 1 = \pi(\alpha)\mathbb{E}_\alpha[\sigma_\alpha], \quad f \in \mathsf{F}_\mathsf{b}^+(\mathsf{X}).$$

Hence,

$$\pi(f) = \frac{\mathbb{E}_\alpha[\sum_{k=1}^{\sigma_\alpha} f(X_k)]}{\mathbb{E}_\alpha[\sigma_\alpha]}$$

showing that the invariant probability measure is unique. ■

►**Q-1.14.** That's a great result! I really like it!

Yes, and you'll see that we will often use both implications, $\Rightarrow$ and $\Leftarrow$. In particular, according to the proposition, if you can prove that there exists an invariant probability measure for your Markov kernel, then all the accessible atoms are positive, which is amazing. The Kac theorem is also remarkable; it's more general than in the context of kernels with accessible atoms. Let's state and prove the theorem, as it will be very useful in many contexts.

THEOREM 1.17 (► *Kac's theorem).* Let P be a Markov chain with an invariant probability measure π. Assume that C is π-accessible, i.e. $\mathbb{P}_x(\sigma_C < \infty)$ for π-almost all $x \in \mathsf{X}$. Then,

$$\pi = \pi_C^0 = \pi_C^1$$

where

$$\pi_C^0(f) = \int_C \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C - 1} f(X_k)\right]$$

$$\pi_C^1(f) = \int_C \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_C} f(X_k)\right]$$

PROOF. We first show that $\pi = \pi_C^0 \Rightarrow \pi = \pi_C^1$. Assume that $\pi = \pi_C^0$. Then, $\pi = \pi P = \pi_C^0 P$. But for any $f \in \mathsf{F}_\mathsf{b}^+$,

$$\pi_C^0(Pf) = \int_C \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} Pf(X_k)\right] = \int_C \pi(\mathrm{d}x) \sum_{k=0}^{\infty} \mathbb{E}_x\left[Pf(X_k)\mathbf{1}_{\{k<\sigma_C\}}\right]$$

$$= \int_C \pi(\mathrm{d}x) \sum_{k=0}^{\infty} \mathbb{E}_x\left[f(X_{k+1})\mathbf{1}_{\{k<\sigma_C\}}\right] = \int_C \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{\ell=1}^{\sigma_C} f(X_\ell)\right] = \pi_C^1(f)$$

Hence, $\pi = \pi_C^0 P = \pi_C^1$ and we have proved that $\pi = \pi_C^0 \Rightarrow \pi = \pi_C^1$. To complete the proof it only remains to show $\pi = \pi_C^0$. By the last-exit decomposition and the Markov property, for all $f \in \mathsf{F}_\mathsf{b}^+(\mathsf{X})$ and all $n \geqslant 1$,

$$\begin{aligned}\pi(f) &= \mathbb{E}_\pi[f(X_n)] = \mathbb{E}_\pi[f(X_n)\mathbf{1}_{\{\sigma_C \leqslant n\}}] + \mathbb{E}_\pi[f(X_n)\mathbf{1}_{\{\sigma_C > n\}}] \\ &= \sum_{\ell=1}^{n} \underbrace{\mathbb{E}_\pi\left[f(X_n)\mathbf{1}_c(X_\ell)\prod_{k=\ell+1}^{n}\mathbf{1}_{c^c}(X_k)\right]}_{\mathbb{E}_\pi\left[\mathbf{1}_C(X_\ell)\mathbb{E}_{X_\ell}\left[f(X_{n-\ell})\prod_{k=1}^{n-\ell}\mathbf{1}_{C^c}(X_k)\right]\right]} + \mathbb{E}_\pi[f(X_n)\mathbf{1}_{\{\sigma_C > n\}}]\end{aligned}$$

Noting that π is invariant and setting $k = n - \ell$, we finally get

$$\begin{aligned}\pi(f) &= \sum_{k=0}^{n-1} \int_C \pi(\mathrm{d}x)\mathbb{E}_x[f(X_k)\mathbf{1}_{\{\sigma_C > k\}}] + \mathbb{E}_\pi[f(X_n)\mathbf{1}_{\{\sigma_C > n\}}] \\ &= \int_C \pi(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{(n-1)\wedge(\sigma_C-1)} f(X_k)\right] + \mathbb{E}_\pi[f(X_n)\mathbf{1}_{\{\sigma_C > n\}}]\,. \\ &= A_n(f) + B_n(f) \end{aligned} \tag{1.5}$$

Using first $B_n(f) \geqslant 0$ and then the monotone convergence, we get

$$\pi(f) \geqslant A_n(f) \to_{n\to\infty} \pi_C^0(f)$$

Hence, $\Delta\pi = \pi - \pi_C^0$ is a non-negative measure, i.e. $\Delta\pi \in \mathscr{M}_+(\mathsf{X})$. To obtain $\Delta\pi = 0$, we only need to show that $\Delta\pi(\mathsf{X}) = 0$. Define $h(x) = \mathbb{P}_x(\sigma_C < \infty)$ and $D = \{h = 0\}$.

- By assumption, $\pi(D) = 0$ and hence $\pi_C^0(D) = 0$ since $\pi \geqslant \pi_C^0$. This implies $\Delta\pi(D) = 0$.
- Note that $D^c = \{h > 0\}$. We will show that $\Delta\pi(D^c) = 0$. We have

$$\begin{aligned} B_n(h) = \mathbb{E}_\pi[h(X_n)\mathbf{1}_{\{\sigma_C > n\}}] &= \mathbb{E}_\pi[\mathbb{P}_{X_n}(\sigma_C < \infty)\mathbf{1}_{\{\sigma_C > n\}}] = \mathbb{E}_\pi[\mathbf{1}_{\{\sigma_C \circ \theta_n < \infty\}}\mathbf{1}_{\{\sigma_C > n\}}] \\ &\leqslant \mathbb{E}_\pi[\mathbf{1}_{\{\sigma_C < \infty\}}\mathbf{1}_{\{\sigma_C > n\}}] \to_{n\to\infty} \mathbb{P}_\pi(\sigma_C < \infty, \sigma_C = \infty) = 0 \end{aligned}$$

where we have used the Markov property at time n for the second equality and the monotone convergence for the last limit. Combining with (1.5) applied to $f = h$ and letting n goes to infinity, we get $\pi(h) = \pi_C^0(h)$ and hence $\Delta\pi(h) = 0$, which, in turn, implies $\Delta\pi(\{h > 0\}) = \Delta\pi(D^c) = 0$.

Finally, $\Delta\pi(\mathsf{X}) = \Delta\pi(D) + \Delta\pi(D^c) = 0$, hence $\Delta\pi$ is the null measure and $\pi = \pi_C^0$ and the proof is completed. ■

►**Q-1.15.** The proof is just a pleasure to follow. Should we use π_C^0 or π_C^1?

In a general context, both are important. However, when $C = \alpha$ is an atom, $\mathbb{E}_x\left[\sum_{k=1}^{\sigma_\alpha} f(X_k)\right]$ does not depend on $x \in \alpha$. In contrast, $\mathbb{E}_x[\sum_{k=0}^{\sigma_\alpha - 1} f(X_k)]$ depends on $x \in \alpha$ because the term corresponding to $k = 0$ in the sum is $f(X_0) = f(x)$, which is **not expected** to be constant for $x \in \alpha$. Therefore, when applying Kac's theorem with an atom, we typically prefer to use π_α^1 rather than π_α^0.

