30. Stationary measures

Having discussed the setting of Markov chain theory, we now move to studying the chains themselves. Here we develop the notion of stationary measures that plays a central role for both theory development and practical considerations.

30.1 Definitions.

Suppose $\{X_n\}_{n\geqslant 0}$ is a Markov chain with initial distribution μ. This tells us the law of X_0 so the natural question is: What is the law of X_n? And, does the law converge as $n \to \infty$? To answer the first question, we introduce the following notations: Given a measure μ and a transition probability p on (S, Σ), denote by $\mu\mathsf{P}$ the measure

$$\mu\mathsf{P}(A) := \int \mu(\mathrm{d}x)p(x, A) \tag{30.1}$$

This can be iterated to define $\mu\mathsf{P}^n$ recursively by $\mu\mathsf{P}^{n+1} := (\mu\mathsf{P}^n)\mathsf{P}$. Note that the total mass of $\mu\mathsf{P}$ is that of μ. Specializing to μ being a Dirac delta, we can also define the iterations of the transition kernel itself. Indeed, we define $p^0(x, \mathrm{d}y) := \delta_x(\mathrm{d}y)$ and, recursively for all $n \geqslant 0$,

$$p^{n+1}(x, A) := \int p^n(x, \mathrm{d}z)p(z, A) \tag{30.2}$$

for all $A \in \Sigma$. When S is countable, we write P^n for the iterations of P defined, again, recursively by

$$\mathsf{P}^{n+1}(x, y) := \sum_{z\in S} \mathsf{P}^n(x, z)\mathsf{P}(z, y) \tag{30.3}$$

with $\mathsf{P}^0(x, y) = \delta_{x,y}$. With these notations in place, we have:

Lemma 30.1 *Let $\{X_n\}_{n\geqslant 0}$ be a Markov chain with transition probability p and initial distribution μ. Then*

$$P^\mu(X_n \in B) = \mu\mathsf{P}^n(B) \tag{30.4}$$

holds for all $B \in \Sigma$ and all $n \geqslant 0$.

Proof. This follows from the calculation

$$\begin{aligned} P^\mu(X_{n+1} \in B) &= E\big(E(1_B(X_{n+1}) \,|\, \mathcal{F}_n)\big) \\ &= E\big(p(X_n, B)\big) = \int P(X_n \in \mathrm{d}x)p(x, B) \end{aligned} \tag{30.5}$$

by plugging $\mu\mathsf{P}^n$ for $P(X_n \in \cdot)$. □

The sequence of measures $\{\mu\mathsf{P}^n\}_{n\geqslant 0}$ thus coincides with the sequence of laws associated with a run of the Markov chain initiated from μ. In order to motivate the next definition, assume that we somehow know that the law $\mu\mathsf{P}^n$ of X_n under P^μ converges to a probability measure ν. Then also the law $\mu\mathsf{P}^{n+1}$ of X_{n+1} converges to ν and so it is reasonable to expect that $\nu\mathsf{P} = \nu$. This naturally leads to:

Definition 30.2 *A non-zero measure ν on (S,Σ) is said to be stationary if $\nu\mathsf{P} = \nu$ or, in full notation,*

$$\forall B \in \Sigma\colon \quad \nu(B) = \int \nu(\mathrm{d}x)p(x,B) \tag{30.6}$$

A stationary measure ν that obeys $\nu(S) = 1$ is called a stationary distribution.

The word "stationary" refers to the fact that, if the chain is started from a stationary distribution, then X_n has the same law for all $n \geqslant 0$. Note, however, that for a measure to be stationary we do not require that ν is a probability, or that it is even finite. We denote the set of stationary measures as

$$\mathscr{I} := \bigl\{\nu \in \mathcal{M}(S)\colon \nu\mathsf{P} = \nu,\, \nu(S) > 0\bigr\} \tag{30.7}$$

and the set of stationary distributions as

$$\mathscr{I}_1 := \bigl\{\nu \in \mathcal{M}(S)\colon \nu\mathsf{P} = \nu,\, \nu(S) = 1\bigr\} \tag{30.8}$$

where $\mathcal{M}(S)$ is the set of all measures on (S,Σ). In the literature one can also find the synonym "invariant" for "stationary."

