31. REVERSIBILITY

Here we continue the discussion of stationary measures with attention focused on the particular case of measures that are reversible.

31.1 Two-sided chain and its reversal.

We start with the observation that the mere existence of a stationary distribution allows us to extend the chain to negative infinity:

Lemma 31.1 (Two-sided Markov chain) *Given a standard Borel space (S,Σ), let p be a transition probability and let $\mu \in \mathscr{I}_1$. Then there exists a probability space $(\Omega,\mathcal{F},P)$ supporting a sequence $\{X_n\}_{n\in\mathbb{Z}}$ of S-valued random variables such that*

$$\forall n \in \mathbb{Z}\colon\ \ P(X_n \in \cdot) = \mu(\cdot) \tag{31.1}$$

and, denoting $\mathcal{F}_n := \sigma(X_k\colon k \leqslant n)$,

$$\forall n \in \mathbb{Z}\colon\ \ P(X_{n+1} \in \cdot \,|\, \mathcal{F}_n) = p(X_n,\cdot) \quad \text{a.s.} \tag{31.2}$$

Proof. We proceed as in the argument sketched at the beginning of the proof of Theorem 29.3. Set $(\Omega,\mathcal{F}) := (S^{\mathbb{Z}},\Sigma^{\otimes\mathbb{Z}})$ and let $X_n\colon S^{\mathbb{Z}} \to S$ be the projection on the n-th coordinate. For $n \geqslant 0$ denote $\mathcal{G}_n := \sigma(X_{-n},\dots,X_n)$ and for $A \in \mathcal{G}_n$ set

$$P_n(A) := \int_A \mu(\mathrm{d}x_{-n}) \prod_{j=-n}^{n-1} p(x_j,\mathrm{d}x_{j+1}) \tag{31.3}$$

Then P_n is a probability measure on $(\Omega,\mathcal{F}_n)$. Using that $\mu \in \mathscr{I}_1$ and that $p(x_n,\cdot)$ is a probability we then verify $P_{n+1}(A) = P_n(A)$ and so $\{P_n\}_{n\geqslant 0}$ is a consistent family of measures. The Kolmogorov Extension Theorem ensures that these are restrictions of a unique probability measure P on $\sigma(\bigcup_{n\geqslant 0}\mathcal{G}_n) = \Sigma^{\otimes\mathbb{Z}}$.

The condition (31.1) is verified readily from (31.3) and the fact that $\mu \in \mathscr{I}_1$. For (31.2) the formula (31.3) gives

$$\forall m \leqslant n\colon \quad P(X_{n+1} \in \cdot \,|\, \mathcal{F}_{m,n}) = p(X_n,\cdot) \quad \text{a.s.} \tag{31.4}$$

with $\mathcal{F}_{m,n} := \sigma(X_k\colon m \leqslant k \leqslant n)$. Taking $m \to -\infty$ with the help of the Lévy Forward Theorem and $\mathcal{F}_n = \sigma(\bigcup_{m\leqslant n}\mathcal{F}_{m,n})$ we then get (31.2). $\square$

With paths of the Markov chain extended to negative times, a natural question is then: What is the law of the reversed path $\{X_{-n}\}_{n\in\mathbb{Z}}$? This is answered easily for countable-stated chains:

Lemma 31.2 *Let S be countable and, given a transition kernel P and an invariant distribution μ, let $\{X_n\}_{n\in\mathbb{Z}}$ be a two-sided Markov chain such that* (31.1) *holds. Then*

$$\mathsf{Q}(x,y) := \begin{cases} \dfrac{\mu(y)}{\mu(x)}\mathsf{P}(y,x), & \text{if } \mu(x) > 0 \\ \delta_{xy}, & \text{if } \mu(x) = 0 \end{cases} \tag{31.5}$$

Then $\{X_{-n}\}_{n\in\mathbb{Z}}$ is Markov chain with transition kernel Q.

