SF3953: Markov Chains and Processes	**Spring 2017**

Lecture 7: Uniqueness of π, Dynamical Systems

Lecturer: Jimmy Olsson — *April 20*

Goals of this lecture

- To find conditions guaranteeing that a Markov chain has a unique invariant probability measure.
- To start up a discussion on limit theorems for Markov chains by introducing some basic concepts related to dynamical systems.

Kac's formula and the existence of a unique π

As usual, let $(\mathsf{X}, \mathcal{X})$ be some measurable space. Recall Definition 4.2 terming a phi-irreducible chain *positive* if it admits an invariant probability measure.

Proposition 7.1 (Kac's formula)**.** *Let P be a transition kernel on $\mathsf{X} \times \mathcal{X}$ that admits an accessible small set $C \in \mathcal{X}$ such that*

$$\sup_{x \in C} \mathbb{E}_x [\sigma_C] < \infty. \tag{7.2}$$

Then the following holds true.

(i) The chain is positive recurrent[1]*, and the unique invariant probability measure π satisfies* Kac's formula*, i.e., for all $A \in \mathcal{X}$,*

$$\pi(A) = \int_C \pi(\mathrm{d}y)\, \mathbb{E}_y \left[\sum_{k=0}^{\sigma_C - 1} \mathbb{1}_A(X_k) \right] = \int_C \pi(\mathrm{d}y)\, \mathbb{E}_y \left[\sum_{k=1}^{\sigma_C} \mathbb{1}_A(X_k) \right]. \tag{7.3}$$

(ii) If $h \in \mathsf{F}_+(\mathcal{X})$ is such that

$$\sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} h(X_k) \right] < \infty, \tag{7.4}$$

then

$$\pi h = \int_C \pi(\mathrm{d}y)\, \mathbb{E}_y \left[\sum_{k=0}^{\sigma_C - 1} h(X_k) \right] = \int_C \pi(\mathrm{d}y)\, \mathbb{E}_y \left[\sum_{k=1}^{\sigma_C} h(X_k) \right] < \infty.$$

[1] By "positive recurrent" we mean simply recurrent and positive.

Proof. First, by Proposition 5.2, P is phi-irreducible. Moreover, by (7.2), σ_C is $\mathbb{P}_x$-a.s. finite for all $x \in C$; in other words, C is Harris recurrent. Thus, by Theorem 6.8, C is recurrent, and Theorem 6.1 hence implies that P is recurrent. By Theorem 6.13 there is an invariant measure μ to which all other invariant measures are proportional and such that $0 < \mu(C) < \infty$. (Indeed, recall that in the proof of Theorem 6.13, μ is constructed from an invariant distribution $\check{\mu}$ on the split space satisfying $\check{\mu}(\check{\boldsymbol{\alpha}}) = 1$, the existence of which is provided by Theorem 4.4. After that, the proof identifies μ such that $\mu^* = \check{\mu}$. Consequently, $1 = \varepsilon\mu(C)$.) Now, define

$$\mu_C : \mathcal{X} \ni A \mapsto \int_C \mu(\mathrm{d}y)\, \mathbb{E}_y \left[\sum_{k=1}^{\sigma_C} \mathbb{1}_A(X_k) \right].$$

By the Harris recurrence of C, $\sum_{k=1}^{\sigma_C} \mathbb{1}_C(X_k) = 1$ $\mathbb{P}_y$-a.s. for all $y \in C$, which implies that $\mu_C(C) = \mu(C)$.

Exercise 7.5. *Establish that $\mu_C = \mu$ by showing that*

1. *the measure μ_C is P-invariant.*
2. *for all $A \in \mathcal{X}$, $\mu(A) \geq \mu_C(A)$.*
3. *the previous implies that $\mu_C = \mu$.*

By Exercise 7.5, μ_C is invariant. In addition, for all $A \in \mathcal{X}$,

$$\int_C \mu(\mathrm{d}y)\, \mathbb{E}_y \left[\mathbb{1}_A(X_0)\right] = \mu(A \cap C) = \mu_C(A \cap C) = \int_C \mu(\mathrm{d}y)\, \mathbb{E}_y \left[\mathbb{1}_A(X_{\sigma_C})\right].$$

