A short course in Markov chains (second part)

Randal Douc

Chapter 1

Geometric ergodicity

Time schedule (Note 1): 3H-Session

Let (X_n) be a Markov chain with Markov kernel P, and assume that P admits an invariant probability measure π. In this chapter, we are interested in identifying conditions under which one can obtain explicit bounds on the error between the marginal distribution of X_n and the target distribution π. To this end, we first introduce an appropriate notion of distance between probability measures, focusing in particular on the total variation distance and the V-norm. We then study the discrepancy between π and $P^n(x,\cdot)$, the distribution of the n-th iterate of the Markov kernel starting from an arbitrary point $x \in \mathsf{X}$, by deriving bounds on the total variation distance or V-norm between these two measures.

1.1 Total variation norm and coupling

We start with the notion of coupling between probability measures. In words, if μ, ν are two probability measures on $(\mathsf{X}, \mathscr{X})$, then a coupling γ of (μ, ν) is a probability measure on the product space $(\mathsf{X}^2, \mathscr{X}^{\otimes 2})$ such that if $(X,Y) \sim \gamma$, then we have the marginal conditions: $X \sim \mu$ and $Y \sim \nu$.

DEFINITION 1.1 . Let $(\mathsf{X}, \mathscr{X})$ be a measurable space and let ν, μ be two probability measures $\mu, \nu \in \mathrm{M}_1(\mathsf{X})$. We define $\mathscr{C}(\mu, \nu)$, the coupling set associated to (μ, ν) as follows

$$\mathscr{C}(\mu,\nu) = \left\{\gamma \in \mathrm{M}_1(\mathsf{X}^2) \,:\, \forall A \in \mathscr{X}, \gamma(A \times \mathsf{X}) = \mu(A), \gamma(\mathsf{X} \times A) = \nu(A)\right\}$$

Any $\gamma \in \mathscr{C}(\mu,\nu)$ is called a coupling of (μ,ν).

To illustrate the definition of a coupling, consider first a simple example. One can construct a coupling of (μ,μ) by sampling $X \sim \mu$ and then setting $Y = X$. The joint distribution of (X,Y) is then a coupling of (μ,μ). This construction yields a maximally dependent coupling, since the two components are equal by design.

Alternatively, one may construct an independent coupling by drawing X and Y independently according to the same distribution μ. In this case, the joint distribution of (X,Y) is again a coupling of (μ,μ), but now with complete independence between the two coordinates.

These examples naturally raise the question of which coupling is the most relevant or useful. As we shall see in what follows, there is a substantial degree of freedom in the choice of a coupling, and the optimal choice depends on the problem under consideration. We will present several couplings that are particularly useful in practice, but there is no universal rule that applies in all situations.

Before going further into coupling techniques, we introduce the total variation norm and explain its fundamental connection with couplings.

DEFINITION 1.2 . Let $(\mathsf{X},\mathscr{X})$ be a measurable space and let ν,μ be two probability measures $\mu,\nu \in \mathrm{M}_1(\mathsf{X})$. Then the total variation norm between μ and ν noted $\|\mu-\nu\|_{\mathrm{TV}}$, is defined by

$$\|\mu-\nu\|_{\mathrm{TV}} = 2\sup\{|\mu(f)-\nu(f)| \;:\; f\in(\mathsf{X}), 0\leqslant f\leqslant 1\} \tag{1.1}$$

$$= \int |\varphi_0-\varphi_1|(x)\zeta(\mathrm{d}x) \tag{1.2}$$

$$= 2\inf\{\mathbb{P}(X\neq Y) \;:\; (X,Y)\sim\gamma \text{ where } \gamma\in\mathscr{C}(\mu,\nu)\} \tag{1.3}$$

where $\mu(\mathrm{d}x)=\varphi_0(x)\zeta(\mathrm{d}x)$ and $\nu(\mathrm{d}x)=\varphi_1(x)\zeta(\mathrm{d}x)$.

Before proving that these different expressions of the total variation norm are indeed equivalent, let us make a few remarks.

(a) The reader might wonder why we can always write $\mu(\mathrm{d}x)=\varphi_0(x)\zeta(\mathrm{d}x)$ and $\nu(\mathrm{d}x)=\varphi_1(x)\zeta(\mathrm{d}x)$ for some suitably chosen measure ζ, and for measurable functions φ_0 and φ_1. Indeed, if $\mu,\nu\in\mathrm{M}_1(\mathsf{X})$, then by setting $\zeta=\mu+\nu$, we obtain the two implications

$$(\zeta(A)=0)\Longrightarrow(\mu(A)=0) \quad \text{and} \quad (\zeta(A)=0)\Longrightarrow(\nu(A)=0)\,.$$

Therefore, the measure ζ dominates both μ and ν. By the Radon–Nikodym theorem, the measures μ and ν admit densities with respect to ζ, which we denote by φ_0 and φ_1 in Definition 1.2.

(b) The first definition, (1.1), is expressed as a supremum over a class of test functions, while the last one is formulated as an infimum over coupling measures. These two characterizations can therefore be interpreted as a duality formula.

(c) The intermediate expression shows that the total variation norm can be written as an L_1-distance between the densities of the two distributions with respect to a common dominating measure.

(d) A direct consequence of the equivalence between these definitions of the total variation norm is the following. If f is a measurable function taking values in $[0,1]$, and if X,Y are random variables such that $(X,Y)\sim\gamma$ with $\gamma\in\mathscr{C}(\mu,\nu)$, then the coupling inequality

$$|\mu(f)-\nu(f)|\leqslant\mathbb{P}(X\neq Y)$$

holds. This inequality will be used repeatedly in what follows, and it highlights why coupling techniques are so powerful in practice.

PROOF. [of the equivalences in Definition 1.2] Call A, B and C the quantities that appear respectively in (1.1), (1.2) and (1.3). Any function f satisfies $0\leqslant f\leqslant 1$ if and only if the function $g=2f-1$ satisfies $|g|\leqslant 1$. This implies immediately

$$A = 2\sup\{|\mu(f)-\nu(f)| \;:\; f\in(\mathsf{X}), 0\leqslant f\leqslant 1\} = \sup\{|\mu(g)-\nu(g)| \;:\; g\in(\mathsf{X}), |g|\leqslant 1\}$$

Moreover, for any $g\in(\mathsf{X})$ such that $|g|\leqslant 1$,

$$|\mu(g)-\nu(g)| = \left|\int(\varphi_1-\varphi_0)(x)g(x)\zeta(\mathrm{d}x)\right| \leqslant \int|\varphi_1-\varphi_0|(x)\underbrace{|g(x)|}_{\leqslant 1}\zeta(\mathrm{d}x) = B$$

Therefore, $A \leqslant B$. Moreover, setting $g^*(x) = \text{sign}(\varphi_0(x) - \varphi_1(x))$, we have $|g^*| = 1$ and therefore,

$$B = \int |\varphi_0 - \varphi_1|(x)\zeta(\mathrm{d}x) = \int (\varphi_0(x) - \varphi_1(x))\, g^*(x)\zeta(\mathrm{d}x)$$
$$= \mu(g^*) - \nu(g^*) \leqslant \sup\{|\mu(g) - \nu(g)| \ :\ g \in (\mathsf{X}), |g| \leqslant 1\} = A$$

Thus $\boxed{A = B}$. Now let $f \in (\mathsf{X})$ be such that $0 \leqslant f \leqslant 1$ and let X, Y be random variables such that $(X,Y) \sim \gamma$ with $\gamma \in \mathscr{C}(\mu,\nu)$, then

$$|\mu(f) - \nu(f)| = |\mathbb{E}[f(X) - f(Y)]| = |\mathbb{E}\left[\{f(X) - f(Y)\}\mathbf{1}_{X \neq Y}\right]| \leqslant \mathbb{E}[\underbrace{|f(X) - f(Y)|}_{\leqslant 1}\mathbf{1}_{X \neq Y}] \leqslant \mathbb{P}(X \neq Y)$$

This shows that $\boxed{A \leqslant C}$. To finish the proof, we will show that $C \leqslant B$ and to do so, we will exhibit an **optimal coupling** of (μ,ν).