Proof. For $x \in S$ such that $\mu(x) > 0$, the fact that μ is stationary gives

$$\sum_{y \in S} \mathsf{Q}(x,y) = \frac{1}{\mu(x)} \sum_{y \in S} \mu(y)\mathsf{P}(y,x) = \frac{1}{\mu(x)}\mu(x) = 1 \tag{31.6}$$

and so $\mathsf{Q}(x,\cdot)$ is a probability mass function. For $\mu(x) = 0$ this follows directly and so Q is indeed a transition kernel.

Given any integers $m < n$ and $x_m, \dots, x_n \in S$, observe that the stationarity of μ forces $\mu(x_i) = 0$ for all $i < n$ once $\mu(x_n) = 0$. Assuming $\mu(x_n) > 0$ we can thus write

$$\begin{aligned} \mu(x_m) \prod_{i=m+1}^{n} \mathsf{P}(x_{i-1}, x_i) &= \mu(x_m)\mathsf{P}(x_m, x_{m+1}) \dots \mathsf{P}(x_{n-1}, x_n) \\ = \frac{\mu(x_m)}{\mu(x_{m+1})}\mathsf{P}(x_m, x_{m+1}) \dots \frac{\mu(x_{n-1}}{\mu(x_n)}\mathsf{P}(x_{n-1}, x_n)\mu(x_n) &= \mu(x_n) \prod_{i=m}^{n-1} \mathsf{Q}(x_{i+1}, x_i) \end{aligned} \tag{31.7}$$

Since the quantities on the extreme ends vanish when $\mu(x_n) = 0$, the equality holds in general. This means that the finite dimensional distributions of $\{X_{-n}\}_{n \in \mathbb{Z}}$ are indeed distributed as a Markov chain with transition kernel Q. Standard extension arguments then extend this to the full distribution on $(S^{\mathbb{Z}}, \Sigma^{\otimes \mathbb{Z}})$. □

We call the chain with the transition kernel Q the *reversed chain*. Note that the definition of Q makes sense, and the reversed chain is thus well defined, even if μ is just a stationary measure (i.e., if $\mu(S) = \infty$) although the connection to the reversal of the Markov chain is less clear (we need to "start" the chain from the infinite measure).

In uncountable state spaces we proceed quite similarly. A key technical issue is the definition of an analogue of Q for which we will need to assume that S has the structure of a standard Borel space:

Lemma 31.3 *Let (S, Σ) be a standard Borel space and let p be a transition probability on S. Given $\mu \in \mathscr{I}$ there exists a transition probability q such that*

$$\forall A, B \in \Sigma\colon \quad \int_A \mu(\mathrm{d}x) p(x, B) = \int_B \mu(\mathrm{d}y) q(y, A) \tag{31.8}$$

Moreover, if $\{X_n\}_{n \in \mathbb{Z}}$ is the two-sided Markov chain such that (31.1–31.2) hold, then $\{X_{-n}\}_{n \in \mathbb{Z}}$ is a Markov chain with transition probability q.

Proof. Let ν denote the joint distribution of (X_0, X_1) on $(S \times S, \Sigma \otimes \Sigma)$. Note that then for each $A, B \in \Sigma$,

$$\nu(A \times B) = \int_A \mu(\mathrm{d}x) p(x, B) \tag{31.9}$$

The fact that S is standard Borel implies that there exists a regular conditional distribution q of X_0 given $\sigma(X_1)$; i.e., a map $q\colon S \times \Sigma \to [0, 1]$ which is Σ-measurable in the first coordinate and a probability measure on (S, Σ) in the second coordinate such that, for each $A, B \in \Sigma$,

$$\int_B \mu(\mathrm{d}y) q(y, A) = \int 1_B(y) q(y, A) \mathrm{d}\mu = \nu(A \times B) \tag{31.10}$$

In conjunction with (31.9) this proves the equality (31.8).