In the special case of S countable, the condition (30.6) is rephrased as

$$\forall x \in S\colon \quad \nu(x) = \sum_{y\in S} \nu(y)\mathsf{P}(y,x) \tag{30.9}$$

where ν is now the probability mass function. (One can similarly rewrite (30.6) using the Radon-Nikodym derivative of ν with respect to a measure on S.)

30.2 Existence and uniqueness.

Once we accept the concept of stationary measures, a natural question is that of existence and uniqueness. Neither of these comes for granted in general, which makes the concept all the more interesting. We start with an observation that builds on a classical result in linear algebra: the *Perron-Frobenius theorem*.

Lemma 30.3 *Let S be finite and let P be a stochastic matrix indexed by S. Then there exists a probability measure ν on S such that $\nu\mathsf{P} = \nu$. In short, $\mathscr{I}_1 \neq \varnothing$.*

Proof. The fact that P is stochastic implies that the constant vector 1 is a right-eigenvector with unit eigenvalue; i.e., $\mathsf{P}1 = 1$. This means that $\det(1-\mathsf{P}) = 0$ and so P admits a left eigenvector h with unit eigenvalue; i.e., $h\mathsf{P} = h$ with $h \neq 0$ or, in "coordinates,"

$$\forall x \in S\colon \quad h(x) = \sum_{y\in S} h(y)\mathsf{P}(y,x) \tag{30.10}$$

This would do the job once if we knew h is non-negative. Although this may not be the case in general, we will instead show that $|h|$ is then such an eigenvector as well. Indeed, (30.10) along with $\mathsf{P}(y,x) \geqslant 0$ give

$$\forall x \in S\colon \quad \bigl|h(x)\bigr| \leqslant \sum_{y\in S} \bigl|h(y)\bigr|\mathsf{P}(y,x) \tag{30.11}$$

which by summing over $x \in S$ and using that $\mathsf{P}(y,\cdot)$ is a probability mass function shows

$$\sum_{x\in S}|h(x)| \leqslant \sum_{x\in S}\sum_{y\in S}|h(y)|\mathsf{P}(y,x) = \sum_{y\in S}|h(y)|\sum_{x\in S}\mathsf{P}(y,x) = \sum_{y\in S}|h(y)| \tag{30.12}$$

Since the left and right hand sides are equal, equality must thus hold in (30.11) and so $|h|\mathsf{P} = |h|$. Taking $\nu(x) := |h(x)|/\sum_{z\in S}|h(z)|$ gives $\nu \in \mathscr{I}_1$, as desired. □

We note that uniqueness fails already for finite-state Markov chains. For instance, take a chain on S with $|S| < \infty$ such that $\mathsf{P}(x,x) = 1$ for all $x \in S$. Then $\delta_x \in \mathscr{I}_1$ for all $x \in S$ and so there are multiple invariant distributions once S is larger than a singleton. A key obstruction here is clearly the lack of the following property:

Definition 30.4 *Consider a Markov chain in a countable (or finite) set S. Writing P for the transition kernel, we say that the chain is* irreducible *if*

$$\forall x,y \in S\,\exists n \geqslant 0\colon \quad \mathsf{P}^n(x,y) > 0 \tag{30.13}$$

As is easy to check, if $\nu \in \mathscr{I}$ and the Markov chain is irreducible, then $\nu(x) > 0$ for all $x \in S$. Using this observation we conclude:

Lemma 30.5 *Suppose that S is finite and P is a transition kernel such that the associated Markov chain is irreducible. Then $\mathscr{I}_1$ is a singleton.*

Proof. Suppose $\nu,\tilde\nu \in \mathscr{I}_1$. Using that S is finite, the fact that the chain is irreducible implies $\epsilon := \min_{x\in S}\nu(x)/\tilde\nu(x) > 0$. If $\tilde\nu \neq \nu$, then $\nu - \epsilon\tilde\nu$ defines a (non-trivial) invariant measure that vanishes at at least one point of S, contradicting irreducibility. Hence $\tilde\nu = \nu$ and, thanks to Lemma 30.3, $\mathscr{I}_1$ is a singleton. □