Define
$$\varepsilon = \int_{\mathsf{X}} \varphi_0 \wedge \varphi_1(x)\zeta(\mathrm{d}x) \quad \text{and} \quad \zeta'(\mathrm{d}x) = \frac{\varphi_0 \wedge \varphi_1(x)}{\varepsilon}\zeta(\mathrm{d}x)$$
Then, it can be readily checked that $\zeta' \in \mathrm{M}_1(\mathsf{X})$, and $\mu(\mathrm{d}x) \geqslant \varepsilon\zeta'(\mathrm{d}x)$ and $\nu(\mathrm{d}x) \geqslant \varepsilon\zeta'(\mathrm{d}x)$. This implies that there exist $\mu_1, \nu_1 \in \mathrm{M}_1(\mathsf{X})$ such that

$$\mu(\mathrm{d}x) = \varepsilon\zeta'(\mathrm{d}x) + (1-\varepsilon)\mu_1(\mathrm{d}x)$$
$$\nu(\mathrm{d}x) = \varepsilon\zeta'(\mathrm{d}x) + (1-\varepsilon)\nu_1(\mathrm{d}x)$$

Define $\gamma(\mathrm{d}x\mathrm{d}y) = \varepsilon\zeta'(\mathrm{d}x)\delta_x(\mathrm{d}y) + (1-\varepsilon)\mu_1(\mathrm{d}x)\nu_1(\mathrm{d}y)$. Obviously, $\gamma \in \mathscr{C}(\mu,\nu)$. Let X, Y be two random variables such that $(X,Y) \sim \gamma$. We can draw (X,Y) in the following way: draw a Bernoulli variable $U \sim \text{Ber}(\varepsilon)$. If $U = 1$, draw $X \sim \zeta'$ and set $Y = X$. If $U = 0$, then draw independently $X \sim \mu_1$ and $Y \sim \nu_1$. Then, clearly

$$\mathbb{P}(X \neq Y) = 1 - \mathbb{P}(X = Y) \leqslant 1 - \varepsilon = 1 - \int_{\mathsf{X}} \varphi_0 \wedge \varphi_1(x)\zeta(\mathrm{d}x)$$
$$= \frac{1}{2}\int_{\mathsf{X}} \underbrace{\varphi_0 + \varphi_1(x) - 2\varphi_0 \wedge \varphi_1(x)}_{|\varphi_0 - \varphi_1(x)|}\zeta(\mathrm{d}x)$$

This shows that $\boxed{C \leqslant B}$ and the proof is completed. ■

EXERCISE 1 . Show that the supremum in (1.1) is attained with a convenient choice of f.

Another equivalent expression of the total variation distance is

$$\|\mu - \nu\|_{\mathrm{TV}} = 2\inf\{\gamma(\Delta) \ :\ \gamma \in \mathscr{C}(\mu,\nu)\} \tag{1.4}$$

where we have used the notation $\Delta(x,y) = \mathbf{1}_{x \neq y}$ (we also say that $\Delta(x,y)$ is the Hamming distance between x and y).

EXAMPLE 1.3 . See Figure 1.1. Let $\mu = \mathscr{N}(-1,1)$ and $\nu = \mathscr{N}(1,1)$. Let $X \sim \mathscr{N}(-1,1)$ and set $Y = X + 2$. Then, (X,Y) is a coupling of (μ,ν) but it is not the optimal coupling for the Hamming distance since $\mathbb{P}(X \neq Y) = 1$, whereas using Definition 1.2,

$$\|\mu - \nu\|_{\mathrm{TV}} = 2\left(1 - \int_{-\infty}^{\infty} \phi(x+1) \wedge \phi(x-1)\mathrm{d}x\right) = 2\left(1 - 2\int_1^{\infty} \frac{\mathrm{e}^{-u^2/2}}{\sqrt{2\pi}}\mathrm{d}u\right),$$

where ϕ the density of the standard Gaussian distribution.

Before moving to geometric ergodicity, we define a generalisation of the total variation norm, namely the V-norm.

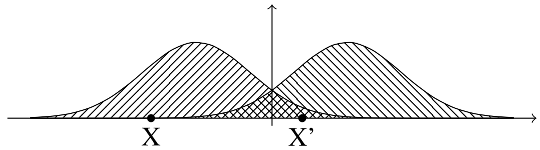

Figure 1.1: An example of coupling of two probability measures.

Definition 1.4 . Let $(\mathsf{X},\mathscr{X})$ be a measurable space and let ν,μ be two probability measures $\mu,\nu \in \mathrm{M}_1(\mathsf{X})$. Let $V:\mathsf{X}\to[1,\infty)$ be a measurable function. Then V-norm between μ and ν noted $\|\mu-\nu\|_{\mathrm{V}}$, is defined by

$$\begin{aligned}\|\mu-\nu\|_{\mathrm{V}} &= 2\sup\{|\mu(f)-\nu(f)| \;:\; f\in(\mathsf{X}), 0\leqslant f\leqslant V\}\\ &= \int|\varphi_0-\varphi_1|(x)\;V(x)\;\zeta(\mathrm{d}x)\\ &= \inf\left\{\mathbb{E}\left[\{V(x)+V(y)\}1_{\{X\neq Y\}}\right] \;:\; (X,Y)\sim\gamma \text{ where } \gamma\in\mathscr{C}(\mu,\nu)\right\}\end{aligned}$$

where $\mu(\mathrm{d}x)=\varphi_0(x)\zeta(\mathrm{d}x)$ and $\nu(\mathrm{d}x)=\varphi_1(x)\zeta(\mathrm{d}x)$.

Hence, the total variation norm is a particular case of the V-norm, corresponding to the choice $V\equiv 1$. We leave the proof of the equivalence between the various definitions of the V-norm to the reader, as it follows closely the arguments used to establish the analogous equivalences for the total variation norm.

Seeing Definition 1.4, an equivalent expression of the V-norm is

$$\|\mu-\nu\|_{\mathrm{V}} = 2\inf\{\gamma(\bar{V}\Delta) \;:\; \gamma\in\mathscr{C}(\mu,\nu)\} \tag{1.5}$$

where we have set $\bar{V}(x,x')=[V(x)+V(x')]/2$. It is also useful to note that the total variation distance is dominated by the V-norm, that is, $\|\mu-\nu\|_{\mathrm{TV}}\leqslant\|\mu-\nu\|_{\mathrm{V}}$.

1.2 Geometric ergodicity

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

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

$$\forall x\in C_d:=\{V\leqslant d\},\quad P(x,\cdot)\geqslant\varepsilon_d\nu_d(\cdot) \tag{1.6}$$

(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, although in specific situations it may also be bounded. Moreover, the level set $\{V\leqslant d\}$ is usually compact when the chain takes values in a topological space. Roughly speaking, Assumption (A1) states that whenever x belongs to the set C_d, the transition measure $P(x,\cdot)$ is uniformly lower bounded by the non-trivial measure $\varepsilon_d \nu_d(\cdot)$.

In many applications, the state space is $\mathsf{X} = \mathbb{R}^n$, and the Markov kernel P is dominated by the Lebesgue measure, that is, $P(x,\mathrm{d}y) = p(x,y)\mathrm{d}y$. In this setting, we typically enforce the minorization condition in the form $P(x,A) \geqslant \varepsilon_d \nu_d(A)$, where

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

In other words, it is sufficient to derive a uniform lower bound on the transition density $p(x,y)$ for $x \in C_d$. When C_d is compact, such a lower bound is often relatively easy to establish. In the terminology of Markov chains, whenever (1.6) holds, the set C_d is called a small set.

The drift condition in Assumption (A2) expresses that, on average, the drift function V is contracted by a factor λ, up to an additive constant b. Intuitively, this means that the Markov kernel P prevents the chain from spending too much time in regions where V takes very large values. Since moderate values of V typically correspond to bounded regions of the state space, this condition ensures that the chain does not escape to infinity too rapidly. It is therefore natural to expect such chains to exhibit good ergodic properties.