1.3 Excursions

Time schedule (**Note** 3): Session 8

THEOREM 1.18 . If P admits a recurrent atom α. Let $Z_0, \ldots, Z_k$ be $\mathscr{F}_{\sigma_\alpha}$-measurable random variables such that $\mathbb{E}_x[Z_i]$ does not depend on $x \in \alpha$. Then, for all $\xi \in \mathscr{M}_1(\mathsf{X})$ such that $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$, we have

$$\mathbb{E}_\xi\left[\prod_{i=0}^k Z_i \circ \theta_{\sigma_\alpha^i}\right] = \mathbb{E}_\xi[Z_0]\prod_{i=1}^k \mathbb{E}_\alpha[Z_i] \tag{1.6}$$

PROOF. First note that $\sigma_\alpha^i = \sigma_\alpha + \sigma_\alpha^{i-1} \circ \theta_{\sigma_i}$. We will prove (1.6) by induction on k. Indeed the case $k = 0$ is clear. Next assume that (1.6) holds with k replaced by $k - 1$. Then,

$$\begin{aligned} \mathbb{E}_\xi\left[\prod_{i=0}^k Z_i \circ \theta_{\sigma_\alpha^i}\right] &= \mathbb{E}_\xi\left[Z_0 \prod_{i=1}^k Z_i \circ \theta_{\sigma_\alpha^{i-1}} \circ \theta_{\sigma_\alpha}\right] \overset{(1)}{=} \mathbb{E}_\xi\left[Z_0 \mathbb{E}_\alpha\left[\prod_{i=1}^k Z_i \circ \theta_{\sigma_\alpha^{i-1}}\right]\right] \\ &= \mathbb{E}_\xi[Z_0]\,\mathbb{E}_\alpha\left[\prod_{\ell=0}^{k-1} Z_{\ell+1} \circ \theta_{\sigma_\alpha^\ell}\right] \\ &\overset{(2)}{=} \mathbb{E}_\xi[Z_0]\prod_{\ell=0}^{k-1}\mathbb{E}_\alpha[Z_{\ell+1}] = \mathbb{E}_\xi[Z_0]\prod_{i=1}^k \mathbb{E}_\alpha[Z_i] \end{aligned}$$

where $\overset{(1)}{=}$ follows from the strong Markov property at σ_α, which is valid since $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$ and $\overset{(2)}{=}$ follows from the induction assumption at $k - 1$, which is valid since $\mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1$ (indeed, the induction assumption is applied with an initial distribution concentrated on α). ■

►**Q-1.16.** It is a bit hard to follow. How do you use this theorem?

We will apply it to get the independence of the excursions between successive visits to the atom.

►**Q-1.17.** Could you be more specific?

Of course. For example, for any $f \in \mathsf{F}_\mathsf{b}(\mathsf{X})$, we can apply the theorem to

$$Z_0 = H_0\left(\sum_{k=0}^{\tau_\alpha} f(X_k)\right)$$
$$Z_n = H_n\left(\sum_{k=1}^{\sigma_\alpha} f(X_k)\right), \quad n \geqslant 1$$

where $\{H_i \,:\, i \in [0:k]\}$ are arbitrary bounded measurable random variables on $\mathbb{R}$. Then, set $\varepsilon_0(\alpha, f) = \sum_{k=0}^{\tau_\alpha} f(X_k)$ and for $n \geqslant 1$, set

$$\varepsilon_n(\alpha, f) = Z_n \circ \theta_{\sigma_\alpha^n} = \sum_{k=\sigma_\alpha^n+1}^{\sigma_\alpha^{n+1}} f(X_k)$$

Corollary 1.19 . Let P be a Markov kernel with a recurrent atom α.

- Under $\mathbb{P}_\alpha$, the random variables $\{\varepsilon_n(\alpha, f) \,:\, n \geqslant 1\}$ are i.i.d.
- If $\mu \in \mathscr{M}_1(\mathsf{X})$ satisfies $\mathbb{P}_\mu(\sigma_\alpha < \infty) = 1$, then, under $\mathbb{P}_\mu$,
  - the random variables $\{\varepsilon_n(\alpha, f) \,:\, n \in \mathbb{N}\}$ are independent
  - the random variables $\{\varepsilon_n(\alpha, f) \,:\, n \geqslant 1\}$ are i.i.d.

►**Q-1.18.** I got the idea, now. We can decompose for example $\sum_{i=0}^{n-1} f(X_i)$ into blocks of the form $\sum_{k=\sigma_\alpha^n+1}^{\sigma_\alpha^{n+1}} f(X_k)$ which are iid... So you can easily treat the LLN and the CLT by using properties on iid random variables...

Exactly... But for the moment, let us deal with $\|P^n(x,\cdot) - \pi\|_{TV}$ which is less directly linked with the blocks $\sum_{k=\sigma_\alpha^n+1}^{\sigma_\alpha^{n+1}} f(X_k)$. Actually, we will need only that the waiting time between successive visits to α are iid. This is again obtained from Corollary 1.19 by setting $f = \mathbf{1}_\mathsf{X}$. In what follows, define the **visit times** $\left\{v_C^n \,:\, n \in \mathbb{N}\right\}$ to the set C by $v_C^0 = \tau_C$ and for $n \geqslant 1$, $v_C^n = v_C^{n-1} + \sigma_C \circ \theta_{v_C^{n-1}}$, and hence

$$\Delta v_\alpha^n = \begin{cases} v_C^0 = \tau_C & \text{if } n = 0 \\ v_C^n - v_C^{n-1} = \sigma_C \circ \theta_{v_C^{n-1}} & \text{if } n \geqslant 1 \end{cases} \tag{1.7}$$

1.3.1 Coupling

Define $\bar{X} = \mathsf{X} \times \mathsf{X}$ and $\bar{\mathscr{X}} = \mathscr{X} \otimes \mathscr{X}$. Let $\bar{P}$ be a Markov kernel on $\bar{\mathsf{X}} \times \bar{\mathscr{X}}$ such that for any $\bar{x} = (x, x') \in \bar{X}$ and $A \in \mathscr{X}$,

$$\bar{P}(\bar{x}, A \times \mathsf{X}) = P(x, A), \quad \bar{P}(\bar{x}, \mathsf{X} \times A) = P(x', A)$$

For any $\bar{\xi} \in \mathscr{M}(\bar{\mathsf{X}})$, we define $\bar{\mathbb{P}}_{\bar{\xi}}$ as the probability measure induced on $(\bar{\mathsf{X}}^\mathbb{N}, \bar{\mathscr{X}}^{\otimes \mathbb{N}})$ by the Markov kernel $\bar{P}$ and initial distribution $\bar{\xi}$. For $\omega = \{\bar{x}_n = (x_n, x_n') \,:\, n \in \mathbb{N}\} \in \bar{\mathsf{X}}^\mathbb{N}$, we define $\bar{X}_n(\omega) = \bar{x}_n$, $X_n(\omega) = x_n$ and $X_n'(\omega) = x_n'$.