Using (31.8) we now check that, for any $m < n$ and any $A_m, \dots, A_n \in \Sigma$, we have

$$\int_{A_m \times \dots \times A_n} \mu(\mathrm{d}x_m) \prod_{i=m+1}^{n} p(x_{i-1}, \mathrm{d}x_i) = \int_{A_m \times \dots \times A_n} \mu(\mathrm{d}x_n) \prod_{i=m}^{n-1} q(x_{i+1}, \mathrm{d}x_i) \tag{31.11}$$

which is the analogue of (31.7). The rest of the argument is the exactly as for countable S and so we omit it. □

Note that $\nu(A \times B) \leqslant \mu(B)$ which means that we can always try to define q as the Radon-Nikodym derivative $\frac{\mathrm{d}\nu(A\times \cdot)}{\mathrm{d}\mu}$. However, while this is automatically a measurable function we cannot generally guarantee that this is a measure as a function of A. This is where the technical assumption on the structure of S enters. As for the countable case, one can make sense of q even when μ is an infinite σ-finite measure. However, we will not attempt to spell out the details.

31.2 Reversible measures.

With the law of the reversed chain identified as a Markov chain, a natural question is: For what Markov chains does the reverse chain have the same law as the primal chain? This leads us to the following concept:

Definition 31.4 (Reversible measures) *Given a transition probability p on (S, Σ), a measure ν on (S, Σ) is said to be* reversible *if*

$$\forall A, B \in \Sigma\colon \quad \int_A \nu(\mathrm{d}x) p(x, B) = \int_B \nu(\mathrm{d}y) p(y, A) \tag{31.12}$$

As is checked by taking both A and B to be singletons, for countable S the above condition reduces to

$$\forall x, y \in S\colon \;\; \nu(x)\mathsf{P}(x,y) = \nu(y)\mathsf{P}(y,x) \tag{31.13}$$

where ν denotes the probability mass function associated with ν and P is the transition kernel. Note that (31.13) holds automatically when $x = y$ so we only need to check it for $x \neq y$. For the same reason, we only need to check this for x and y such that at least one of $\mathsf{P}(x,y) \neq 0$ and $\mathsf{P}(y,x) \neq 0$ hold. Another observation is:

Lemma 31.5 *A reversible measure is automatically stationary.*

Proof. Take $A := S$ in (31.12) and use that $p(y, S) = 1$. □

This suggests that, in order to identify stationary measures of a chain, we first search for the reversible once. This gives us:

Another proof of Lemma 30.9. Observe that ν defined in (30.26) obeys

$$\nu(k+1) = \nu(k)\frac{\alpha_k}{\beta_{k+1}} = \nu(k)\frac{\mathsf{P}(k,k+1)}{\mathsf{P}(k+1,k)} \tag{31.14}$$

Since all transition happens between pairs of the form $(k, k+1)$, this shows that ν is reversible and, by Lemma 31.5, thus stationary. □

Another example is the random walk on a weighted graph; see Example 29.13. Here (using the notations there), $\pi(x)\mathsf{P}(x,y) = a(x,y)$ and π is a reversible measure as soon as $a(x,y) = a(y,x)$ for all pairs (x,y). We will return to this example when we discuss connection of Markov chains to electric networks.

While a passage through reversible measures is attractive, we should bear in mind that reversibility is a special condition and that most Markov chains do not admit reversible measures, period. To show that the situation is even more interesting, recall the example of a biased simple random walk on $\mathbb{Z}$ discussed in Lemma 30.11 where we showed that every stationary measure is a linear combination $\alpha\nu + \beta\nu'$ (with non-negative α and β) of the constant measure

$$\forall k \in \mathbb{Z}\colon\ \ \nu(k) = 1 \tag{31.15}$$

and the exponentially tilted measure

$$\forall k \in \mathbb{Z}\colon\ \ \nu'(k) = \left(\frac{1-p}{p}\right)^k \tag{31.16}$$

We now readily check that ν is not reversible unless $p = 1/2$ (when it coincides with ν') while ν' is reversible for all $p \in (0,1)$.