Before stating the main result, it is worth emphasizing that, in practice, there is no general recipe for choosing a drift function V satisfying Assumption (A2) for a given Markov kernel P. One usually has to experiment with several candidate functions. For instance, if $X_{k+1} = \alpha X_k + \varepsilon_k$, where (ε_k) are i.i.d. random variables and $\alpha \in (0,1)$, then assuming $\mathbb{E}[|\varepsilon_1|^r] < \infty$, a natural choice is $V(x) = |x|^r$. If instead $\mathbb{E}[\mathrm{e}^{\beta\varepsilon_1}] < \infty$, one may try $V(x) = \mathrm{e}^{\beta x}$. For Metropolis–Hastings algorithms, it is also common to consider negative powers of the target density. In all cases, however, the choice of V is highly model dependent, and, in some sense, this step provides an opportunity for a certain amount of creativity.

We now show that Assumptions (A1) and (A2) imply that the Markov kernel P is geometrically ergodic in the following sense.

THEOREM 1.5 . [Geometric ergodicity] Assume (A1) and (A2) for some measurable function $V \geqslant 1$. Then, there exists a constant $(\rho,\alpha) \in (0,1) \times \mathbb{R}^+$ such that for all $x,x' \in \mathsf{X}$ and all $n \in \mathbb{N}$,

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

REMARK 1.6 . 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 1.5 then shows that

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

for some constants $(\rho,\alpha) \in (0,1) \times \mathbb{R}^+$. Such a Markov chain is usually said to be uniformly ergodic.

The proof proceeds in several steps. In order to bound $\|P^n(x,\cdot) - P^n(x',\cdot)\|_{\mathrm{V}}$, we construct a bivariate Markov chain (X_k, X'_k) with the property that the first component (X_k) evolves marginally as a Markov chain with kernel P started from x, while the second component (X'_k) evolves marginally as a Markov chain with kernel P started from x'. This coupling construction allows us to study the two marginal distributions simultaneously and to control their discrepancy through their joint evolution. Let us now describe this construction more precisely.

In what follows, we choose d sufficiently large so that

$$\boxed{\bar{\lambda} := \lambda + \frac{2b}{1+d} < 1} \tag{1.7}$$

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

$$\bar{P}((x,x'),\cdot) \in \mathscr{C}(P(x,\cdot),P(x',\cdot)) \tag{1.8}$$

This will indeed imply by induction that for any $n \in \mathbb{N}$,

$$\bar{P}^n((x,x'),\cdot) \in \mathscr{C}(P^n(x,\cdot),P^n(x',\cdot)) \tag{1.9}$$

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 constructed so as to attempt to couple its two components with probability ε_d each time it enters the set $\bar{C}_d$. Once coupling occurs (that is, when $X_k = X'_k$), the two components remain together forever, meaning that $X_n = X'_n$ for all $n \geqslant k$.

Some useful 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 1.5.

PROOF. [of Theorem 1.5] For any $\eta > 0$, define

$$\rho_\eta = \max\left(\frac{(\lambda+b)d+\eta(1-\varepsilon)}{d+\eta}, \frac{\bar{\lambda}+\eta}{1+\eta}\right) \tag{1.10}$$

Since $\lim_{\eta\to\infty}\rho_\eta = 1-\varepsilon$, we may choose $\eta^\star$ such that $\rho_{\eta^\star} < 1$. This ensures that the contraction factor $\rho_{\eta^\star}$ is strictly less than 1, which is crucial for proving geometric convergence.

Set

$$W = (\bar{V}+\eta^\star)\Delta.$$

We first show that, for any $x, x' \in \mathsf{X}$,

$$\bar{P}W(x,x') \leqslant \rho_{\eta^\star} W(x,x') \tag{1.11}$$

By definition of the Markov kernel $\bar{P}$, we have $\bar{P}W(x,x) = 0 = \rho_{\eta^\star}W(x,x)$, so it suffices to prove (1.11) for $\boxed{x \neq x'}$.

Now, for any $x,x' \in \mathsf{X}$ such that $x \neq x'$, using the fact that $\Delta \leqslant 1$, we get

$$\begin{aligned}
\bar{P}W(x,x') &\leqslant [\bar{P}\bar{V} + \eta\bar{P}\Delta](x,x') \\
&\leqslant \begin{cases} [(\lambda+b)\bar{V}+\eta(1-\varepsilon)](x,x') & \text{if } (x,x') \in \bar{C}_d \\ [\bar{\lambda}\bar{V}+\eta](x,x') & \text{if } (x,x') \notin \bar{C}_d \end{cases}
\end{aligned}$$

The right-hand side can be bounded by considering two cases:

Case 1: $(x,x') \in \bar{C}_d$

Observe that the function $v \mapsto \frac{(\lambda+b)v+\eta(1-\varepsilon)}{v+\eta}$ is monotone. It equals $1-\varepsilon < 1$ at $v = 0$ and converges to $\lambda + b > 1$ as $v \to \infty$, implying that the function is non-decreasing. Hence, for $(x,x') \in \bar{C}_d$ with $x \neq x'$,

$$\frac{(\lambda+b)\bar{V}(x,x')+\eta(1-\varepsilon)}{\bar{V}(x,x')+\eta} \leqslant \frac{(\lambda+b)d+\eta(1-\varepsilon)}{d+\eta} \leqslant \rho_{\eta^\star}$$

where the last inequality follows directly from (1.10).

Case 2: $(x,x') \notin \bar{C}_d$

Similarly, the function $v \mapsto \frac{\bar{\lambda}v+\eta}{v+\eta}$ is monotone. It equals 1 at $v = 0$ and converges to $\bar{\lambda} < 1$ as $v \to \infty$, so it is non-increasing. Since $\bar{V} \geqslant 1$, we have for $(x,x') \notin \bar{C}_d$ with $x \neq x'$

$$\frac{\bar{\lambda}\bar{V}(x,x')+\eta}{\bar{V}(x,x')+\eta} \leqslant \frac{\bar{\lambda}+\eta}{1+\eta} \leqslant \rho_{\eta^\star}$$

again using (1.10).

Combining the two cases, we have for any $x \neq x'$,

$$\bar{P}W(x,x') \leqslant \rho_{\eta^\star}\left[\bar{V}(x,x')+\eta\right] = \rho_{\eta^\star}W(x,x')$$

where we used $\Delta(x,x') = 1$ in the last equality. Therefore, (1.11) holds for all $\boxed{x,x' \in \mathsf{X}}$.

Now, for any $x,x' \in \mathsf{X}$,

$$\left\|P^n(x,\cdot)-P^n(x',\cdot)\right\|_{\mathrm{V}} \overset{(1)}{\leqslant} 2\bar{P}^n[\bar{V}\Delta](x,x') \leqslant 2\bar{P}^nW(x,x') \overset{(2)}{\leqslant} 2\rho_{\eta^\star}^nW(x,x') \overset{(3)}{\leqslant} \alpha\rho_{\eta^\star}^n(V(x)+V(x'))$$

Here, $\overset{(1)}{\leqslant}$ follows from (1.9) and (1.5), $\overset{(2)}{\leqslant}$ from (1.11), and $\overset{(3)}{\leqslant}$ holds by setting $\alpha = 1+\eta^\star$. This chain of inequalities establishes the desired geometric convergence in the V-norm.