Lemma 1.20 (► *Coupling inequality for atomic chains*). Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with an atom α. Then for any $\xi, \xi' \in \mathscr{M}_1(\mathsf{X})$,

$$\|\xi P^n - \xi' P^n\|_{TV} \leqslant 2\bar{\mathbb{P}}_{\xi \otimes \xi'}(\sigma_{\bar{\alpha}} \geqslant n)$$

Proof. Define $\bar\alpha = \alpha \times \alpha$. Let f be a measurable function such that $|f| \leqslant 1$. Write

$$\begin{aligned}
|\xi P^n f - \xi' P^n f| &= |\mathbb{E}_\xi[f(X_n)] - \mathbb{E}_{\xi'}[f(X_n)]| = |\bar{\mathbb{E}}_{\xi\otimes\xi'}\left[f(X_n) - f(X'_n)\right]| \\
&= \left|\sum_{k=1}^{n-1} \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbf{1}_{\{\sigma_{\bar\alpha}=k\}}\left(f(X_n) - f(X'_n)\right)\right] + \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbf{1}_{\{\sigma_{\bar\alpha}\geqslant n\}}\left(f(X_n) - f(X'_n)\right)\right]\right| \\
&\leqslant \left|\sum_{k=1}^{n-1} \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbf{1}_{\{\sigma_{\bar\alpha}=k\}}\left(P^{n-k}f(\alpha) - P^{n-k}f(\alpha)\right)\right]\right| + \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbf{1}_{\{\sigma_{\bar\alpha}\geqslant n\}}|f(X_n) - f(X'_n)|\right] \\
&\leqslant 2\bar{\mathbb{P}}_{\xi\otimes\xi'}(\sigma_{\bar\alpha} \geqslant n)
\end{aligned}$$

The proof is completed. ■

1.4 Residual life time kernel

Define $\mathsf{X} = \mathbb{N}^*$ and $\mathscr{X} = \mathscr{P}(\mathsf{X})$. Let $\nu \in \mathscr{M}_1(\mathsf{X})$. Let Q be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ defined by: for $k \in \mathbb{N}^*$

$$\begin{aligned}
Q(k, k-1) &= 1\,, \quad \text{if } k > 1 \\
Q(1, k) &= \nu(k)
\end{aligned}$$

where by abuse of notation, we write for any kernel Q on a discrete space X, $Q(k,\ell) = Q(k,\{\ell\})$ for all $k,\ell \in \mathsf{X}$. Define the independent coupling kernel $\bar{Q}(\bar{x},\bar{y}) = Q(x,y)Q(x',y')$ for all $(\bar{x},\bar{y}) \in \bar{X}^2$ where we have used the notation $\bar{x} = (x,x')$, $\bar{y} = (y,y')$ and $\bar{\mathsf{X}} = \mathsf{X} \times \mathsf{X}$. We write $\bar{\mathscr{X}} = \mathscr{X} \otimes \mathscr{X}$.

Then, $\bar{Q}$ is a Markov kernel on $\bar{\mathsf{X}} \times \bar{\mathscr{X}}$ and we define by $\bar{\mathbb{Q}}_\xi$ the probability measure induced on $(\bar{\mathsf{X}}^{\mathbb{N}}, \bar{\mathscr{X}}^{\otimes\mathbb{N}})$ by the Markov kernel $\bar{Q}$ and initial distribution ξ on $(\bar{\mathsf{X}}, \bar{\mathscr{X}})$. As usual, for convenience, for any $\bar{x} \in \bar{\mathsf{X}}$, we write $\bar{\mathbb{Q}}_{\bar{x}} = \bar{\mathbb{Q}}_{\delta_{\bar{x}}}$ and $\bar{\mathbb{E}}_{\bar{x}} = \bar{\mathbb{E}}_{\delta_{\bar{x}}}$. To alleviate notation, we write $\sigma_{\bar{x}}$ or $N_{\bar{x}}$ for any $\bar{x} \in \bar{\mathsf{X}}$ instead of $\sigma_{\{\bar{x}\}}$ or $N_{\{\bar{x}\}}$.

Proposition 1.21 . Assume that

(a) $\mathrm{g.c.d}\{\mathrm{k} \in \mathbb{N}^* \ :\ \nu(\mathrm{k}) \neq 0\} = 1$

(b) $m_\nu = \sum_{k=1}^\infty k\nu(k) < \infty$

Then, for any $\bar{k} \in \mathbb{N}^* \times \mathbb{N}^*$, $\bar{\mathbb{Q}}_{\bar{k}}(\sigma_{(1,1)} < \infty) = 1$.

Proof. Let $M = \sup\{k \in \mathbb{N}^* \ :\ \nu(k) > 0\}$ and define $B = [1:M]$ and $\bar{B} = B \times B$. Assume first that

($\star$) any singleton in $\bar{B}$ is an accessible recurrent atom for $\bar{Q}$.

Since $(1,1) \in \bar{B}$, the singleton $\{(1,1)\}$ is accessible by ($\star$) and we have $\bar{\mathbb{Q}}_{\bar{k}}(\sigma_{(1,1)} < \infty) > 0$ for any $\bar{k} = (k,k') \in \bar{X}$. We can then apply the solidarity lemma (Lemma 1.1), which yields

$$\bar{\mathbb{Q}}_{\bar{k}}(N_{\bar{k}} = \infty) \leqslant \bar{\mathbb{Q}}_{\bar{k}}(N_{(1,1)} = \infty)$$

If in addition $\bar{k} \in \bar{B}$, the lhs is equal to 1 since by ($\star$), $\{\bar{k}\}$ is a recurrent atom and we finally have $\bar{\mathbb{Q}}_{\bar{k}}(N_{(1,1)} = \infty) = 1$, which in turn show that

$$\bar{\mathbb{Q}}_{\bar{k}}(\sigma_{(1,1)} < \infty) = 1\,, \quad \bar{k} \in \bar{B} \tag{1.8}$$

The rest of the proof consists in showing ($\star$) and in extending (1.8) to any $\bar{k} \in \bar{\mathsf{X}}$. We start with ($\star$) which will be proved through several steps.

- **Step 1: $\bar{Q}$ admits an invariant probability measure $\bar\pi$.** Since for any $k \in \mathbb{N}^*$, $Q^{k-1}(k,1) = 1 > 0$, the singleton $\{1\}$ is an accessible atom for Q. Moreover,

$$\mathbb{E}_1[\sigma_{\{1\}}] = \mathbb{E}_1[X_1] = \sum_{k=1}^\infty k\nu(k) = m_\nu < \infty$$

showing that $\{1\}$ is positive. According to Proposition 1.16, Q admits a (unique) invariant probability measure π. It implies that $\bar{\pi} = \pi \otimes \pi$ is an invariant probability measure for the independent coupling kernel $\bar{Q}$.

- **Step 2: $\{(1,1)\}$ is an accessible atom for $\bar{Q}$.** Since $\mathbb{Q}_1(\sigma_1 = k) = \mathbb{Q}_1(X_1 = k) = \nu(k)$, assumption (a) and Lemma 1.11 imply that Q is aperiodic. Then, Lemma 1.12 show that there exists $m \in \mathbb{N}$ such that for any $n \geqslant m$, $Q^n(1,1) > 0$. Moreover, since $\{1\}$ is accessible for Q, for any $(k,k') \in \bar{B}$, there exist $\ell, \ell' \in \mathbb{N}$ such that $Q^\ell(k,1) > 0$ and $Q^{\ell'}(k',1) > 0$. Hence, choosing $n, n' \geqslant m$ such that $\ell + n = \ell' + n'$, we have

$$\bar{Q}^{\ell+n}((k,k'),(1,1)) = Q^{\ell+n}(k,1)Q^{\ell'+n'}(k',1) \geqslant Q^\ell(k,1)Q^n(1,1)Q^{\ell'}(k',1)Q^{n'}(1,1) > 0$$

where the rhs is positive as a product of positive terms.