As it turns out, for countable state spaces, the existence of reversible measures can be linked to the geometry of the graph naturally associated with the Markov chain:

Theorem 31.6 (Kolmogorov's cycle conditions) *Let S be countable and let P be an irreducible transition kernel. Then there exists a non-vanishing reversible measure if and only if*

$$\forall n \geqslant 1 \forall x_0,\dots,x_n \in S\colon\ \ x_n = x_0 \ \Rightarrow\ \prod_{i=1}^{n} \mathsf{P}(x_{i-1},x_i) = \prod_{i=0}^{n-1} \mathsf{P}(x_{i+1},x_i) \tag{31.17}$$

Under this condition, the reversible measure is determined uniquely modulo normalization.

Proof. If ν is a reversible measure then irreducibility implies $\nu(x) > 0$ for all $x \in S$. Multiplying the product on the left in (31.17) by $\nu(x_0)$ shows

$$\nu(x_0)\prod_{i=1}^{n} \mathsf{P}(x_{i-1},x_i) = \nu(x_n)\prod_{i=1}^{n} \mathsf{P}(x_i,x_{i-1}) = \nu(x_n)\prod_{i=0}^{n-1} \mathsf{P}(x_{i+1},x_i) \tag{31.18}$$

When $x_n = x_0$ we can cancel $\nu(x_0) = \nu(x_n)$ and get equality of the product.

For the converse, suppose that (31.17) hold and pick $x_0,\dots,x_n \in S$ and $x'_0,\dots,x'_m \in S$ with $x'_0 = x_0$ and $x'_m = x_n$ and

$$\prod_{i=1}^{n} \mathsf{P}(x_{i-1},x_i) > 0 \ \wedge\ \prod_{j=1}^{m} \mathsf{P}(x'_{j-1},x'_j) > 0 \tag{31.19}$$

The cycle condition (31.17) then implies

$$\left(\prod_{i=1}^{n} \mathsf{P}(x_{i-1},x_i)\right)\left(\prod_{j=1}^{m} \mathsf{P}(x'_j,x'_{j-1})\right) = \left(\prod_{j=1}^{m} \mathsf{P}(x'_{j-1},x'_j)\right)\left(\prod_{i=1}^{n} \mathsf{P}(x_i,x_{i-1})\right) \tag{31.20}$$

and so all four products are strictly positive. This now rewrites as

$$\prod_{i=1}^{n} \frac{\mathsf{P}(x_{i-1}, x_i)}{\mathsf{P}(x_{i-1}, x_i)} = \prod_{j=1}^{m} \frac{\mathsf{P}(x'_{j-1}, x'_j)}{\mathsf{P}(x'_j, x'_{j-1})} \tag{31.21}$$

implying that the product does not depend on the choice of the path from x_0 to x_n. We now fix $\nu(x_0) > 0$ and define

$$\nu(x) := \nu(x_0) \prod_{i=1}^{n} \frac{\mathsf{P}(x_{i-1}, x_i)}{\mathsf{P}(x_{i-1}, x_i)} \tag{31.22}$$

for any choice of the path $x_0, \dots, x_n = x$ with $\prod_{i=1}^{n} \mathsf{P}(x_{i-1}, x_i) > 0$, which by our previous argument ensures $\prod_{i=1}^{n} \mathsf{P}(x_i, x_{i-1}) > 0$. Such a path exists by the assumed irreducibility and so $\nu(x) > 0$ for all $x \in S$.

For any distinct $x, y \in S$ with $\mathsf{P}(y, x) > 0$ we then get

$$\nu(y) = \nu(x) \frac{\mathsf{P}(x, y)}{\mathsf{P}(y, x)} \tag{31.23}$$

by concatenating y at the end of the path $x_0, \dots, x_n = x$. This shows that $\mathsf{P}(x, y) > 0$ as well and that (31.13) holds and ν is thus a reversible measure. To get the uniqueness modulo normalization, observe that any reversible measure will obey (31.22) and so $\nu(x)/\nu(x_0)$ does not depend on the choice of the measure. □

31.3 Functional-analytic connection.

The existence of stationary measures opens up another line of approach to Markov chains based on tools from functional analysis. We start with:

Lemma 31.7 *Let p be the transition probability on (S, Σ) and let $\mu \in \mathscr{I}$. Let $\alpha \in [1, \infty]$. For each $f \in L^\alpha(\mu)$ the integral defining function $\mathsf{P}f$ in* (30.14) *is well defined and finite μ-a.e. The map $f \mapsto \mathsf{P}f$ defines a bounded linear operator $\mathsf{P}\colon L^\alpha(\mu) \to L^\alpha(\mu)$ with $\|\mathsf{P}f\|_\alpha \leqslant \|f\|_\alpha$.*

Proof. Note that for any measurable f, we have $|\mathsf{P}f(x)| \leqslant (\mathsf{P}|f|)(x)$ with the integral defining $\mathsf{P}f$ well-defined and convergent whenever $(\mathsf{P}|f|)(x) < \infty$, so for existence it suffices to show that $\mathsf{P}|f| < \infty$ μ-a.e. for all $f \in L^\alpha$. We will prove this along with the bound on the L^α-norm. Indeed, $\alpha < \infty$ we invoke Jensen's inequality to get

$$|\mathsf{P}f|^\alpha \leqslant \mathsf{P}\bigl(|f|^\alpha\bigr) \tag{31.24}$$

Combining with stationarity of μ, this now yields

$$\int |\mathsf{P}f|^\alpha \mathrm{d}\mu \leqslant \int \mathsf{P}\bigl(|f|^\alpha\bigr) \mathrm{d}\mu = \int \mu(\mathrm{d}x) p(x, \mathrm{d}y) \bigl|f(y)\bigr|^\alpha = \int \mu(\mathrm{d}y) f(y)^\alpha \tag{31.25}$$

proving the desired inequality $\|\mathsf{P}f\|_\alpha \leqslant \|f\|_\alpha$. For $\alpha = \infty$ we argue directly $\|\mathsf{P}f\|_\infty \leqslant \|f\|_\infty$. It follows that P is, on each $L^\alpha(\mu)$, an everywhere-defined linear operator with operator norm at most 1. □

We now link to the topics discussed earlier via:

Lemma 31.8 *Given a standard Borel space (S,Σ), a transition probability p and a stationary distribution μ, let q be the transition probability of the reversed chain. Then the adjoint P^+ of the operator P from Lemma 31.7 on $L^2(\mu)$ takes the form*

$$\mathsf{P}^+ f(x) = \int q(x,\mathrm{d}y) f(y), \quad \mu\text{-a.e.} \tag{31.26}$$

In particular, μ is reversible if and only if P is self-adjoint.

Proof. The space $L^2(\mu)$ is endowed with the canonical inner product $\langle f,g\rangle := \int fg\mathrm{d}\mu$. A calculation gives

$$\begin{aligned}\langle f,\mathsf{P}g\rangle &= \int \mu(\mathrm{d}x)p(x,\mathrm{d}y)f(x)g(y)\\ &= \int \mu(\mathrm{d}y)q(y,\mathrm{d}x)f(x)g(y) = \langle \mathsf{Q}f,g\rangle\end{aligned} \tag{31.27}$$

where

$$\mathsf{Q}f(y) := \int q(y,\mathrm{d}x)f(x) \tag{31.28}$$

and where the middle equality follows by a routine extension from (31.8) via the Monotone Class Theorem. (One first proves (31.27) for positive f and g using Tonelli's theorem and then argues the general case using Fubini's theorem.) The functional-analytic definition of the adjoint gives us $\mathsf{Q} = \mathsf{P}^+$ which now applies the claim. □

Clearly, the adjoint P^+ exists in all circumstances (as P is bounded) and so the main point of the representation (31.26) is that P^+ is associated with a transition probability of a Markov chain. The connection with self-adjointness is technically advantageous because it enables arguments based on the spectral theorem. On the other hand, the conclusions are then restricted to events of positive μ-measure which may not be always advantageous. (This is why the theory of Markov processes is rather built over the Banach space of continuous functions.)