■

COROLLARY 1.7 . 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 $(\rho,\alpha) \in (0,1)\times\mathbb{R}^+$ such that for all $\mu \in \mathrm{M}_1(\mathsf{X})$ and all $n \in \mathbb{N}$,

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

PROOF. For any $\mu, \nu \in \mathrm{M}_1(\mathsf{X})$ and any $h \in (\mathsf{X})$ such that $|h| \leqslant V$, we have, using Theorem 1.5,

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

Thus,

$$\|\mu P^n - \nu P^n\|_{\mathrm{V}} \leqslant \alpha\rho^n[\mu(V) + \nu(V)] \tag{1.12}$$

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

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

This implies that $\{P^n(x,\cdot)\}$ is a Cauchy sequence and since $(\mathrm{M}_1(\mathsf{X}), \|\cdot\|_{\mathrm{TV}})$ is complete, it converges to a limit $\pi \in \mathrm{M}_1(\mathsf{X})$. Then, for all $x \in \mathsf{X}$ and all $h \in (\mathsf{X})$ 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 (1.12), we get that $\lim_{n\to\infty} \|P^n(x,\cdot) - P^n(x',\cdot)\|_{\mathrm{TV}} = 0$. 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 (1.12) with $\nu = \pi$, we get

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

where we have set $\alpha' = \alpha[1 + \pi(V)] < \infty$. ■

Chapter 2

Central limit theorems

Time schedule (**Note** 2): 3H-Session

2.1 The Poisson Equation

2.1.1 Definition

We start with a general definition of a Poisson equation.

DEFINITION 2.1 . For a given measurable function h such that $\pi|h| < \infty$, the Poisson equation is defined by

$$\hat{h} - P\hat{h} = h - \pi(h) \tag{2.1}$$

A solution to Poisson equation (2.1) is a function $\hat{h}$ such that $P|\hat{h}|(x) < \infty$ for all $x \in \mathsf{X}$ and for all $x \in \mathsf{X}$, $\hat{h}(x) - P\hat{h}(x) = h(x) - \pi(h)$.

The following result holds under the set of assumptions (A1) and (A2).

THEOREM 2.2 . Assume (A1) and (A2) hold for some measurable function $V \geqslant 1$. Then, for any measurable function h such that $|h| \leqslant V$, the function

$$\hat{h} = \sum_{n=0}^{\infty} \{P^n h - \pi(h)\} \tag{2.2}$$

is well-defined. Moreover, $\hat{h}$ is a solution of the Poisson equation associated to h and there exists a constant γ such that for all $x \in \mathsf{X}$,

$$|\hat{h}(x)| \leqslant \gamma V(x)$$

PROOF. To see the existence of a solution to the Poisson equation under (A1) and (A2), note that by Theorem 1.5, $\sum_{n=0}^{\infty} \{P^n h(x) - \pi(h)\}$ converges for any $|h| \leqslant V$ and we can thus define

$$\hat{h}(x) = \sum_{n=0}^{\infty} \{P^n h(x) - \pi(h)\}$$

Then,

$$P\hat{h}(x) = \sum_{n=1}^{\infty} \{P^n h(x) - \pi(h)\}$$

which immediately shows (2.1). Moreover, setting $\hat{h}$ as in (2.2), Theorem 1.5 shows that for all $x \in \mathsf{X}$,

$$|\hat{h}(x)| \leqslant \frac{\alpha}{1-\rho} V(x)$$

■

2.1.2 Poisson equation and martingales

The interest of Poisson equation is that it allows to link quantities of interest of our Markov chain with a well-chosen martingale. Then, we apply limiting results on martingales and the impact of those results to our Markov chain.

We start with a refresher on martingales.

A refresh on martingales

Let $(M_n)_{n\in\mathbb{N}}$ be a sequence of random variables on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and let $(\mathscr{F}_n)_{n\in\mathbb{N}}$ be a filtration (ie for all $n \in \mathbb{N}$, $\mathscr{F}_n \subset \mathscr{F}_{n+1} \subset \mathscr{F}$). We say that $(M_n)_{n\in\mathbb{N}}$ is a $(\mathscr{F}_n)$-martingale if for all $n \in \mathbb{N}$, M_n is integrable and for all $n \geqslant 1$,

$$\mathbb{E}[M_n|\mathscr{F}_{n-1}] = M_{n-1}$$

The *increment process* of the martingale is by definition $(M_{n+1} - M_n)_{n\in\mathbb{N}}$.

The following CLT result holds for martingales with stationary increments. It is stated without proof.

Theorem 2.3 . If a sequence $(M_n)_{n\in\mathbb{N}}$ is a $(\mathscr{F}_n)$-martingale with stationary and square integrable increments, then

$$n^{-1/2} M_n \overset{\mathscr{L}_{\mathbb{P}}}{\rightsquigarrow} \mathscr{N}\left(0, \mathbb{E}[(M_1 - M_0)^2]\right)$$

Link with martingales

Define

$$S_n(h) = \sum_{k=0}^{n-1} \{h(X_k) - \pi(h)\}$$

The solution of the Poisson equation allows us to relate $S_n(h)$ to a martingale by writing:

$$S_n(h) = M_n(\hat{h}) + \hat{h}(X_0) - \hat{h}(X_n) \tag{2.3}$$

where

$$M_n(\hat{h}) = \sum_{k=1}^{n} \left\{\hat{h}(X_k) - P\hat{h}(X_{k-1})\right\} \tag{2.4}$$

Note that $\left\{M_n(\hat{h})\right\}_{n\in\mathbb{N}}$ is indeed a $(\mathscr{F}_k)$-martingale where $\mathscr{F}_k = \sigma(X_0, \ldots, X_k)$ since:

$$\mathbb{E}[M_n(\hat{h})|\mathscr{F}_{n-1}] - M_{n-1}(h) = \mathbb{E}[\hat{h}(X_n) - P\hat{h}(X_{n-1})|\mathscr{F}_{n-1}] = P\hat{h}(X_{n-1}) - P\hat{h}(X_{n-1}) = 0$$

This link with martingales allows to obtain LLN and Central Limit theorems for our Markov chain from limiting results on martingales. Since LLN has been already studied in different approaches in previous chapters, we only focus here on CLT.

2.1.3 Central Limit theorems

THEOREM 2.4 . Let P be a Markov kernel with a unique invariant probability measure π. Let $h \in \mathrm{L}_2(\pi)$. Assume that there exists a solution $\boxed{\hat{h} \in \mathrm{L}_2(\pi)}$ to the Poisson equation $\hat{h} - P\hat{h} = h$. Then

$$n^{-1/2} \sum_{k=0}^{n-1} \{h(X_k) - \pi(h)\} \overset{\mathscr{L}_{\mathbb{P}_\pi}}{\rightsquigarrow} \mathscr{N}(0, \sigma_\pi^2(h)) ,$$

where

$$\sigma_\pi^2(h) = \mathbb{E}_\pi[\{\hat{h}(X_1) - P\hat{h}(X_0)\}^2] \tag{2.5}$$

PROOF. Without loss of generality, we assume $\pi(h) = 0$. The sequence $(M_n(\hat{h}))_{n\in\mathbb{N}}$ defined in (2.4) is such that

$$\begin{aligned}\mathbb{E}_\pi[(M_n(\hat{h}) - M_{n-1}(\hat{h}))^2] &= \mathbb{E}_\pi[\{\hat{h}(X_1) - P\hat{h}(X_0)\}^2] \leqslant 2\mathbb{E}_\pi[\hat{h}^2(X_1) + (P\hat{h}(X_0))^2] \\ &= 2\left[\pi(\hat{h}^2) + \pi((P\hat{h})^2)\right] \overset{(1)}{\leqslant} 2\left[\pi(\hat{h}^2) + \underbrace{\pi P}_{\pi}(\hat{h})^2\right] = 2\pi(\hat{h}^2) < \infty\end{aligned}$$

where $\overset{(1)}{\leqslant}$ follows from Cauchy-Schwarz inequality. Therefore, the sequence $(M_n(\hat{h}))_{n\in\mathbb{N}}$ is a martingale with stationary and square integrable increments under $\mathbb{P}_\pi$. By Theorem 2.3, we have

$$n^{-1/2} M_n(\hat{h}) \overset{\mathscr{L}_{\mathbb{P}_\pi}}{\rightsquigarrow} \mathscr{N}\left(0, \mathbb{E}_\pi[\{\hat{h}(X_1) - P\hat{h}(X_0)\}^2]\right) . \tag{2.6}$$

Since the Markov chain $(X_k)_{k\in\mathbb{N}}$ is stationary under $\mathbb{P}_\pi$, we get $\mathbb{E}_\pi[|\hat{h}(X_0) + \hat{h}(X_n)|] \leqslant 2\pi(|\hat{h}|)$ which implies that

$$n^{-1/2}\{\hat{h}(X_0) + \hat{h}(X_n)\} \overset{\mathbb{P}_\pi-prob}{\longrightarrow} 0 .$$

Combining it with (2.6) and (2.3) and using Slutsky's lemma gives:

$$n^{-1/2} \sum_{k=0}^{n-1} h(X_k) \overset{\mathscr{L}_{\mathbb{P}_\pi}}{\rightsquigarrow} \mathscr{N}(0, \sigma_\pi^2(h))$$

■

THEOREM 2.5 . Assume that (A1) and (A2) hold for some function V. Then, for all measurable functions h such that $|h|^2 \leqslant V$,

$$n^{-1/2} \sum_{k=0}^{n-1} \{h(X_k) - \pi(h)\} \overset{\mathscr{L}_{\mathbb{P}_\pi}}{\rightsquigarrow} \mathscr{N}(0, \sigma_\pi^2(h)) ,$$

where

$$\sigma_\pi^2(h) = \mathbb{E}_\pi[\{\hat{h}(X_1) - P\hat{h}(X_0)\}^2] \tag{2.7}$$

and $\hat{h}$ is defined as in (2.2).