We will now generalize the existence part of the claim to compact state spaces. Given a measurable function f, let $\mathsf{P}f$ be the function

$$\mathsf{P}f(x) := \int p(x,\mathrm{d}y)f(y) \tag{30.14}$$

whenever the integral is meaningful. If S is a topological space, write $C_{\mathrm{b}}(S)$ for the set of bounded continuous functions on S. Note that, in general, $\mathsf{P}f$ is just measurable regardless of how regular f may be. The following concept boosts the regularity of $\mathsf{P}f$ in a way that is useful both for theory building and applications:

Definition 30.6 (Feller property) *Let p be a transition probability of a Markov chain on a topological space S endowed with the σ-algebra $\Sigma := \mathcal{B}(S)$ of its Borel sets. We say that p is* Feller *or has the* Feller property *if*

$$\forall f \in C_{\mathrm{b}}(S)\colon\; \mathsf{P}f \in C_{\mathrm{b}}(S) \tag{30.15}$$

With this in hand we get:

Theorem 30.7 (Krylov and Bogoliubov, 1937) *Suppose that S is a compact space and let p be a transition probability on $(S,\mathcal{B}(S))$ with the Feller property. Then $\mathscr{I}_1 \neq \emptyset$.*

Proof. Pick $x_0 \in S$ and, given $f \in C(S)$ and $n \geqslant 1$, define

$$\phi_n(f) := E^{x_0}\left(\frac{1}{n}\sum_{k=0}^{n-1} f(X_k)\right) = \frac{1}{n}\sum_{k=0}^{n-1} \mathsf{P}^k f(x_0) \tag{30.16}$$

Note that $|\phi_n(f)| \leqslant \|f\| := \sup_{x\in S}|f(x)|$ (which is finite thanks to the assumed compactness) and so $\{\phi_n\}_{n\geqslant 1}$ is an equicontinuous family of (continuous) linear functionals on $C(S)$. The compactness of S also implies that $C(S)$ is separable meaning that there exists a countable set $\mathcal{D} \subseteq C(S)$ whose closure is all of $C(S)$. Cantor's diagonal argument allows us to pick $n_k \to \infty$ such that $\{\phi_{n_k}(f)\}_{k\geqslant 1}$ converges for all $f \in \mathcal{D}$. Taking any $f \in C(S)$, for each $\epsilon > 0$ we can find $f_\epsilon \in \mathcal{D}$ such that $\|f - f_\epsilon\| < \epsilon$. But then the uniform bound on the norm of ϕ_n implies

$$\begin{aligned}
&\limsup_{k\to\infty} \phi_{n_k}(f) - \liminf_{k\to\infty} \phi_{n_k}(f) \\
&\qquad\leqslant \limsup_{k\to\infty} \phi_{n_k}(f_\epsilon) + \|f - f_\epsilon\| - \liminf_{k\to\infty} \phi_{n_k}(f_\epsilon) - \|f - f_\epsilon\| \\
&\qquad= 2\|f - f_\epsilon\| < 2\epsilon
\end{aligned} \tag{30.17}$$

As ϵ is arbitrary, this proves that

$$\phi(f) := \lim_{k\to\infty} \phi_{n_k}(f) \tag{30.18}$$

exists for all $f \in C(S)$.

The definition of ϕ_n gives

$$\phi_n(\mathsf{P}f) = \phi_n(f) + \frac{1}{n}\big[\mathsf{P}^n f(x_0) - f(x_0)\big] \tag{30.19}$$

The term on the right is bounded by $\frac{2}{n}\|f\|$ and so it converges to zero. Since the Feller property gives $\mathsf{P}f \in C(S)$, we thus get

$$\forall f \in C(S)\colon \quad \phi(\mathsf{P}f) = \phi(f) \tag{30.20}$$

As $|\phi(f)| \leqslant \|f\|$ and $\phi(1) = 1$, the Riesz Representation Theorem implies existence of a probability measure ν on $(S, \mathcal{B}(S))$ such that

$$\forall f \in C(S)\colon \quad \phi(f) = \int f \mathrm{d}\nu \tag{30.21}$$

Using the Fubini-Tonelli theorem, the identity (30.20) then translates into

$$\forall f \in C(S)\colon \qquad \int f \mathrm{d}(\nu\mathsf{P}) = \int f \mathrm{d}\nu \tag{30.22}$$

