1 Setting

A Markov chain on a state space S and with transition probabilities $K(\cdot\,,\cdot)$ is a sequence of random variables

$$X_0, X_1, X_2, \dots \tag{1}$$

with values from S satisfying

$$\begin{aligned}\mathbf{P}(X_{t+1} = a_{t+1} \,|\, X_0 = a_0, X_1 = a_1, \dots, X_t = a_t) &= \mathbf{P}(X_{t+1} = a_{t+1} \,|\, X_t = a_t) &(2)\\ &= K(a_t, a_{t+1}) &(3)\end{aligned}$$

for every $t \geq 0$ and $a_0, a_1, \dots, a_{t+1} \in S$. The first equality is the *Markov property*, and the fact that the transition probabilities $\mathbf{P}(X_{t+1} = b \,|\, X_t = a) = K(a, b)$ do not depend on time is referred to as *time-homogeneity*. The state space S is assumed to be finite or at most countable.

Given a probability distribution μ on S, we denote by μK the distribution after one step if the current state of Markov chain has probability distribution μ. That is, $\mu K(b) \triangleq \sum_{a\in S} \mu(a)K(a, b)$, which is compatible with the image of K as a matrix and μ as a row vector.

If $f : S \to \mathbb{R}$ is a function associating real numbers to each state in S and π a probability distribution on S, we write $\pi(f)$ or πf for the expected value of f with respect to the distribution π. Therefore, $\pi(f) = \sum_{a\in S} \pi(a)f(a)$. Similarly, we write Kf for the function on S defined by $Kf(a) \triangleq \sum_{b\in S} K(a, b)f(b)$. That is, $Kf(a)$ is the conditional expected value of $f(X_{t+1})$ given $X_t = a$.

For any state $a \in S$, we use the shorthands $\mathbf{P}_a$ and $\mathbf{E}_a$ for the conditional probability and expectation given $X_0 = a$. Extending this notation, we write $\mathbf{P}_\mu$ and $\mathbf{E}_\mu$ for the probability and expectation if the initial state state X_0 has distribution μ.

2 Equilibrium

A Markov chain

$$X_0, X_1, X_2, \dots \tag{4}$$

is in *(statistical) equilibrium* if its statistical properties do not change with a shift in time. More precisely, the chain $(X_t)_{t\geq 0}$ is in equilibrium at time n if

$$X_n \overset{\mathscr{D}}{\sim} X_{n+1} \overset{\mathscr{D}}{\sim} X_{n+2} \overset{\mathscr{D}}{\sim} \cdots . \tag{5}$$

An equilibrium (statistical) state is therefore associated with a *stationary* distribution, that is, a probability distribution π on the state space S satisfying $\pi K = \pi$. A Markov chain is in equilibrium at time n if and only if the distribution of the state at time n is stationary.

Three main questions regarding statistical equilibrium:

Q1: (Existence) Does a given Markov chain have any equilibrium state?

- Positive examples: Random walk on a cycle $\mathbb{Z}_n$ (uniform); Ehrenfest urn (binomial); Drunkard's walk ($\lambda\delta_{\text{home}} + (1-\lambda)\delta_{\text{pub}}$); Galton-Watson model ($\delta_0$)
- Negative example: Random walk on $\mathbb{Z}$

Q2: (Uniqueness) Is the equilibrium state unique? (Provided it exists.)

- Positive examples: Random walk on $\mathbb{Z}_n$; Ehrenfest urn; Card shuffling (hopefully!)
- Negative example: Drunkard's walk

Q3: (Asymptotics) If not in equilibrium, does the chain evolve towards equilibrium?

- Positive example: Random walk on $\mathbb{Z}_n$ (for n odd), Card shuffling (hopefully!)
- Negative example: Random walk on $\mathbb{Z}_n$ (for n even), Ehrenfest urn

2.1 Existence

Theorem 1. *Every finite-state Markov chain has at least one stationary distribution.*

Proof. Let μ be an arbitrary distribution on the state space S. For $n \geq 0$, set

$$\overline{\mu}_n \triangleq \frac{1}{n}(\mu + \mu K + \mu K^2 + \cdots \mu K^{n-1}) . \tag{6}$$

This is the uniform mixture of the distributions of $X_0, X_1, \dots, X_{n-1}$ if we start the chain with X_0 having distribution μ. Obviously, $\overline{\mu}$ itself is a probability distribution on S. Moreover, if n is large, $\overline{\mu}$ is "close to being stationary".