PROOF. Assume that (A1) and (A2) hold for some function V. Then, (A1) also holds with V replaced by $V^{1/2}$. Moreover, since $PV \leqslant \lambda V + b$, we have by Cauchy-Schwarz,

$$P(V^{1/2}) \leqslant (PV)^{1/2} \leqslant (\lambda V + b)^{1/2} \leqslant \lambda^{1/2} V^{1/2} + b^{1/2}$$

Finally, (A1) and (A2) hold for the function $V^{1/2}$. We can therefore apply Theorem 2.2 with V replaced by $V^{1/2}$. Then, for all $h \leqslant V^{1/2}$, the function $\hat{h}$ defined by (2.2) is solution to the Poisson equation and there exists a constant $\gamma > 0$ such that $\hat{h} \leqslant \gamma V^{1/2}$. This implies that $\pi(\hat{h}^2) \leqslant \gamma\pi(V) < \infty$ by Corollary 1.7. Therefore $\hat{h} \in \mathrm{L}_2(\pi)$ and Theorem 2.4 applies. The proof is completed. ■

Under the assumptions of Theorem 2.5, the CLT holds under $\mathbb{P}_\pi$. Moreover, this result can be extended to $\mathbb{P}_\nu$ for any probability measure ν in $\mathrm{M}_1(\mathsf{X})$.

This extension follows directly from the following useful proposition.

PROPOSITION 2.6 . Let P be a Markov kernel on $(\mathsf{X}, \mathscr{X})$ with invariant probability measure π. Assume that there exists a constant $\sigma^2 \in (0,\infty)$ and a distribution ξ such that

$$n^{-1/2}\sum_{k=0}^{n-1}\{h(X_k)-\pi(h)\} \overset{\mathscr{L}_{\mathbb{P}_\pi}}{\rightsquigarrow} \mathscr{N}(0,\sigma^2)\,,$$
$$\lim_{n\to\infty}\|\xi P^n-\pi\|_{\mathrm{TV}}=0$$

Then, we have

$$n^{-1/2}\sum_{k=0}^{n-1}\{h(X_k)-\pi(h)\} \overset{\mathscr{L}_{\mathbb{P}_\xi}}{\rightsquigarrow} \mathscr{N}(0,\sigma^2)\,.$$

PROOF.

Without loss of generality, we assume that $\pi(h)=0$. Our goal is to show that

$$\limsup_{n\to\infty}\left|\mathbb{E}_\xi\left[\mathrm{e}^{\mathrm{i}u\left\{n^{-1/2}\sum_{k=0}^{n-1}h(X_k)\right\}}\right]-\mathrm{e}^{-\sigma^2/2}\right|=0. \tag{2.8}$$

Fix $p>0$. For any $n\geqslant p$, we decompose the error as

$$\left|\mathbb{E}_\xi\left[\mathrm{e}^{\mathrm{i}u\left\{n^{-1/2}\sum_{k=0}^{n-1}h(X_k)\right\}}\right]-\mathrm{e}^{-\sigma^2/2}\right|\leqslant A_n+B_n+C_n+D_n,$$

where

$$A_n=\mathbb{E}_\xi\left[\left|\mathrm{e}^{\mathrm{i}u\left\{n^{-1/2}\sum_{k=0}^{n-1}h(X_k)\right\}}-\mathrm{e}^{\mathrm{i}u\left\{n^{-1/2}\sum_{k=p}^{n-1}h(X_k)\right\}}\right|\right],$$
$$B_n=\left|\mathbb{E}_\xi\left[\mathrm{e}^{\mathrm{i}u\left\{n^{-1/2}\sum_{k=p}^{n-1}h(X_k)\right\}}\right]-\mathbb{E}_\pi\left[\mathrm{e}^{\mathrm{i}u\left\{n^{-1/2}\sum_{k=p}^{n-1}h(X_k)\right\}}\right]\right|,$$
$$C_n=\mathbb{E}_\pi\left[\left|\mathrm{e}^{\mathrm{i}u\left\{n^{-1/2}\sum_{k=p}^{n-1}h(X_k)\right\}}-\mathrm{e}^{\mathrm{i}u\left\{n^{-1/2}\sum_{k=0}^{n-1}h(X_k)\right\}}\right|\right],$$
$$D_n=\left|\mathbb{E}_\pi\left[\mathrm{e}^{\mathrm{i}u\left\{n^{-1/2}\sum_{k=0}^{n-1}h(X_k)\right\}}\right]-\mathrm{e}^{-\sigma^2/2}\right|.$$

By assumption, we have $\lim_{n\to\infty}D_n=0$. Moreover,

$$A_n=\mathbb{E}_\xi\left[\left|\mathrm{e}^{\mathrm{i}u\left\{n^{-1/2}\sum_{k=0}^{p-1}h(X_k)\right\}}-1\right|\right]\xrightarrow[n\to\infty]{}0,$$

by the dominated convergence theorem. By the same argument, we also have $\lim_{n\to\infty}C_n=0$.

Next, define the function

$$\phi(x)=\mathbb{E}_x\left[\mathrm{e}^{\mathrm{i}u\left\{n^{-1/2}\sum_{\ell=0}^{n-p-1}h(X_\ell)\right\}}\right].$$

With this notation, the Markov property allows to rewrite B_n as

$$B_n=|\mathbb{E}_\xi[\phi(X_p)]-\mathbb{E}_\pi[\phi(X_p)]|.$$

Writing $\phi=\mathscr{R}e\,(\phi)+\mathrm{i}\mathscr{I}m\,(\phi)$ and noting that $|\phi|\leqslant 1$, we deduce that

$$B_n\leqslant 2\,\|\xi P^p-\pi\|_{\mathrm{TV}}\,.$$

Finally, we obtain

$$\limsup_{n\to\infty}\left|\mathbb{E}_\xi\left[\mathrm{e}^{\mathrm{i}u\left\{n^{-1/2}\sum_{k=0}^{n-1}h(X_k)\right\}}\right]-\mathrm{e}^{-\sigma^2/2}\right|\leqslant 2\,\|\xi P^p-\pi\|_{\mathrm{TV}}\,.$$

Since p is arbitrary and $\lim_{p\to\infty}\|\xi P^p-\pi\|_{\mathrm{TV}}=0$, this proves (2.8) and concludes the proof. ■

Chapter 3

Spectral theory

Time schedule (**Note** 3): Last session.

3.1 Markov operator

We define $L_2(\pi)$ as the set of measurable functions f on X such that $\pi(f^2) < \infty$. The space $(L_2(\pi), \|\cdot\|)$, where the norm is induced by the inner product on $L_2(\pi)$, $\langle f, g\rangle = \int \pi(\mathrm{d}x) f(x) g(x)$, is a Hilbert space.

Bounded linear operator

LEMMA 3.1 . Let P be a Markov kernel on $(\mathsf{X}, \mathscr{X})$ admitting π as an invariant probability measure. Then

$$P : f \mapsto Pf$$

is a bounded linear operator on $L_2(\pi)$. Moreover, $|||P||| = 1$.