Thus, since for all $y \in C$,

$$\sum_{k=1}^{\sigma_C} \mathbb{1}_A(X_k) = \sum_{k=0}^{\sigma_C - 1} \mathbb{1}_A(X_k) - \mathbb{1}_A(X_0) + \mathbb{1}_A(X_{\sigma_C}), \quad \mathbb{P}_y\text{-a.s.},$$

it holds for all $A \in \mathcal{X}$,

$$\mu(A) = \mu_C(A) = \int_C \mu(\mathrm{d}y)\, \mathbb{E}_y \left[\sum_{k=1}^{\sigma_C} \mathbb{1}_A(X_k) \right] = \int_C \mu(\mathrm{d}y)\, \mathbb{E}_y \left[\sum_{k=0}^{\sigma_C - 1} \mathbb{1}_A(X_k) \right]. \tag{7.6}$$

Now,

$$\mu(\mathsf{X}) = \int_C \mu(\mathrm{d}y)\, \mathbb{E}_y \left[\sum_{k=0}^{\sigma_C - 1} \mathbb{1}_{\mathsf{X}}(X_k) \right] \leq \mu(C) \sup_{y \in C} \mathbb{E}_y \left[\sigma_C\right] < \infty,$$

implying that any invariant measure is finite. As a consequence, $\pi = \mu/\mu(\mathsf{X})$ is the unique invariant probability distribution. Indeed, let π' be another invariant probability distribution; then there exists $c \in \mathbb{R}_+$ such that $\pi' = c\mu$, which implies that $1 = \pi'(\mathsf{X}) = c\mu(\mathsf{X})$, or, equivalently, $c = 1/\mu(\mathsf{X})$. Thus, $\pi' = \pi$. In addition, by dividing (7.6) by $\mu(\mathsf{X})$, we conclude that (7.3) holds true. This completes the proof of (i).

We turn to (ii). Under the condition (7.4), by monotone convergence (cf. Exercise 4.3),

$$\pi h = \int_C \pi(\mathrm{d}y)\, \mathbb{E}_y \left[\sum_{k=0}^{\sigma_C - 1} h(X_k) \right] \leq \pi(C) \sup_{y \in \mathsf{X}} \mathbb{E}_y \left[\sum_{k=0}^{\sigma_C - 1} h(X_k) \right] < \infty,$$

showing (ii). □

Even though this condition appears to be hard to check directly, it is indeed implied by the following considerably more straightforward *drift condition*.

Theorem 7.7. *Assume that there exist $C \in \mathcal{X}$, $V \in \mathsf{F}(\mathcal{X})$ and $h \in \mathsf{F}(\mathcal{X})$ such that $1 \leq h \leq V$, and a constant $b \in \mathbb{R}_+$ such that*

$$PV \leq V - h + b\mathbb{1}_C. \tag{7.8}$$

Then for all $x \in \mathsf{X}$,

$$\mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} h(X_k) \right] \leq V(x) + b\mathbb{1}_C(x). \tag{7.9}$$

Thus, if C is an accessible small set and V is bounded on C, then the chain is positive recurrent and $\pi h < \infty$.

Proof. For all $n \in \mathbb{N}^*$, set

$$M_n = \left(V(X_n) + \sum_{k=0}^{n-1} h(X_k) \right) \mathbb{1}_{\{\sigma_C \geq n\}}.$$

We show that $\{M_n : n \in \mathbb{N}^*\}$ is a super-martingale. Indeed, pick arbitrarily $x \in \mathsf{X}$ and write, using that $\{\sigma_C \geq n+1\} = \{\sigma_C \leq n\}^\complement \in \mathcal{F}_n$, $\mathbb{P}_x$-a.s.,

$$\begin{aligned}
\mathbb{E}_x \left[M_{n+1} \mid \mathcal{F}_n \right] &= \left(PV(X_n) + \sum_{k=0}^{n} h(X_k) \right) \mathbb{1}_{\{\sigma_C \geq n+1\}} \\
&\leq \left(V(X_n) - h(X_n) + b\mathbb{1}_C(X_n) + \sum_{k=0}^{n} h(X_k) \right) \mathbb{1}_{\{\sigma_C \geq n+1\}} \\
&= \left(V(X_n) + \sum_{k=0}^{n-1} h(X_k) \right) \mathbb{1}_{\{\sigma_C \geq n+1\}} \\
&\leq M_n,
\end{aligned}$$

as $\mathbb{1}_C(X_n)\mathbb{1}_{\{\sigma_C \geq n+1\}} = 0$. For any $n \in \mathbb{N}^*$, $\sigma_C \wedge n$ is a bounded stopping time, and by applying Doob's optional stopping theorem we obtain, using again the drift condition (7.8),

$$\mathbb{E}_x [M_{\sigma_C \wedge n}] \leq \mathbb{E}_x [M_1] = PV(x) + h(x) \leq V(x) + b\mathbb{1}_C(x). \tag{7.10}$$