- **Step 3: any singleton in $\bar{B}$ is an accessible atom for $\bar{Q}$.** Let $\bar{k} = (k,k') \in \bar{B}$. Considering **Step 2** and Lemma 1.4, it is sufficient to show that $\bar{\mathbb{Q}}_{(1,1)}(\sigma_{\bar{k}} < \infty) > 0$. To do so, let $k \in B$, then by definition of M, there exists $\ell \geqslant k$ such that $\nu(\ell) > 0$. Hence

$$Q^{\ell-k+1}(1,k) \geqslant \mathbb{Q}_1(X_1 = \ell, X_{\ell-k+1} = k) = \mathbb{Q}_1(X_1 = \ell) = \nu(\ell) > 0.$$

Hence for any $\bar{k} = (k,k') \in B^2$, there exist $(\ell,\ell') \in \mathbb{N}^* \times \mathbb{N}^*$ such that $Q^\ell(1,k) > 0$ and $Q^{\ell'}(1,k') > 0$. Using again Lemma 1.12 with the aperiodic kernel Q, there exists $m \in \mathbb{N}$ such that for any $n \geqslant m$, $Q^n(1,1) > 0$. Hence, choosing $n, n' \geqslant m$ such that $n + \ell = n' + \ell'$, we obtain

$$\bar{Q}^{n+\ell}((1,1),\bar{k}) \geqslant Q^{n+\ell}(1,k)Q^{n'+\ell'}(1,k') \geqslant Q^n(1,1)Q^\ell(1,k)Q^{n'}(1,1)Q^{\ell'}(1,k') > 0$$

showing that $\bar{\mathbb{Q}}_{(1,1)}(\sigma_{\bar{k}} < \infty) > 0$ and the proof of **Step 3** is completed.

Considering **Step 1** and **Step 2**, Proposition 1.16 applied to $P = \bar{Q}$ shows that any singleton in $\bar{B}$ is positive and hence recurrent. This proves $(\star)$. To complete the proof, it remains to show that (1.8) holds for all $\bar{k} \in \bar{\mathsf{X}}$.

First note that B is absorbing for Q. Indeed if $k \in B$ and $k > 1$, then $k-1 \in B$ and hence $Q(k,B) \geqslant Q(k,k-1) = 1$. Moreover, with $k = 1 \in B$, $Q(1,B) = \nu(B) = \nu(\mathsf{X}) = 1$. Finally, $Q(k,B) = 1$ for all $k \in B$ and B is therefore absorbing. Since $Q^k(k,1) = 1$ and $1 \in B$, we deduce for all $n \geqslant k$, $Q^n(k,B) = 1$. Hence for all $n \geqslant k \vee k'$, $\bar{Q}^n(\bar{k},\bar{B}) = 1$. This implies (using (1.8)) that for all $\bar{k} \in \bar{\mathsf{X}}$, taking $n \geqslant k \vee k'$,

$$\begin{aligned}\mathbb{Q}_{\bar{k}}(\sigma_{(1,1)} < \infty) &\geqslant \mathbb{Q}_{\bar{k}}(\bar{X}_n \in \bar{B}, \sigma_{(1,1)} \circ \theta_n < \infty) \\ &= \sum_{\bar{\ell} \in \bar{B}} \bar{Q}^n(\bar{k},\bar{\ell}) \underbrace{\mathbb{Q}_{\bar{\ell}}(\sigma_{(1,1)} < \infty)}_{1} = \bar{Q}^n(\bar{k},\bar{B}) = 1\end{aligned}$$

which completes the proof. ■

►**Q-1.19.** Amazing! If I understand well, the assumption on the aperiodicity of Q is only used to get accessibility for the coupling kernel $\bar{Q}$?

You are perfectly right. If, by a convenient choice of the kernel coupling $\bar{Q}$, you can obtain that $\bar{Q}$ is accessible, then, there is no need of the aperiodicity assumption on Q.

THEOREM 1.22 . Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with an aperiodic, accessible positive atom α. Then for any $\xi \in \mathscr{M}_1(\mathsf{X})$ such that $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$,

$$\lim_{n\to\infty} \|\xi P^n - \pi\|_{TV} = 0$$

PROOF. Choosing the independent coupling kernel and applying Lemma 1.20, we obtain

$$\|\xi P^n - \pi\|_{TV} \leqslant 2\bar{\mathbb{P}}_{\xi\otimes\xi'}(\sigma_{\alpha\times\alpha} \geqslant n)$$

Hence, to obtain to that the rhs converges to 0, we will show that $\bar{\mathbb{P}}_{\xi\otimes\xi'}(\sigma_{\alpha\times\alpha} < \infty) = 1$. The event $\{\sigma_{\alpha\times\alpha} < \infty\}$ only depend on the visit times of each marginal chain. Denote by $\{\nu_\alpha^k : k \in \mathbb{N}\}$, resp. $\{\nu'^k_\alpha : k \in \mathbb{N}\}$, the visit times to the set α (see (1.7)) for $\{X_n : n \in \mathbb{N}\}$ and $\{X'_n : n \in \mathbb{N}\}$. Define the residual life-time processes (also called the age processes) by

$$A_n = \inf\left\{ \nu_\alpha^k \ : \ \nu_\alpha^k > n \text{ and } k \in \mathbb{N} \right\} - n$$
$$A'_n = \inf\left\{ \nu'^k_\alpha \ : \ \nu'^k_\alpha > n \text{ and } k \in \mathbb{N} \right\} - n$$

Then, under $\bar{\mathbb{P}}_{\xi \otimes \xi'}$, the age processes are independent and each of them is marginally a Markov chain with transition kernel Q defined by

$$Q(k, k-1) = 1\,, \quad \text{if } k > 1$$
$$Q(1, k) = \mathbb{P}_\alpha(\sigma_\alpha = k)$$

starting from a_ξ for $\{A_n \ : \ n \in \mathbb{N}\}$ and $a_{\xi'}$ for $\{A'_n \ : \ n \in \mathbb{N}\}$ where a_ξ is the probability defined on $\mathbb{N}^*$ by $a_\xi(k) = \mathbb{P}_\xi(A_0 = k) = \mathbb{P}_\xi(\sigma_\alpha = k)$. Since $X_n \in \alpha \Leftrightarrow A_{n-1} = 1$, we obtain that

$$\inf\left\{ n \ : \ (X_n, X'_n) \in \alpha \times \alpha \right\} = \inf\left\{ k \ : \ (A_k, A'_k) = (1,1) \right\} + 1$$

and consequently, $\bar{\mathbb{P}}_{\xi \otimes \xi'}(\sigma_{\alpha \times \alpha} < \infty) = \bar{\mathbb{Q}}_{a_\xi \otimes a_{\xi'}}(\sigma_{(1,1)} < \infty) = \sum_{\bar{k}} a_\xi \otimes a_{\xi'}(\bar{k}) \bar{\mathbb{Q}}_{\bar{k}}(\sigma_{(1,1)} < \infty) = 1$. The latter equality follows from Proposition 1.21, which is applicable because α, being an aperiodic positive atom, satisfies the following:

(a) $\text{g.c.d}\,\{\text{k} \in \mathbb{N}^* \ : \ \mathbb{P}_\alpha(\sigma_\alpha = \text{k}) \neq 0\} = 1$

(b) $\sum_{k=1}^{\infty} k \mathbb{P}_\alpha(\sigma_\alpha = k) = \mathbb{E}_\alpha[\sigma_\alpha] < \infty$

The proof is complete. ■

Chapter 2

Small sets, splitting, irreducibility, geometric ergodicity

Time schedule (**Note** 4): Session 9

Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$.