Hence, P induces a bounded linear operator on $L_2^0(\pi)$ and for notational convenience, in what follows, we use the same notation P, seen either as a Markov kernel or as an operator on $L_2^0(\pi)$.

PROOF. For any $f \in L_2(\pi)$, we have $\pi\left[(Pf)^2\right] \leqslant \pi(P[f^2]) = \pi(f^2)$, which shows that P maps $L_2(\pi)$ into itself. The operator is clearly linear. The previous inequality can also be written as $\|Pf\|^2 \leqslant \|f\|^2$, and therefore $|||P||| \leqslant 1$. This shows that P is a bounded linear operator on $L_2(\pi)$. Since $P1 = 1$, we obtain $\|P1\| = \|1\|$, and hence $|||P||| = 1$. ■

Most of the time, we work with real-valued functions. When studying the spectrum and the resolvent set, we implicitly consider the complexification of $L_2(\pi)$, in which case, the inner-product will be $\langle f, g\rangle = \int \pi(\mathrm{d}x) \bar{f}(x) g(x)$. Moreover, in these lecture notes, most of the results, although stated for the Markov operator P actually hold more generally for any bounded linear operator. We focus on Markov operators only to avoid unnecessary generality. We denote by $\mathsf{BL}_2(\pi)$ the set of bounded linear operators on $L_2(\pi)$. We define:

- $\mathrm{Spec}(P) = \{\lambda \in \mathbb{C} : \lambda I - P \text{ is not invertible}\}$, the spectrum of P.
- $\mathrm{Spec}_p(P) = \{\lambda \in \mathbb{C} : \mathrm{Ker}[\lambda I - P] \neq \{0\}\}$, the point spectrum of P.

Clearly, $\mathrm{Spec}_p(P) \subset \mathrm{Spec}(P)$. Moreover, if $S \in \mathsf{BL}_2(\pi)$ with $|||S||| < 1$, then the series $\sum_k S^k$ is normally convergent and can be shown to be the inverse of $I - S$. It follows that for any $\lambda \in \mathbb{C}$ such that $|\lambda| > 1$, $\lambda I - P = \lambda(I - P/\lambda)$ is invertible, and therefore $\mathrm{Spec}(P) \subset \bar{\mathrm{B}}(0, 1)$.

The resolvent set of P is defined by $\mathrm{Res}(P) = \mathrm{Spec}(P)^c$. It is an open set. Indeed, if S is invertible, then writing $T = S(S^{-1}(T-S)+I)$ and taking T sufficiently close to S, we see that T is invertible with inverse $(I+S^{-1}(T-S))^{-1}S^{-1}$.

By definition, an eigenvalue λ of P is an element of the point spectrum; its multiplicity is $\mathrm{Dim}(\lambda I - P)$.

Note that 1 is an eigenvalue of P with multiplicity 1. Indeed, assume that there exists a function $f \in L_2(\pi)$ satisfying $Pf = f$. Then $\pi(f^2) = \pi([Pf]^2) \leqslant \pi(P[f^2]) = \pi(f^2)$ which implies that we have equality in the Cauchy Schwarz inequality: for π-almost all $x \in \mathsf{X}$, $(Pf(x))^2 = P[f^2](x)$ and hence, f is $\pi - a.s.$ constant. (Concerning this argument, see also the comments in the appendix).

We thus obtain the orthogonal decomposition $L_2(\pi) = \mathrm{Span}(1) \overset{\perp}{\oplus} L_2^0(\pi)$, where $L_2^0(\pi) = \{f \in L_2(\pi) : \pi(f) = 0\}$ is a closed subspace, invariant under P.

We are therefore interested in the asymptotic behaviour of

$$\sup_{f \in L_2^0(\pi)} \|P^n f\| = |||P^n|||_{L_2^0(\pi)} = \sup_{h \in L_2(\pi)} \|P^n h - \pi(h)\|.$$

For convenience, we set $H = L_2(\pi)$ and $H_0 = L_2^0(\pi)$.

THEOREM 3.2 . Defining the spectral radius by $\mathrm{Spec.Rad.}(P|_{H_0}) = \{|\lambda| : \lambda \in \mathrm{Spec}(P|_{H_0})\}$, we have

$$\mathrm{Spec.Rad.}(P|_{H_0}) = \lim_n |||P^n|||_{H_0}^{1/n}.$$

PROOF. Let A denote the left-hand side and B the right-hand side. The existence of the limit appearing in the expression of B follows from Fekete's lemma, since by setting $a_n = |||P^n|||$, one has $\log a_{p+q} \leqslant \log a_p + \log a_q$, which implies that $\lim \log a_n/n$ converges, its limit being equal to $\inf_n \log a_n/n$, a limit which may in fact be $-\infty$.

Let us now show that $A \leqslant B$, which is the easier direction. If we choose $\lambda \in \mathbb{C}$ such that $|\lambda| > B$, then the series $\sum_n (P/\lambda)^n$ converges normally and is the inverse of $I - P/\lambda$, which shows that $\lambda I - P$ is invertible. Hence λ belongs to the resolvent set. Therefore $A \leqslant |\lambda|$. Finally, $A \leqslant B$.

We now show $B \leqslant A$. Let us take $\lambda \in \mathbb{C}$ such that $|\lambda| > A$. Then one may define $\phi(z) = (I - zP)^{-1}$ for all $|z| < \lambda^{-1}$. We now prove the Cauchy integral formula. Readers may safely skip this proof on a first reading; it is included for completeness and for its elegance and usefulness. For any $r < |\lambda|^{-1}$, and any $z_0 \in \mathbb{C}$ with $|z_0| < r$, define $g(\beta) = \int_0^{2\pi} \frac{\phi(\beta re^{i\theta} + (1-\beta)z_0)}{re^{i\theta} - z_0} re^{i\theta}\,\mathrm{d}\theta$. Since

$$g'(\beta) = \int_0^{2\pi} \phi'(\beta re^{i\theta} + (1-\beta)z_0)\, re^{i\theta}\,\mathrm{d}\theta = \left[\frac{\phi(\beta re^{i\theta} + (1-\beta)z_0)}{i\beta}\right]_0^{2\pi} = 0,$$

we deduce that g is constant and in particular that $g(0) = g(1)$, which can be rewritten as

$$\phi(z_0) \int_0^{2\pi} \frac{1}{1 - (z_0/r)e^{-i\theta}}\,\mathrm{d}\theta = \int_0^{2\pi} \frac{\phi(re^{i\theta})}{1 - (z_0/r)e^{-i\theta}}\,\mathrm{d}\theta.$$

Expanding inside the integral, $(1 - (z_0/r)e^{-i\theta})^{-1} = \sum_n (z_0/r)^n e^{-i\theta n}$, and interchanging (legitimately) the series and the integral, we obtain, for z_0 in a neighborhood of 0,

$$\phi(z_0) = \sum_{n=0}^{\infty} z_0^n \frac{1}{2\pi} \frac{\int_0^{2\pi} \phi(re^{i\theta}) e^{-in\theta}\,\mathrm{d}\theta}{r^n} = (I - z_0 P)^{-1} = \sum_{n=0}^{\infty} z_0^n P^n.$$

At this point, we may equate the Taylor expansions, which yields

$$P^n = \frac{1}{2\pi} \frac{\int_0^{2\pi} \phi(re^{i\theta}) e^{-in\theta}\,\mathrm{d}\theta}{r^n}$$

for all $n \in \mathbb{N}$. Since ϕ is continuous and therefore bounded on any compact set, there exists a constant C such that $|||P^n||| \leqslant C/r^n$, and hence $\limsup_n |||P^n|||^{1/n} \leqslant 1/r$. As this holds for any $r < |\lambda|^{-1}$, we obtain $B \leqslant |\lambda|$. Finally, $B \leqslant A$, and the proof is complete. ∎

Comment on the proof. To be precise, a careful reading of this proof shows that the mapping $\phi : \mathbb{C} \to L_2^0(\pi)$ should be complex differentiable, that is, holomorphic on the ball $B(0, |\lambda|^{-1})$. This property can be verified directly on the resolvent set of P. Indeed, writing

$$I - (z+h)P = (I - zP)\left[(I - zP)^{-1}(-hP) + I\right]$$

we see that, for h sufficiently small, the inverse of $I - (z+h)P$ is given by

$$(I - (z+h)P)^{-1} = \left[(I - zP)^{-1}(-hP) + I\right]^{-1}(I - zP)^{-1} = \sum_{k=0}^{\infty}(-h)^k\left[(I - zP)^{-1}P\right]^k(I - zP)^{-1}.$$

This expansion implies $\phi(z+h) = \phi(z) - h(I - zP)^{-1}P(I - zP)^{-1} + o(|h|)$, showing that ϕ is indeed holomorphic at z.