This yields

$$\mathbb{E}_x \left[\sum_{k=0}^{\sigma_C \wedge n - 1} h(X_k) \right] \leq V(x) + b\mathbb{1}_C(x)$$

which implies (7.9) by monotone convergence. If V is bounded on C, using the bound (7.9) with $h = 1$ implies that $\sup_{x \in C} \mathbb{E}_x[\sigma_C] < \infty$, which means that the condition (7.2) is satisfied. Consequently, the chain is positive by Proposition 7.1. In addition, if V is bounded on C, (7.9) implies (7.4) and finiteness of πh follows from Proposition 7.1. □

The drift condition (7.8) is indeed weak and may be checked for a large set of models.

Exercise 7.11. *The* first-order auto regressive *(AR(1))* process *on $\mathbb{R}$ is defined iteratively by*

$$X_{k+1} = \phi X_k + \sigma \varepsilon_{k+1}, \quad k \in \mathbb{N},$$

where ϕ and σ are constants such that $|\phi| < 1$ and $\sigma > 0$ and $\{\varepsilon_k : k \in \mathbb{N}^*\}$ is a sequence of mutually independent variables with common density Γ with respect to Lebesgue measure ν. For simplicity, assume that $\Gamma(x) > 0$ for all $x \in \mathbb{R}$.

(a) Show that all compact sets are small for this model.

(b) Check that this chain satisfies the drift condition (7.8) for some suitable choices of V and h.

Dynamical systems

We now direct focus towards limit theorems for positive recurrent chains. During the coming lecture, our aim will be to establish, for any π-integrable function $h \in \mathsf{F}(\mathcal{X})$, (a significantly more general version of the) the *law of large numbers*

$$\lim_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} h(X_k) = \pi h, \quad \mathbb{P}_x\text{-a.s.}$$

(where $x \in \mathsf{X}$ is arbitrary), as well as, given that $h \in L^2(\pi)$, the *central limit theorem*

$$\sqrt{n} \left(\frac{1}{n} \sum_{k=0}^{n-1} h(X_k) - \pi h \right) \overset{\mathbb{P}_\pi}{\Longrightarrow} \sigma Z,$$

where Z is standard normally distributed and $\sigma^2 > 0$ is an asymptotic variance. The derivation of the latter result will be driven by the so-called *Poisson equation*.

The approach that we take is based on the theory of *dynamical systems*, and in the following we introduce some useful definitions and properties related to the same.

Definition 7.12 (dynamical system)**.** *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space.*

- *A measurable map T from $(\Omega, \mathcal{A})$ to $(\Omega, \mathcal{A})$ is a* measure-preserving transformation *if for all $A \in \mathcal{A}$, $\mathbb{P}(T^{-1}(A)) = \mathbb{P} \circ T^{-1}(A) = \mathbb{P}(A)$. The probability $\mathbb{P}$ is then said to be* invariant *under the transformation T and $(\Omega, \mathcal{A}, \mathbb{P}, T)$ is called a* dynamical system*.*
- *The application T is said to be an* invertible measure-preserving transformation *if it is measure-preserving, invertible, and its inverse T^{-1} is measurable.*

Exercise 7.13. *Show that*

(a) if T is an invertible measure-preserving transformation, then T^{-1} is also measure-preserving.

(b) if T is measure-preserving, then T^n is measure-preserving for all $n \in \mathbb{N}^*$.

From previous lectures we recall the canonical space $(\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}})$, the coordinate process $\{X_k : k \in \mathbb{N}\}$, and the shift-operator $\theta : \mathsf{X}^{\mathbb{N}} \ni (\omega_0, \omega_1, \dots) \mapsto (\omega_1, \omega_2, \dots) \in \mathsf{X}^{\mathbb{N}}$.