Standard approximation arguments now give equality $\nu = \nu\mathsf{P}$ on open sets and, by Dynkin's π/λ-theorem, on all Borel sets. Hence, $\nu \in \mathscr{I}_1$. □

The above proof gives some clear indications what may go wrong with the argument in non-compact case. First, "mass" can spread to "infinity" resulting in vanishing limit in (30.18). Perhaps more importantly, the technique will at best produce a stationary measure of finite mass and, as we shall see, there are Markov chains that admit no such measures yet they do have measures of infinite mass.

30.3 Examples and counterexamples.

We now move to treat some specific examples of stationary measures as well as counterexamples to their existence and uniqueness. We start with the Ehrenfest urn model:

Lemma 30.8 *Consider the Ehrenfest chain with state space* $S := \{0, 1, \dots, n\}$ *and let the transition matrix be as in* (29.18). *Then*

$$\nu(k) := \binom{n}{k} 2^{-n}, \qquad k = 0, \dots, n, \tag{30.23}$$

defines a stationary distribution.

Proof. We have to show that ν satisfies (30.9). First we note that

$$\sum_{k \in S} \nu(k) \mathsf{P}(k, l) = \nu(l-1) \mathsf{P}(l-1, l) + \nu(l+1) \mathsf{P}(l+1, l). \tag{30.24}$$

Then we calculate

$$\begin{aligned} \text{rhs of (30.24)} &= 2^{-n} \Big[\binom{n}{l-1} \frac{n-l+1}{n} + \binom{n}{l+1} \frac{l+1}{n} \Big] \\ &= 2^{-n} \binom{n}{l} \Big[\frac{l}{n-l+1} \frac{n-l+1}{n} + \frac{n-l}{l+1} \frac{l+1}{n} \Big]. \end{aligned} \tag{30.25}$$

The proof is finished by noting that, after a cancellation, the bracket equals one. □

Concerning the birth-death chain, we state the following:

Lemma 30.9 *Consider the birth-death chain on* $\mathbb{N}$ *characterized by sequences* $\{\alpha_k\}_{k \geqslant 0}$, $\{\beta_k\}_{k \geqslant 0}$ *and* $\{\gamma_k\}_{k \geqslant 0}$ *as in* (29.19). *Suppose that* $\beta_n > 0$ *for all* $n \geqslant 1$. *Then* $\nu(0) := 1$ *and*

$$\nu(n) := \prod_{k=1}^{n} \frac{\alpha_{k-1}}{\beta_k}, \qquad n \geqslant 1, \tag{30.26}$$

defines a stationary measure of the chain.

Proof. The condition (30.9) boils down to showing that

$$\nu(k-1)\alpha_{k-1} + \nu(k)\gamma_k + \nu(k+1)\beta_{k+1} = \nu(k) \tag{30.27}$$

for each $k \geqslant 0$ (with $\nu(-1) := 0$). Now check that $\nu(k-1)\alpha_{k-1} = \nu(k)\beta_k$ and that $\nu(k+1)\beta_{k+1} = \alpha_k \nu(k)$ and apply (29.20) to get the claim. □

The reader might wonder how does one come up with the expression (30.26) in the first place. This will be explained elegantly using the concept of reversibility later.