3.2 Reversibility and self-adjointness.

We now assume that $\boxed{P \text{ is } \pi\text{-reversible}}$, that is $\pi(\mathrm{d}x)P(x, \mathrm{d}y) = \pi(\mathrm{d}y)P(y, \mathrm{d}x)$. Then, obviously P is self-adjoint, i.e., $\langle Pf, g\rangle = \langle f, Pg\rangle$. Note that since P is self-adjoint, $\langle Pf, f\rangle \in \mathbb{R}$.

THEOREM 3.3 . If P is reversible, then

$$\begin{aligned}|||P|||_{L_2^0(\pi)} &= \sup_{\|f\|\leqslant 1, f\in L_2^0(\pi)} \sqrt{\langle Pf, Pf\rangle} = \sup_{\|f\|\leqslant 1, f\in L_2^0(\pi)} |\langle Pf, f\rangle| \\ &= \lim_{n\to\infty} |||P^n|||_{L_2^0(\pi)}^{1/n} = \sup\{|\lambda| : \lambda \in \operatorname{Spec}(P|_{L_2^0(\pi)})\}.\end{aligned}$$

PROOF. Let us denote the previous equalities by $A = B = C = D = E$. By definition of the triple norm, we clearly have $A = B$. Moreover, $C \leqslant B$ follows from the Cauchy–Schwarz inequality. To show that $B \leqslant C$, write $\langle Pf, \underbrace{Pf/\|Pf\|}_{g}\rangle$ for f of norm 1 and express this quantity in terms of $\langle P(f \pm g), f \pm g\rangle$ using the parallelogram identity (and the fact that P is self-adjoint). We obtain, noting that $\langle Pf, g\rangle = \|Pf\| \in \mathbb{R}$,

$$\|Pf\| = |\langle Pf, g\rangle| = \left|\frac{1}{4}\left[\langle P(f+g), f+g\rangle - \langle P(f-g), f-g\rangle\right]\right| \leqslant \frac{C}{4}\left(\|f+g\|^2 + \|f-g\|^2\right) \leqslant C.$$

Hence $B = C$. Thus $A = B = C$ and $D = E$ (by Theorem 3.2).

Finally, it remains to show that $A = D$. Using the identity $\langle Pf, Pf\rangle = \langle P^2 f, f\rangle$ in the equalities $A = B = C$, we obtain $|||P|||^2 = |||P^2|||$. By induction, this yields $|||P|||^{2^k} = |||P^{2^k}|||$. Therefore, $|||P||| = |||P^{2^k}|||^{1/2^k}$, and letting k tend to infinity, we conclude that $A = D$.

■

The following theorem is stated for the Markov operator P, but we emphasize that it applies more generally to any self-adjoint bounded operator.

THEOREM 3.4 . If P is self-adjoint, then its eigenvalues are real and

$$\operatorname{Spec}(P|_{L_2^0(\pi)}) \subset [m, M],$$

where

$$m = \inf_{f\in L_2^0(\pi), \|f\|\leqslant 1} \langle Pf, f\rangle, \qquad M = \sup_{f\in L_2^0(\pi), \|f\|\leqslant 1} \langle Pf, f\rangle,$$

and both bounds m and M belong to the spectrum of P.

PROOF. Let $z \notin [m,M]$ and let us show that it belongs to the resolvent set. Let $f \in H_0$. Choose α such that $\langle(\alpha I - P)f, f\rangle = 0$. Then

$$\|(zI-P)f\|^2 = \|(\alpha I - P)f\|^2 + |z-\alpha|^2\|f\|^2 \geqslant |z-\alpha|^2\|f\|^2 \geqslant \Delta\|f\|^2,$$

where we have set $\Delta = d(z,[m,M]) > 0$. This simple inequality shows that z belongs to the resolvent set. Indeed, it successively implies that $\mathrm{Ker}(zI-P) = \{0\}$, that $\mathrm{Ran}(zI-P)$ is closed, and that if $g \in \mathrm{Ran}(zI-P)^\perp$ then $(\bar{z}I - P)g = 0$. Applying the above inequality with $(\bar{z}, g)$ instead of (z, f), we obtain that $g = 0$. Hence $zI - P$ is invertible and, moreover, its inverse is bounded (again by the same inequality). Therefore, z belongs to the resolvent set. This proves the first part of the theorem.

Finally, suppose that $M = |||P|||$. Choose f_n of norm 1 such that $\langle Pf_n, f_n\rangle \to M$. Then

$$\begin{aligned}\|(MI-P)f_n\|^2 &= M^2 + \|Pf_n\|^2 - 2M\langle f_n, Pf_n\rangle \\ &\leqslant 2M^2 - 2M\langle f_n, Pf_n\rangle \to 2M^2 - 2M^2 = 0.\end{aligned}$$

Thus $MI - P$ is not invertible (otherwise, $1 = \|f_n\|^2 \leqslant |||(MI-P)^{-1}|||\,\|(MI-P)f_n\| \to 0$). Therefore $M \in \mathrm{Spec}(P)|_{H_0}$.

So far, the proof was written with P but this also holds for any self-adjoint bounded operator. This remark allows to replace P by $Q = MI - P$, we obtain

$$\sup_{f \in L_2^0(\pi),\ \|f\| \leqslant 1} \langle Qf, f\rangle = M - m,$$

$$\inf_{f \in L_2^0(\pi),\ \|f\| \leqslant 1} \langle Qf, f\rangle = 0.$$

It follows that $|||Q||| = M - m$ and consequently (by the previous argument applied with P replaced by $MI - P$) that $M - m \in \mathrm{Spec}(MI-P)|_{H_0}$. This means that

$$(M-m)I - (MI-P) = -(mI-P)$$

is not invertible. Hence we have shown that $m \in \mathrm{Spec}(P)$. If we now suppose that $-m = |||P|||$, we apply the same reasoning by replacing P with $-P$.

■

A careful inspection of the proof actually shows that

- $\mathrm{Spec}(P|_{L_2^0(\pi)}) \subset \overline{\{\langle f, Pf\rangle\ :\ f \in L_2^0(\pi), \|f\| \leqslant 1\}}$.

3.2.1 Spectral measure

THEOREM 3.5 . If P is self-adjoint, then for any $f \in L_2^0(\pi)$ there exists a finite (nonnegative) measure μ_f supported on $\mathrm{Spec}(P|_{L_2^0(\pi)}) \subset [-1,1]$ such that, for all $n \in \mathbb{N}$,

$$\langle f, P^n f\rangle = \int_{-1}^{1} x^n \mu_f(\mathrm{d}x).$$

Taking $n = 0$ yields $\mu_f([-1,1]) = \pi(f^2) = \mathrm{Var}_\pi(f)$.

The proof of the theorem is omitted; we only sketch the main ideas. We first give a precise meaning to the map $\phi \mapsto \langle f, \phi(P)f\rangle$: it is initially defined for polynomials and then extended to any continuous function ϕ on $\mathrm{Spec}(P|L_2^0(\pi))$ by density of the polynomials, using the Stone–Weierstrass theorem. This construction yields a nonnegative continuous linear functional on the space of continous functions on the compact set $\mathrm{Spec}(P|L_2^0(\pi))$, equipped with the supremum norm. The existence of the spectral measure then follows from the Riesz representation theorem.