Lemma 7.14. *A probability measure $\mathbb{P}$ on $(\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}})$ is invariant for θ if and only if the coordinate process is stationary under $\mathbb{P}$.*

Exercise 7.15. *Prove Lemma 7.14.*

Definition 7.16 (invariant variables and events)**.** *Let $T : \Omega \to \Omega$ be measurable.*

- *An $\bar{\mathbb{R}}$-valued random variable Y on $(\Omega, \mathcal{A})$ is invariant for T if $Y \circ T = Y$.*
- *A set $A \in \mathcal{A}$ is invariant for T if $A = T^{-1}(A)$ or, equivalently, if the indicator function $\mathbb{1}_A$ is invariant for T.*

Proposition 7.17. *Let $T : \Omega \to \Omega$ be measurable.*

(i) The collection $\mathcal{J}$ of invariant sets for T is a sub-σ-field of $\mathcal{A}$.

(ii) Let $(\mathsf{E}, \mathcal{E})$ be a measurable space such that singletons are measurable. Let $Y : (\Omega, \mathcal{A}) \to (\mathsf{E}, \mathcal{E})$ be a measurable mapping. Then, Y is invariant for T if and only if Y is $\mathcal{J}$-measurable.

Proof.

Exercise 7.18. *Establish the first claim (i).*

We prove (ii). If Y is invariant for T, then for all $B \in \mathcal{E}$,

$$T^{-1}(Y^{-1}(B)) = (T \circ Y)^{-1}(B) = Y^{-1}(B).$$

Thus, $Y^{-1}(B) \in \mathcal{J}$, which means that Y is $\mathcal{J}$-measurable.

Conversely, assume that Y is $\mathcal{J}$-measurable. Then for all $B \in \mathcal{E}$,

$$Y^{-1}(B) = T^{-1}(Y^{-1}(B)) = (Y \circ T)^{-1}(B).$$

Now, since singletons are measurable, we may use the previous identity with $B = \{x\}$ for any $x \in \mathsf{X}$. Thus, if $Y(\omega) = x$, we may write

$$\omega \in Y^{-1}(\{x\}) = (Y \circ T)^{-1}(\{x\}),$$

which implies that $Y \circ T(\omega) = x$. Thus, $Y = Y \circ T$, which shows that Y is invariant. □

Remark 7.19. *Consider again the coordinate process $\{X_k : k \in \mathbb{N}\}$ on $(\Omega, \mathcal{F}) = (\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes\mathbb{N}})$. In this setting, the values of a random variable that is invariant for θ are preserved when outcomes are shifted one or several steps. Thus, we may expect such variables, and hence the invariant σ-field, to be related to asymptotics in some sense. Indeed, in this case, $\mathcal{J}$ is contained in the* tail σ-field, *i.e. $\mathcal{J} \subset \cap_{k=0}^{\infty} \sigma(X_\ell : \ell \geq k)$. Moreover, for all $h \in \mathsf{F}(\mathcal{B}(\bar{\mathbb{R}}))$, $\limsup_{n\to\infty} h(X_n)$, $\liminf_{n\to\infty} h(X_n)$, $\limsup_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} h(X_k)$, and $\liminf_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} h(X_k)$ are all invariant.*

Proof. For each $k \in \mathbb{N}$, set $\mathcal{G}_k = \sigma(X_\ell : \ell \geq k)$ and let $\mathcal{E}_k$ be the finite rectangles generating $\mathcal{G}_k$. If $B \in \mathcal{E}_k$, then $\theta^{-1}(B) \in \mathcal{E}_{k+1}$. Thus, $\theta : (\Omega, \mathcal{G}_{k+1}) \to (\Omega, \mathcal{G}_k)$ is measurable. Now, let $A \in \mathcal{F} = \mathcal{G}_0$ be invariant for θ. If also $A \in \mathcal{G}_k$, then $A = \theta^{-1}(A) \in \mathcal{G}_{k+1}$. Consequently, by induction, $A \in \cap_{k=0}^{\infty} \mathcal{G}_k$, which was to be shown.

To prove the remaining statements, let $\omega \in \Omega$ be fixed. In addition, let $L \subset \bar{\mathbb{R}}$ be the set of points $x \in \bar{\mathbb{R}}$ for which there is a subsequence of $\{h(X_n) : n \in \mathbb{N}\}$ converging to x. Then, $\limsup_{n\to\infty} h(X_n) = \sup L$. On the other hand, the converging subsequences of $\{h(X_n) \circ \theta : n \in \mathbb{N}\}$ are exactly those of $\{h(X_n) : n \in \mathbb{N}\}$ (just shifted one step). Thus, $\limsup_{n\to\infty} h(X_n) \circ \theta = \sup L$, which shows that $\limsup_{n\to\infty} h(X_n)$ is invariant. The same argument applies to the rest of the quantities. □