DEFINITION 2.1 . We say that C is a small set for P if there exist a integer $n \in \mathbb{N}^*$, a constant $\varepsilon > 0$, a non-null measure $\nu \in \mathscr{M}_1(\mathsf{X})$ such that

$$P^n(x, \cdot) \geqslant \varepsilon\nu(\cdot), \quad \forall x \in C. \tag{2.1}$$

In that case, we also say that C is a n-small set or a $(n, \varepsilon\nu)$-small set for P.

The existence of small sets allows to work with atomic Markov chains by adding a second component to the Markov chain. This construction is known as the *splitting* construction. Assume that C is a $(1, \varepsilon\nu)$-small set, i.e., there exist $\varepsilon > 0$ and $\nu \in \mathscr{M}_1(\mathsf{X})$ such that

$$P(x, \cdot) \geqslant \varepsilon\nu(\cdot), \quad \forall x \in C.$$

Let us consider an informal description of the transition associated to the splitting Markov chain on $\mathsf{X} \times [0,1]$: at time n, we have $\check{X}_n = (X_n, U_n)$ and we wish to draw $\check{X}_{n+1}$ for $\check{X}_n$. Define for $x \in C$

$$R(x, \mathrm{d}y) = \frac{P(x, \mathrm{d}y) - \varepsilon\nu(\mathrm{d}y)}{1 - \varepsilon}.$$

Then, the transition can be described as below:

Condition	**Action**
$\check{X}_n \in C \times [0, \varepsilon]$	Draw $\check{X}_{n+1} \sim \nu \otimes \mathrm{Unif}([0,1])$.
$\check{X}_n \in C \times (\varepsilon, 1]$	Draw $\check{X}_{n+1} \sim R(X_n, \cdot) \otimes \mathrm{Unif}([0,1])$.
$X_n \notin C$	Draw $\check{X}_{n+1} \sim P(X_n, \cdot) \otimes \mathrm{Unif}([0,1])$.

Table 2.1: Transition of the splitting chain.

More formally, with the notation $\check{x} = (x, u)$ and $\check{x}' = (x', u')$, we define the Markov kernel $\check{P}$ by

$$\check{P}(\check{x}, \mathrm{d}\check{x}') = \left[\mathbf{1}_{C\times[0,\varepsilon]}(\check{x})\nu(\mathrm{d}x') + \mathbf{1}_{C\times(\varepsilon,1]}(\check{x})R(x,\mathrm{d}x') + \mathbf{1}_{C^c}(x)P(x,\mathrm{d}x')\right]\mathbf{1}_{[0,1]}(u)\mathrm{d}u$$

Obviously, $\check{\alpha} = C \times [0, \varepsilon]$ is an atom for $\check{P}$. It is easy to see that when $\{(X_n, U_n)\ :\ n \in \mathbb{N}\}$ is a Markov chain with Markov kernel $\check{P}$ and initial distribution $\xi \otimes \mathrm{Unif}([0, 1])$, then the sequence $\{X_n\ :\ n \in \mathbb{N}\}$ is a Markov chain with Markov kernel P and initial distribution ξ.

►**Q-2.1.** Since $\{(X_n, U_n)\ :\ n \in \mathbb{N}\}$ forms an atomic Markov chain, we can leverage all the results from the previous chapter! This is fantastic.

I couldn't agree more. The splitting technique allows us to work first with the atomic Markov kernel $\check{P}$ and then transfer the derived properties to P. While we won't delve further into this approach to focus on other methods, I must admit it is an exceptionally powerful tool.

►**Q-2.2.** Before we move on from splitting techniques, what happens if C is an n-small set rather than a 1-small set?

In that case, we work directly with the Markov kernel P^n instead of P, and subsequently transfer the properties from P^n back to P.

2.1 Irreducibility and uniqueness of the invariant probability measure.

DEFINITION 2.2 . The Markov kernel P is said to be irreducible if and only if it admits an accessible small set.

DEFINITION 2.3 . The Markov kernel P is said to be ϕ-irreducible if and only if ϕ is a non-null measure on $(\mathsf{X}, \mathscr{X})$ such that any set $A \in \mathscr{X}$ such that $\phi(A) > 0$ is accessible for P.

PROPOSITION 2.4 . If $\mathscr{X}$ is countably generated, the Markov kernel P is irreducible if and only if there exists $\phi \in \mathscr{M}_+(\mathscr{X})$ such that P is ϕ-irreducible.

PROOF. We only prove $\Rightarrow$. The proof of the other implication $\Leftarrow$ is more involved and will be admitted in this course. Assume that P is irreducible. Then, it admits an accessible $(n, \varepsilon\nu)$-small set that we call C. Let $A \in \mathscr{X}$ such that $\nu(A) > 0$. We will show that A is accessible. Let $x \in \mathsf{X}$. Since C is accessible, there exists $m \in \mathbb{N}^*$ such that $P^m(x, C) > 0$. Then,

$$P^{m+n}(x, A) \geqslant \int P^m(x, \mathrm{d}y)\mathbf{1}_C(y)P^n(y, A) \geqslant \varepsilon P^m(x, C)\nu(A)$$

and the rhs is positive as the product of positive terms. Finally, A is accessible and hence, P is ν-irreducible. ■

In what follows, we always assume that $\mathscr{X}$ is countably generated.

We now turn to a very simple lemma that will be useful for finding sufficient conditions for uniqueness.

LEMMA 2.5 . If P admits two distinct invariant probability measures, it also admits distinct invariant probability measures π_0 and π_1 that are mutually singular, i.e., such that there exists $A \in \mathscr{X}$ such that $\pi_0(A) = \pi_1(A^c) = 0$.

PROOF. Let ζ_0, ζ_1 be two distinct invariant probability measures for P. Both have densities with respect to some common dominating measure (for example, taking $\zeta = \zeta_1 + \zeta_2$, we have that ζ dominates both ζ_0 and ζ_1, which can be seen from the implication $\zeta(A) = 0 \Rightarrow (\zeta_1(A) = 0$ and $\zeta_2(A) = 0)$ for any $A \in \mathscr{X}$ and according to the Radon Nikodym theorem, if a measure dominates another one, the latter has a density with respect to the former). Write then $\zeta_0(\mathrm{d}x) = f_0(x)\zeta(\mathrm{d}x)$ and $\zeta_1(\mathrm{d}x) = f_1(x)\zeta(\mathrm{d}x)$ where f_0, f_1 are non-negative measurable functions on X. Define the positive part $(\zeta_1 - \zeta_0)^+$and the negative part $(\zeta_1 - \zeta_0)^-$ of the signed measure $\zeta_1 - \zeta_0$ by $(\zeta_1 - \zeta_0)^+(\mathrm{d}x) = [f_1(x) - f_0(x)]^+\zeta(\mathrm{d}x)$ and $(\zeta_1 - \zeta_0)^-(\mathrm{d}x) = [f_1(x) - f_0(x)]^-\zeta(\mathrm{d}x)$. Then,

$$\begin{aligned}(\zeta_1 - \zeta_0)^+ P\mathbf{1}_A &= \int_{\mathsf{X}} \zeta(\mathrm{d}x)[f_1(x) - f_0(x)]^+ P(x,A) \\ &\geqslant \int_{\mathsf{X}} \zeta(\mathrm{d}x)[f_1(x) - f_0(x)] P(x,A) \\ &\geqslant \zeta_1 P(A) - \zeta_0 P(A) = \zeta_1(A) - \zeta_0(A) = [\zeta_1 - \zeta_0](A).\end{aligned}$$