This theorem allows one to replace P^n by the scalar x^n, which greatly simplifies many arguments and is justified by the spectral theorem. Note that μ_f may charge the points $\{1\}$ or $\{-1\}$.

We define

- $\mathrm{Abs.Spec.Gap}(P) = 1 - \sup\{|\lambda| : \lambda \in \mathrm{Spec}(P|_{H_0})\}$,
- $\mathrm{SpecGap}(P) = 1 - \sup\{\lambda : \lambda \in \mathrm{Spec}(P|_{H_0})\}$.

Moreover, we have the following result.

PROPOSITION 3.6 . Let P be a reversible Markov kernel.

$$\begin{aligned}\mathrm{SpecGap}(P) &= \inf_{f\in H_0, \|f\|\leqslant 1} \langle f,f\rangle - \langle Pf,f\rangle \\ &= \inf_{f\in H_0, \|f\|\leqslant 1} \langle (I-P)f,f\rangle \\ &= \inf_{f\in H_0, \|f\|\leqslant 1} \frac{1}{2}\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)(f(y)-f(x))^2.\end{aligned}$$

Some comments. To see the first equality, recall that $I-P$ being reversible, applying Theorem 3.4, with P replaced by $I-P$,

$$\mathrm{SpecGap}(P) = \inf\{\lambda \in \mathbb{C} \,:\, \lambda \in \mathrm{Spec}(I-P)|_{H_0}\} = \inf\left\{\langle f,(I-P)f\rangle \,:\, f \in L_2^0(\pi), \|f\| \leqslant 1\right\}$$

The second equality is immediate. The third follows by expanding $\frac{1}{2}\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)(f(y)-f(x))^2$ and using that P is π-invariant. The standard notation for the Dirichlet form is $\mathscr{E}(f,g) = \langle f,(I-P)g\rangle$. We thus have two equivalent expressions for the Dirichlet form $\mathscr{E}(f,f)$:

$$\mathscr{E}(f,f) = \langle f,(I-P)f\rangle = \frac{1}{2}\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)(f(y)-f(x))^2.$$

If P is positive, that is, $\langle f,Pf\rangle \geqslant 0$ for all $f \in L_2^0(\pi)$, then $\mathrm{Spec}(P|_{H_0}) \subset [0,1]$ and the spectral gap coincides with the absolute spectral gap, which allows to combine Theorem 3.3 and Proposition 3.6.

3.3 Comparison of asymptotic behavior for two Markov kernels

As a byproduct of the different expressions of the spectral gap in Proposition 3.6, we have

COROLLARY 3.7 . Let P and Q be reversible kernels and assume that $P \succeq Q$ in the sense of covariance ordering, that is, $\langle Pf,f\rangle \leqslant \langle Qf,f\rangle$ for all $f \in H_0$. Then:

- $\mathrm{SpecGap}(P) \geqslant \mathrm{SpecGap}(Q)$.
- $\lim_{n\to\infty} \mathrm{Var}_P\left[\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(X_k)\right] \leqslant \lim_{n\to\infty} \mathrm{Var}_Q\left[\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(X_k)\right]$, where both chains are started from the stationnary distribution π.

The first bullet follows immediately from covariance ordering: $\langle Pf,f\rangle \leqslant \langle Qf,f\rangle$ is equivalent to $\langle (I-Q)f,f\rangle \leqslant \langle (I-P)f,f\rangle$. If the spectral gaps are positive, then P converges geometrically to π at a faster rate than Q.

The second is more delicate and corresponds to the proof of Tierney (1998). It shows that the Monte Carlo estimator of $\pi(f)$ has a smaller asymptotic variance when using P rather than Q.

How can covariance ordering be verified? In many cases, it suffices to show that for all $(x,A) \in \mathsf{X} \times \mathscr{X}$,

$$P(x, A \setminus \{x\}) \geqslant Q(x, A \setminus \{x\}),$$

which is known as Peskun ordering. Indeed,

$$\frac{1}{2}\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)(f(y)-f(x))^2 \geqslant \frac{1}{2}\int \pi(\mathrm{d}x)Q(x,\mathrm{d}y)(f(y)-f(x))^2,$$

which implies $\langle (I-P)f, f\rangle \geqslant \langle (I-Q)f,f\rangle$ and hence $P \succeq Q$.

The following exercise allows to prove the second bullet in Corollary 3.7.

EXERCISE 2 . Let P be a reversible Markov kernel and let $f \in H_0$. Define $A_n = \mathrm{Var}_P\left[\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(X_k)\right]$ where the chain starts from the stationary distribution π.

1. Show that
$$A_n = \langle f,f\rangle + 2\sum_{\ell=1}^{n-1}\frac{n-\ell}{n}\langle f, P^\ell f\rangle$$
2. Deduce that there exists a finite non-negative measure μ_f on $[-1,1]$ such that
$$A_n = \int_{[-1,1]} w_n(x)\mu_f(\mathrm{d}x) \quad \text{with} \quad w_n(x) = \frac{1+x}{1-x} - \frac{2x}{(1-x)^2}\frac{1-x^n}{n}$$
3. By splitting the integral on $[-1,0]$ and $(0,1]$ show that $\lim_{n\to\infty} A_n$ exists and is equal to $-\langle f,f\rangle + 2\int_{-1}^{1}\frac{1}{1-x}\mu_f(\mathrm{d}x)$.

We now consider two π-reversible kernels P_0, P_1 such that $P_0 \succeq P_1$ according to the covariance ordering. Define $P_\alpha = (1-\alpha)P_0 + \alpha P_1$ for $\alpha \in (0,1]$ and for any $\lambda \in (0,1)$, write $H_\lambda(\alpha) = (I - \lambda P_\alpha)^{-1}$.

4. Show that H'_λ (the right derivative of H_λ) is equal to
$$H'_\lambda(\alpha) = \lambda(I-\lambda P_\alpha)^{-1}(P_1 - P_0)(I - \lambda P_\alpha)^{-1}.$$
5. Using that P_0, P_1 are π-reversible, show that $\langle f, H'_\lambda(\alpha) f\rangle \geqslant 0$.
6. Deduce $\langle f, H_\lambda(0) f\rangle \leqslant \langle f, H_\lambda(1) f\rangle$.
7. Letting $\lambda \to 1$, deduce that
$$\lim_{n\to\infty}\mathrm{Var}_{P_0}\left[\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(X_k)\right] \leqslant \lim_{n\to\infty}\mathrm{Var}_{P_1}\left[\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(X_k)\right].$$

A Appendix

The following lemma can be used to show that 1 has multiplicity 1. The proof, however, is more involved (although also more general since it applies to $f \in L_1(\pi)$ rather than $f \in L_2(\pi)$) than the nice elementary argument presented in the Lecture Notes (originally due to an MDA student in 2025). We include the proof for $f \in L_1(\pi)$ below for completeness.

LEMMA .8 . If P admits a unique invariant probability measure π, then any harmonic function $f \in L_1(\pi)$ is $\mathbb{P}_\pi - a.s.$ constant.

PROOF. From $Pf = f$, we deduce that $\{f(X_n)\ :\ n \in \mathbb{N}\}$ is a martingale and that $\sup_{n\in\mathbb{N}} \mathbb{E}_\pi[f(X_n)^+] = \pi(f^+) < \infty$, so that it converges $\mathbb{P}_\pi$-almost surely.

We argue by contradiction. If f is not $\mathbb{P}_\pi$-almost surely constant, then there exist $a < b$ such that $\pi(f < a) > 0$ and $\pi(f > b) > 0$. Then, $\mathbb{P}_\pi$-almost surely,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{\{f(X_k)<a\}} = \pi(f < a) > 0.$$

Hence $\#\{k : f(X_k) < a\} = \infty$, $\mathbb{P}_\pi - a.s.$, and similarly $\#\{k : f(X_k) > b\} = \infty$, $\mathbb{P}_\pi - a.s.$, which contradicts the almost sure convergence of $\{f(X_n)\ :\ n \in \mathbb{N}\}$. ■