Argument. We have

$$\overline{\mu}_n K - \overline{\mu}_n = \frac{1}{n}(\mu K + \mu K^2 + \cdots \mu K^n) - \frac{1}{n}(\mu + \mu K + \cdots \mu K^{n-1}) = \frac{1}{n}(\mu K^n - \mu) . \tag{7}$$

If we measure the closeness of two distributions ν_1 and ν_2 by

$$\|\nu_2 - \nu_1\| \triangleq \sum_{a\in S} |\nu_2(a) - \nu_1(a)| , \tag{8}$$

we can write

$$\|\overline{\mu}_n K - \overline{\mu}_n\| = \frac{1}{n}\|\mu K^n - \mu\| = \frac{1}{n}\sum_a |\mu K^n(a) - \mu(a)| \tag{9}$$

$$\leq \frac{1}{n}\sum_a (\mu K^n(a) + \mu(a)) \tag{10}$$

$$= \frac{2}{n} \tag{11}$$

which goes to 0 as $n \to \infty$.

Now, consider the sequence

$$\overline{\mu}_1, \overline{\mu}_2, \ldots \tag{12}$$

whose elements asymptotically become closer and closer to being stationary. If we knew that the sequence converges, we could try out the limit distribution to see if it is stationary. Unfortunately, we still do not know that. However, the space of all probability distributions on S

$$\mathscr{P}(S) \triangleq \{\nu : S \to [0,1] \,:\, \sum_{a\in S} \nu(a) = 1\} \tag{13}$$

is compact. (*Warning*: not true if S infinite.)

Argument. The space $[0,1]^S$ of all functions $\alpha : S \to [0,1]$ is a product of finitely many compact spaces, and hence is compact (in which topology?). The set $\mathscr{P}(S)$ is closed in $[0,1]^S$, and therefore is compact (why?).

Therefore, there is a subsequence

$$\overline{\mu}_{n_1}, \overline{\mu}_{n_2}, \ldots \tag{14}$$

that converges. We claim that the limit distribution $\overline{\mu}$ is stationary.

Argument. Comparing $\overline{\mu}$ and $\overline{\mu}K$ with $\overline{\mu}_n$ and $\overline{\mu}_n K$ and using the triangular inequality, we can write

$$\|\overline{\mu}K - \overline{\mu}\| \leq \underbrace{\|\overline{\mu}K - \overline{\mu}_n K\|}_{\to 0} + \underbrace{\|\overline{\mu}_n K - \overline{\mu}\|}_{< \frac{2}{n} \to 0} + \underbrace{\|\overline{\mu}_n - \overline{\mu}\|}_{\to 0}\,, \tag{15}$$

which goes to 0 as $n \to \infty$. (That the first term goes to 0 is by the continuity of $\nu \mapsto \nu K$.)

□

Remark 2. In the above proof, we used the fact that the mapping $\nu \mapsto \nu K$ is continuous with respect to the topology of the metric $\|\nu_2 - \nu_1\| \triangleq \sum_{a\in S} |\nu_2(a) - \nu_1(a)|$. In fact, more is true: $\nu \mapsto \nu K$ is contractive.

Exercise 3. Verify that $\|\nu_2 K - \nu_1 K\| \leq \|\nu_2 - \nu_1\|$.

◇

Remark 4. There are other approaches to prove the existence of stationary distributions for finite-state Markov chains. For example, the classic approach is to note that the stochastic matrix K has eigenvalue 1 (why?). Therefore, $\alpha K = \alpha$ has a non-zero solution for α. One can then verify that if α is a solution, so is $|\alpha|$, which is obtained by taking the absolute value of the elements of α. Normalizing $|\alpha|$ gives a stationary distribution.

The idea of using the averages

$$\frac{1}{n}(\mu + \mu K + \mu K^2 + \cdots \mu K^{n-1})\,. \tag{16}$$

is however applicable to many other (more complicated) situations. For example, the same idea can be used to show that every continuous dynamical system on a compact space has an invariant measure, or that every Markov kernel on a compact state space has a stationary measure, provided it acts on the measures continuously (the *Feller* property). ◇