Hence, setting $B = \{f_1 > f_2\}$, we have

$$(\zeta_1 - \zeta_0)^+(A) = (\zeta_1 - \zeta_0)(A \cap B) \leqslant (\zeta_1 - \zeta_0)^+ P\mathbf{1}_{A\cap B} \leqslant (\zeta_1 - \zeta_0)^+ P\mathbf{1}_A$$

Hence, $(\zeta_1 - \zeta_0)^+ \leqslant (\zeta_1 - \zeta_0)^+ P$. The measure $(\zeta_1 - \zeta_0)^+ P - (\zeta_1 - \zeta_0)^+$ is therefore non-negative and we have

$$[(\zeta_1 - \zeta_0)^+ P - (\zeta_1 - \zeta_0)^+](\mathsf{X}) = \int_{\mathsf{X}} (\zeta_1 - \zeta_0)^+(\mathrm{d}x) \underbrace{P(x,\mathsf{X})}_{1} - (\zeta_1 - \zeta_0)^+(\mathsf{X}) = 0.$$

Finally, $(\zeta_1 - \zeta_0)^+ = (\zeta_1 - \zeta_0)^+ P$. The probability measure $\pi_0 = \frac{(\zeta_1 - \zeta_0)^+}{(\zeta_1 - \zeta_0)^+(\mathsf{X})}$ is thus an invariant probability measure for P. Replacing $(\zeta_1 - \zeta_0)^+$ by $(\zeta_1 - \zeta_0)^-$, we obtain in the same way that $\pi_1 = \frac{(\zeta_1 - \zeta_0)^-}{(\zeta_1 - \zeta_0)^-(\mathsf{X})}$ is an invariant probability measure. We can easily check that taking $A = \{f_0 \geqslant f_1\}$, we have $\pi_0(A) = \pi_1(A^c) = 0$, showing that these probability measures are mutually singular. ■

PROPOSITION 2.6 . Any Markov kernel P that is ϕ-irreducbile admits at most one invariant probability measure.

PROOF. The proof is by contradiction. Assume that there exists two distinct invariant probability measures. According to Lemma 2.5, we can consider two invariant probability measures π_1 and π_2 that are mutually singular. Under the assumptions of the Proposition, let $A \in \mathscr{X}$ such that $\phi(A) > 0$. Then, for any $i \in \{1,2\}$,we have

$$0 < \int_{\mathsf{X}} \pi_i(\mathrm{d}x) \underbrace{\sum_{n=0}^{\infty} P^n(x,A)}_{>0} = \sum_{n=0}^{\infty} \pi_i P^n(A) = \sum_{n=0}^{\infty} \pi_i(A),$$

which in turn implies that $\pi_i(A) > 0$. The contraposed implication gives that if for some $i \in \{1,2\}$, $\pi_i(A) = 0$, then $\phi(A) = 0$. Now, since $\{\pi_i : i \in \{1,2\}\}$ are mutually singular, there exists $A \in \mathscr{X}$ such that $\pi_1(A) = \pi_2(A^c) = 0$ and this shows that $\phi(A) = \phi(A^c) = 0$ which is impossible. ■

2.2 Geometric ergodicity

In what follows, we assume that for **some measurable function** $V : \mathsf{X} \to [1, \infty)$, we have

(A1) [**Minorisation condition**] for all $d > 0$, there exists $\varepsilon_d > 0$ and a probability measure ν_d such that

$$\forall x \in C_d := \{V \leqslant d\}, \quad P(x,\cdot) \geqslant \varepsilon_d \nu_d(\cdot) \tag{2.2}$$

(A2) **[Drift condition]** there exists a constants $(\lambda, b) \in (0,1) \times \mathbb{R}^+$ such that for all $x \in \mathsf{X}$,

$$PV(x) \leqslant \lambda V(x) + b$$

Typically, the function V is unbounded (but in particular situations, it can also be bounded) and the sublevel set $\{V \leqslant d\}$ is typically compact (when the chain takes value a topological space)... Roughly speaking, (A1) tells you that wherever x moves in a set C_d, the measure $P(x,\cdot)$ is lower bounded by the non-trivial measure $\varepsilon_d \nu_d(\cdot)$. In many cases, $\mathsf{X} = \mathbb{R}^n$, and P is dominated by the Lebesgue measure: $P(x, \mathrm{d}y) = p(x,y)\mathrm{d}y$. In that case, we usually take $P(x,A) \geqslant \varepsilon_d \nu_d(A)$ where

$$\varepsilon_d = \int_{\mathsf{X}} \left[\inf_{x \in C_d} p(x,y) \right] \mathrm{d}y, \quad \nu_d(A) = \frac{\int_A \inf_{x \in C_d} p(x,y) \mathrm{d}y}{\varepsilon_d}$$

i.e. we only need to bound from below the kernel density $p(x,y)$ when $x \in C_d$. If C_d is compact, then it is quite easy to check such lower-bound. In the Markov chain terminology, if (2.2) holds, we say that C_d is a 1-small set.

The drift condition (A2) tells you that in the mean sense, the drift function V is shrinked by a factor λ up to the additive constant b... Intuitively speaking, the Markov kernel P does not bring to regions where V is too large so that the chain does not go to infinity too quickly (since limited values for V typically correspond to bounded sets). And we can easily imagine that such chains will have nice ergodic properties.

Before stating the result, we must say that, in practise, for a given Markov kernel P, there is no general rule for guessing the expression of a drift function V that satisfies (A2), and we have to try different functions V for checking the assumptions... For example, if $X_{k+1} = \alpha X_k + \varepsilon_k$ where (ε_k) are iid and $\alpha \in (0,1)$. If we know that $\mathbb{E}[|\varepsilon_1|^r] < \infty$, then we can try a drift function $V(x) = |x|^r$ and if $\mathbb{E}[\mathrm{e}^{\beta \varepsilon_1}] < \infty$, then we can try $V(x) = \mathrm{e}^{\beta x}$. For MH algorithms, we also sometimes use a negative power of the target density. But once again, the choice of V is very model specific (and in some sense, this is a good opportunity to be imaginative!!!). We now show that assumptions (A1) and (A2) imply that the Markov kernel P is "geometrically ergodic" in the following sense.

THEOREM 2.7 (▶ *Geometric ergodicity).* Assume (A1) and (A2) for some measurable function $V \geqslant 1$. Then, there exists a constant $\rho \in (0,1)$ such that for all $x, x' \in \mathsf{X}$ and all $n \in \mathbb{N}$,

$$\left\| P^n(x,\cdot) - P^n(x',\cdot) \right\|_{TV} \leqslant \rho^n \left[V(x) + V(x') \right].$$

REMARK 2.8 . Assume that there exist a constant $\varepsilon > 0$ and a probability measure ν such that for all $x \in \mathsf{X}$, $P(x,\cdot) \geqslant \varepsilon \nu(\cdot)$. In that case, (A1) and (A2) are satisfied with the constant function $V(x) = 1$ and Theorem 2.7 then shows that

$$\left\| P^n(x,\cdot) - P^n(x',\cdot) \right\|_{TV} \leqslant 2\rho^n.$$

for some constant $\rho \in (0,1)$. Such a Markov chain is usually said to be **uniformly ergodic**.