Next we observe:

Lemma 30.10 *Consider the card shuffling Markov chain defined in* (29.25). *Then* $\nu(\sigma) := \frac{1}{n!}$ *is an stationary distribution.*

Proof. This is checked directly from

$$\sum_{\sigma\in\mathcal{S}_n}\nu(\sigma)\mathsf{P}(\sigma,\tilde\sigma)=\frac{1}{n!}\sum_{\sigma\in\mathcal{S}_n}q(\tilde\sigma\circ\sigma^{-1})=\frac{1}{n!}\sum_{\sigma'\in\mathcal{S}_n}q(\sigma')=\frac{1}{n!}=\nu(\tilde\sigma) \tag{30.28}$$

where we changed variables to $\sigma' := \tilde\sigma\circ\sigma^{-1}$ and observed that the sum then still goes over all elements in $\mathcal{S}_n$. ☐

For the card shuffles that have a chance to mix the deck, i.e., those that are irreducible, the uniform distribution is the unique invariant distribution by Lemma 30.5. This is actually true for general random walks on groups, where the left-invariant (and thus "uniform") measure is always stationary whenever it exists. However, as the next lemma shows, other stationary measures may exist as well:

Lemma 30.11 (Biased simple random walk) *Let $S := \mathbb{Z}$ and, given $p\in(0,1)$, define P by*

$$\mathsf{P}(k,k+1)=p \quad\text{and}\quad \mathsf{P}(k,k-1)=1-p \tag{30.29}$$

For each $\alpha,\beta\geqslant 0$ set

$$\nu_{\alpha,\beta}(k):=\alpha+\beta\left(\frac{1-p}{p}\right)^k \tag{30.30}$$

Then

$$\mathscr{I}=\{\nu_{\alpha,\beta}\colon \alpha,\beta\geqslant 0\} \tag{30.31}$$

In particular, $\mathscr{I}$ has at least two distinct extreme points whenever $p\neq 1/2$.

Proof. The condition (30.9) reduces to

$$\nu(k+1)p+\nu(k-1)(1-p)=\nu(k)=\nu(k)p+\nu(k)(1-p) \tag{30.32}$$

which rewrites as

$$\nu(k+1)-\nu(k)=\frac{1-p}{p}\big[\nu(k)-\nu(k-1)\big] \tag{30.33}$$

This is clearly solved by constant ν but also by ν whose increments are such that

$$\nu(k+1)-\nu(k)=a\left(\frac{1-p}{p}\right)^k \tag{30.34}$$

for a a positive constant. This now readily yields (30.30). ☐

The example in Lemma 30.11 shows that non-uniqueness of the stationary measure (which is determined only up an overall multiple) may arise naturally in Markov chains on infinite state spaces, in spite of irreducibility being in force. We may also be in the situation when a small change to the transition probabilities of the chain changes the existence of stationary measures:

Lemma 30.12 *Consider the Renewal chain with transition kernel defined using the sequence $\{\alpha_k\}_{k\geqslant 0}$ as in (29.21–29.22). Then*

$$\mathscr{I}\neq\varnothing \quad\Leftrightarrow\quad \prod_{k\geqslant 0}(1-\alpha_k)=0 \tag{30.35}$$

Proof. The condition (30.9) translates into

$$\forall k \geqslant 1\colon \quad \nu(k-1)(1-\alpha_{k-1}) = \nu(k) \tag{30.36}$$

and

$$\sum_{k\geqslant 0} \alpha_k \nu(k) = \nu(0) \tag{30.37}$$

The first condition is resolved as

$$\forall k \geqslant 1\colon \quad \nu(k) = \nu(0) \prod_{j=0}^{k-1} (1-\alpha_j) \tag{30.38}$$

which by plugging into the second condition gives

$$\nu(0) = \nu(0)\left(\alpha_0 + \sum_{k\geqslant 1} \alpha_k \prod_{j=0}^{k-1}(1-\alpha_j)\right) = \nu(0)\left(1 - \prod_{j\geqslant 0}(1-\alpha_j)\right) \tag{30.39}$$

where the second equality follows by truncating the sum, proving the formula for this case and removing the trunction by a limit. The upshot of (30.39) is that $\nu(0)$ must vanish unless the infinite product vanishes, in which case we can set $\nu(0)$ to any non-negative value and satisfy both (30.36) and (30.37). □

The upshot of the above examples is that, for infinite-state Markov chains, the existence and uniqueness of stationary measures are quite problem dependent. We will return to the example of the Renewal chain when we discuss the connection between existence of stationary measures and recurrence of the chain.