The proof needs several steps. To bound $\|P^n(x,\cdot) - P^n(x',\cdot)\|_{TV}$, we will construct a bivariate Markov chain (X_k, X'_k) such that first component process (X_k) behaves marginally as a Markov chain starting from x with Markov kernel P, while the second component process (X_k) behaves marginally as a Markov chain starting from x' with Markov kernel P. Let us be more specific... In what follows, we choose d sufficiently large so that

$$\bar{\lambda} := \lambda + \frac{2b}{1+d} < 1 \tag{2.3}$$

Definition of the joint kernel $\bar{P}$

Define $Q(x_k, \mathrm{d}x_{k+1}) = \frac{P(x_k,\mathrm{d}x_{k+1}) - \varepsilon_d \nu_d(\mathrm{d}x_{k+1})}{1-\varepsilon_d}$ and set

$$\begin{aligned}\bar{P}((x_k,x'_k),\mathrm{d}x_{k+1}\mathrm{d}x'_{k+1}) &= \mathbf{1}_{x_k=x'_k} P(x_k,\mathrm{d}x_{k+1})\delta_{x_{k+1}}(x'_{k+1})\\ &+\mathbf{1}_{x_k\neq x'_k}\mathbf{1}_{(x_k,x'_k)\notin C_d^2}\left[P(x_k,\mathrm{d}x_{k+1})P(x'_k,\mathrm{d}x'_{k+1})\right]\\ &+\mathbf{1}_{x_k\neq x'_k}\mathbf{1}_{(x_k,x'_k)\in C_d^2}\left[\varepsilon_d\nu_d(\mathrm{d}x_{k+1})\delta_{x_{k+1}}(x'_{k+1})+(1-\varepsilon_d)Q(x_k,\mathrm{d}x_{k+1})Q(x'_k,\mathrm{d}x'_{k+1})\right]\end{aligned}$$

Actually, $\bar{P}$ is a Markov kernel on $\mathsf{X}^2 \times \mathscr{X}^{\otimes 2}$ and it can be easily checked that

$$\bar{P}((x,x'),\cdot) \in \mathscr{C}(P(x,\cdot),P(x',\cdot)) \tag{2.4}$$

where for $\mu,\nu \in \mathscr{M}_1(\mathsf{X})$, we write $\mathscr{C}(\mu,\nu)$ for the set of probability measures γ on $\mathsf{X}\times\mathsf{X}$ satisfying the marginal conditions: for any $A \in \mathscr{X}$,

$$\gamma(A\times\mathsf{X}) = \mu(A)\,, \quad \text{and} \quad \gamma(\mathsf{X}\times A) = \nu(A)\,.$$

By induction on $n \in \mathbb{N}$, we can show that (2.4) implies

$$\bar{P}^n((x,x'),\cdot) \in \mathscr{C}(P^n(x,\cdot),P^n(x',\cdot)) \tag{2.5}$$

Interpretation of the joint kernel $\bar{P}$

Set $\bar{X}_k = (X_k, X'_k)$ and $\bar{C}_d = C_c \times C_d$. If $(\bar{X}_k)_{k\in\mathbb{N}}$ is a Markov chain with the Markov kernel $\bar{P}$, the transition from $\bar{X}_k = (x_k, x'_k)$ to $\bar{X}_{k+1} = (X_{k+1}, X'_{k+1})$ can be seen as follows

- If $x_k = x'_k$, draw $X_{k+1} \sim P(x_k,\cdot)$ and set $X'_{k+1} = X_{k+1}$.
- Otherwise,
  - If $(x_k,x'_k) \notin \bar{C}_d$, then
    - Draw independently $X_{k+1} \sim P(x_k,\cdot)$ and $X'_{k+1} \sim P(x'_k,\cdot)$
  - If $(x_k,x'_k) \in \bar{C}_d$, then
    - Draw $U \sim \mathrm{Ber}(\varepsilon_d)$.
    - If $U = 1$, draw $X_{k+1} \sim \nu_d$ and set $X'_{k+1} = X_{k+1}$.
    - If $U = 0$, draw independently $X_{k+1} \sim Q(x_k,\cdot)$ and $X'_{k+1} \sim Q(x'_k,\cdot)$.
- Set $\bar{X}_{k+1} = (X_{k+1}, X'_{k+1})$.

Therefore, the bivariate Markov chain $(\bar{X}_k)_{k\in\mathbb{N}} = (X_k, X'_k)_{k\in\mathbb{N}}$ is such that it tries to couple its two components with probability ε_d each time it falls into $\bar{C}_d$ and once it couples (ie $X_k = X'_k$) then, it stays together for ever (ie for all $n \geqslant k$, $X_n = X'_n$). Define $T = \inf\left\{k \,:\, X_k = X'_k\right\}$ and let $\bar{\mathbb{P}}_\xi$ denote the probability induced on $((\mathsf{X}\times\mathsf{X})^{\mathbb{N}}, (\mathscr{X}\otimes\mathscr{X})^{\otimes\mathbb{N}})$ by the Markov kernel $\bar{P}$ and initial distribution $\xi \in \mathscr{M}_1(\mathsf{X}\times\mathsf{X})$. The associated expectation operator is denoted by $\bar{\mathbb{E}}_\xi$ and by convention we simply write $\bar{\mathbb{E}}_{x,x'}$ or $\bar{\mathbb{P}}_{x,x'}$ when $\xi = \delta_{(x,x')}$

Exact coupling inequality

Similarly to Lemma 1.20, the coupling inequality, in this context, writes:

$$\left\|P^n(x,\cdot) - P^n(x',\cdot)\right\|_{TV} \leqslant 2\bar{\mathbb{P}}_{x,x'}(X_n \neq X'_n) = 2\mathbb{P}_{x,x'}(T > n) \tag{2.6}$$

Indeed for any measurable function f such that $|f| \leqslant 1$, we have

$$\begin{aligned}|P^n f(x) - P^n f(x')| = |\mathbb{E}_{x,x'}[f(X_n) - f(X'_n)]| &= |\mathbb{E}_{x,x'}[(f(X_n) - f(X'_n))\mathbf{1}_{X_n\neq X'_n}]|\\ &\leqslant 2\bar{\mathbb{P}}_{x,x'}(X_n \neq X'_n) = 2\mathbb{P}_{x,x'}(T > n)\end{aligned}$$

where the last equality follows from the fact that if $X_k = X'_k$, then $\mathbb{P}_{x,x'}$-a.s., $X_n = X'_n$ for all $n \geqslant k$.

Some nice properties of $\bar{P}$

The following inequalities are immediate:

1. Set $\Delta(x,x') = \mathbf{1}_{x \neq x'}$, then
   (a) if $(x,x') \in \bar{C}_d$, $\bar{P}\Delta(x,x') \leqslant (1-\varepsilon_d)\Delta(x,x')$
   (b) if $(x,x') \notin \bar{C}_d$, $\bar{P}\Delta(x,x') \leqslant \Delta(x,x')$
2. Setting $\bar{V}(x,x') = (V(x)+V(x'))/2$, we have $\bar{P}\bar{V}(x,x') = 2^{-1}(PV(x)+PV(x')) \leqslant \lambda\bar{V}(x,x')+b$. This implies
   (a) if $(x,x') \in \bar{C}_d$, $\bar{P}\bar{V}(x,x') \leqslant (\lambda+b)\bar{V}(x,x')$
   (b) if $(x,x') \notin \bar{C}_d$, $\bar{P}\bar{V}(x,x') \leqslant \underbrace{\left(\lambda+\frac{2b}{1+d}\right)}_{\bar{\lambda}}\bar{V}(x,x')$

We now have all the tools for proving Theorem 2.7.

PROOF. [of Theorem 2.7]
For any $\beta \in (0,1)$, define

$$\rho_\beta = \max((1-\varepsilon_d)^{1-\beta}(\lambda+b)^\beta, \bar{\lambda}^\beta) \tag{2.7}$$

The expression of ρ_β may seem a bit complicated (we will understand why we choose ρ_β like this in (2.8) below) but, since $\bar{\lambda}$ and $1-\varepsilon_d$ are both in $(0,1)$, we can always pick β sufficiently small (but positive) so that $\boxed{\rho_\beta \in (0,1)}$.
This ρ_β being chosen, set $W = \Delta^{1-\beta}\bar{V}^\beta$. Then, using Holder's inequality and the inequalities in the section **Some nice properties of $\bar{P}$**, we have for all $(x,x') \in \mathsf{X}^2$,

$$\begin{aligned}
\bar{P}W(x,x') = \bar{P}(\Delta^{1-\beta}\bar{V}^\beta)(x,x') &\leqslant (\bar{P}\Delta(x,x'))^{1-\beta}(\bar{P}\bar{V}(x,x'))^\beta \\
&\leqslant (\Delta^{1-\beta}\bar{V}^\beta)(x,x') \times \begin{cases} (1-\varepsilon_d)^{1-\beta}(\lambda+b)^\beta & \text{if } (x,x') \in C_d^2 \\ \bar{\lambda}^\beta & \text{if } (x,x') \notin C_d^2 \end{cases} \\
&\leqslant \rho_\beta W(x,x')
\end{aligned}$$

This implies by induction that for all $n \in \mathbb{N}$ and all $(x,x') \in \mathsf{X}^2$,

$$\bar{P}^n W(x,x') \leqslant \rho_\beta^n W(x,x') \tag{2.8}$$

Then

$$\left\|P^n(x,\cdot)-P^n(x',\cdot)\right\|_{TV} \overset{(1)}{\leqslant} 2\bar{P}^n\Delta(x,x') \overset{(2)}{\leqslant} 2\bar{P}^n W(x,x') \overset{(3)}{\leqslant} 2\rho_\beta^n W(x,x') \overset{(4)}{\leqslant} \rho_\beta^n(V(x)+V(x'))$$

where (1) comes from (2.5) and (2.6), (2) from $\Delta(x,x') = \Delta^{1-\beta}(x,x') \leqslant W(x,x')$ because $V \geqslant 1$, (3) from (2.8) and (4) from

$$W(x,x') \leqslant \left(\frac{V(x)+V(x')}{2}\right)^\beta \leqslant \frac{V(x)+V(x')}{2}$$

since $V \geqslant 1$ and $\beta \in (0,1)$. ■

COROLLARY 2.9 . Assume that (A1) and (A2) hold for some measurable function $V \geqslant 1$. Then, the Markov kernel P admits a unique invariant probability measure π. Moreover, $\pi(V) < \infty$ and there exists constants $(\rho,\beta) \in (0,1) \times \mathbb{R}^+$ such that for all $\mu \in \mathrm{M}_1(\mathsf{X})$ and all $n \in \mathbb{N}$,

$$\|\mu P^n - \pi\|_{TV} \leqslant \beta\rho^n\mu(V).$$

PROOF. For any $\mu, \nu \in \mathsf{M}_1(\mathsf{X})$ and any $h \in \mathsf{F}(\mathsf{X})$ such that $|h| \leqslant 1$, we have, using Theorem 2.7,

$$|\mu P^n h - \nu P^n h| = |\int_{\mathsf{X}^2} \mu(\mathrm{d}x)\nu(\mathrm{d}y)[P^n h(x) - P^n h(y)]| \leqslant \int_{\mathsf{X}^2} \mu(\mathrm{d}x)\nu(\mathrm{d}y)|[P^n h(x) - P^n h(y)]| \leqslant \rho^n[\mu(V) + \nu(V)]$$

Thus,

$$\|\mu P^n - \nu P^n\|_{TV} \leqslant \rho^n[\mu(V) + \nu(V)] \tag{2.9}$$

Replacing μ by δ_x and ν by $P(x,\cdot)$, we get for all $x \in \mathsf{X}$,

$$\left\| P^n(x,\cdot) - P^{n+1}(x,\cdot) \right\|_{TV} \leqslant \rho^n[V(x) + PV(x)] \leqslant \rho^n[(1+\lambda)V(x) + b]$$

This implies that $\{P^n(x,\cdot)\}$ is a Cauchy sequence and since $(\mathsf{M}_1(\mathsf{X}), \|\cdot\|_{TV})$ is complete, it converges to a limit $\pi \in \mathsf{M}_1(\mathsf{X})$. Then, for all $x \in \mathsf{X}$ and all $h \in \mathsf{F}(\mathsf{X})$ such that $|h| \leqslant 1$, we also have $|Ph| \leqslant 1$ and therefore

$$\pi(Ph) = \lim_{n\to\infty} P^n(Ph)(x) = \lim_{n\to\infty} P^{n+1}h(x) = \pi(h)$$

showing that π is P-**invariant**. We now show **uniqueness** of an invariant probability measure. To see this, note that π actually does not depend on the choice of x. Indeed, replacing μ by δ_x and ν by $\delta_{x'}$ in (2.9), we get that $\lim_{n\to\infty} \|P^n(x,\cdot) - P^n(x',\cdot)\|_{TV} = 0$. Therefore, for all $x \in \mathsf{X}$, $\lim_{n\to\infty} P^n h(x) = \pi(h)$. Let π' be an invariant probability measure for P, then

$$\pi'(h) = \pi' P^n(h) = \int \pi'(\mathrm{d}x) \underbrace{P^n h(x)}_{\to \pi(h)} \to_{n\to\infty} \pi(h)$$

where the last equality comes from Lebesgue's dominated convergence theorem. Since $PV \leqslant \lambda V + b$, we have by induction for all $n \in \mathbb{N}$,

$$P^n V(x) \leqslant \lambda^n V(x) + b\left(\sum_{k=0}^{n-1} \lambda^k\right) \leqslant \lambda^n V(x) + \frac{b}{1-\lambda}$$

Therefore, for any $M > 0$, by Jensen's inequality applied to the convex function $u \mapsto u \wedge M$, we have $P^n(V \wedge M)(x) \leqslant (P^n V(x)) \wedge M \leqslant \left(\lambda^n V(x) + \frac{b}{1-\lambda}\right) \wedge M$. We then integrate wrt π and use $\pi = \pi P^n$:

$$\pi(V \wedge M) = \pi P^n(V \wedge M) \leqslant \int \pi(\mathrm{d}x) \left(\lambda^n V(x) + \frac{b}{1-\lambda}\right) \wedge M$$

The Lebesgue dominated convergence theorem then shows by letting n to infinity, $\pi(V \wedge M) \leqslant \frac{b}{1-\lambda} \wedge M$. Then, letting M to infinity, we get $\pi(V) \leqslant b/(1-\lambda) < \infty$. To complete the proof, apply (2.9) with $\nu = \pi$, we get

$$\|\mu P^n - \underbrace{\pi P^n}_{\pi}\|_{\mathrm{TV}} \leqslant \rho^n[\mu(V) + \pi(V)] \leqslant \beta\rho^n[\mu(V)$$

with $\beta = 1 + \pi(V) < \infty$. ■

Index