| **SF3953: Markov Chains and Processes** | **Spring 2017** |
|---|---|
| Lecture 8: Limit theorems for Markov chains | |
| *Lecturer: Jimmy Olsson* | *April 21* |

## Goals of this lecture

- To establish a Birkhoff theorem for positive recurrent Markov chains.
- To establish a central limit theorem for such chains under the assumption that the Poisson equation has a solution.

## Ergodic dynamical systems

We will make use of the following result, which is fundamental in ergodic theory.

**Theorem 8.1** (Birkhoff's ergodic theorem)**.** *Let* $(\Omega, \mathcal{A}, \mathbb{P}, T)$ *be a dynamical system and* $Y$ *a random variable such that* $\mathbb{E}[|Y|] < \infty$*. Then,* $\mathbb{P}$*-a.s.,*

$$\lim_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} Y \circ T^k = \mathbb{E}[Y \mid \mathcal{J}]. \tag{8.2}$$

*Moreover, the convergence holds in* $L^1(\Omega, \mathcal{A}, \mathbb{P})$*.*

**Definition 8.3.** *A dynamical system* $(\Omega, \mathcal{A}, \mathbb{P}, T)$ *is* ergodic *if the invariant* $\sigma$*-field* $\mathcal{J}$ *is trivial for* $\mathbb{P}$*, i.e. for all* $A \in \mathcal{J}$*,* $\mathbb{P}(A) \in \{0, 1\}$*.*

**Exercise 8.4.** *Let* $(\Omega, \mathcal{A}, \mathbb{P}, T)$ *be an ergodic dynamical system and* $Y$ *a random variable for which* $\mathbb{E}[Y]$ *is well-defined. Then, show that the limit in* (8.2) *is equal to* $\mathbb{E}[Y]$*.*

## Ergodic Markov chains

We now cast the problem of deriving a *Birkhoff theorem for Markov chains* into the framework of ergodic dynamical systems. Thus, in following we consider a Markov kernel $P$ on some measurable space $(\mathsf{X}, \mathcal{X})$ and the coordinate process $\{X_k : k \in \mathbb{N}\}$ on the canonical space $(\Omega, \mathcal{F}) = (\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}})$, endowed with the family $\{\mathbb{P}_\mu : \mu \in \mathsf{M}_1(\mathcal{X})\}$ of probability measures under which the coordinate process is a Markov chain with kernel $P$ and initial distribution $\mu$.

**Definition 8.5.** *A probability measure $\pi \in \mathsf{M}_1(\mathcal{X})$ is $P$-*ergodic *if it is invariant with respect to $P$ and if the dynamical system $(\Omega, \mathcal{F}, \mathbb{P}_\pi, \theta)$ is ergodic.*

**Theorem 8.6** (a Birkhoff theorem for Markov chains)**.** *Let $P$ be a Markov kernel on $\mathsf{X} \times \mathcal{X}$ and let $\pi$ be a $P$-ergodic probability measure. Let $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P}_\pi)$. Then for $\pi$-almost all $x \in \mathsf{X}$,*

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \theta^k = \mathbb{E}_\pi[Y], \quad \mathbb{P}_x\text{-a.s.}$$

*Proof.* Since $\pi$ is $P$-ergodic, the invariant $\sigma$-field $\mathcal{J}$ is trivial. Thus, by Theorem 8.1 and Exercise 8.4,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \theta^k = \mathbb{E}_\pi[Y], \quad \mathbb{P}_\pi\text{-a.s.} \tag{8.7}$$

Let $A \in \mathcal{F}$ be the set of $\omega \in \Omega$ for which the convergence (8.7) holds. Then $\mathbb{P}_\pi(A) = \int \mathbb{P}_x(A)\, \pi(\mathrm{d}x) = 1$, which implies, since $\mathbb{P}_x(A) \le 1$, that $\mathbb{P}_x(A) = 1$ for $\pi$-almost all $x \in \mathsf{X}$. This completes the proof. □

We now seek conditions guaranteeing $P$-ergodicity, and it turns out that uniqueness of $\pi$ is such a condition.

The following proposition shows that if a Markov kernel $P$ has an invariant distribution and if the invariant $\sigma$-field is not trivial for $\mathbb{P}_\pi$, then one may construct two mutually singular invariant distributions.

**Proposition 8.8.** *Let $\pi$ be a $P$-invariant probability measure. Assume that there exists $A \in \mathcal{J}$ such that $\alpha = \mathbb{P}_\pi(A) \notin \{0, 1\}$. Then, there exists $B \in \mathcal{X}$ such that $\pi(B) = \alpha$ and the probability measures*

$$\pi_B(\cdot) = \alpha^{-1}\pi(B \cap \cdot) \quad \text{and} \quad \pi_{B^\complement}(\cdot) = (1-\alpha)^{-1}\pi(B^\complement \cap \cdot)$$

*are invariant for $P$ and*

$$\mathbb{P}_{\pi_B}\Big(X_k \in B \text{ for all } k \in \mathbb{N}\Big) = \mathbb{P}_{\pi_{B^\complement}}\Big(X_k \in B^\complement \text{ for all } k \in \mathbb{N}\Big) = 1.$$

This proposition has an important consequence: if a Markov kernel $P$ has a unique invariant distribution $\pi$, then the invariant $\sigma$-field is necessarily trivial for $\mathbb{P}_\pi$. The dynamical system $(\Omega, \mathcal{F}, \mathbb{P}_\pi, \theta)$ is hence ergodic.

**Corollary 8.9.** *If the Markov kernel $P$ admits a unique invariant probability $\pi$, then $\pi$ is $P$-ergodic.*

We preface the proof of Proposition 8.8 by the following interesting lemma.

**Lemma 8.10.** *Let $\pi \in \mathsf{M}_1(\mathcal{X})$ be invariant with respect to $P$ and $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P}_\pi)$ be invariant for $\theta$. Then $\mathbb{E}_x[|Y|] < \infty$ $\pi$-a.s., the function $\mathsf{X} \ni x \mapsto \mathbb{E}_x[Y]$ is $\pi$-integrable, and $Y = \mathbb{E}_{X_0}[Y]$, $\mathbb{P}_\pi$-a.s.*

*Proof.* Since $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P}_\pi)$, $\mathbb{E}_\pi[|Y|] = \int \mathbb{E}_x[|Y|]\, \pi(\mathrm{d}x) < \infty$, which implies that $\mathbb{E}_x[|Y|] < \infty$ $\pi$-a.s. and that the function $\mathsf{X} \ni x \mapsto \mathbb{E}_x[Y]$ is $\pi$-integrable. Finally, by the Markov property and the invariance of $Y$, $\mathbb{P}_\pi$-a.s.,

$$\mathbb{E}_{X_k}[Y] = \mathbb{E}_\pi[Y \circ \theta_k \mid \mathcal{F}_k] = \mathbb{E}_\pi[Y \mid \mathcal{F}_k].$$

Thus, by Theorem 8.18,

$$\lim_{k\to\infty} \mathbb{E}_{X_k}[Y] = \mathbb{E}_\pi[Y \mid \mathcal{F}] = Y, \tag{8.11}$$

$\mathbb{P}_\pi$-a.s. and in $L^1(\Omega, \mathcal{F}, \mathbb{P}_\pi)$. Since for all random variables $Z$ on $(\Omega, \mathcal{F})$, $\mathbb{E}_\pi[Z \circ \theta_k] = \mathbb{E}_\pi[Z]$ (can you prove it?),

$$\begin{aligned}\mathbb{E}_\pi[|Y - \mathbb{E}_{X_0}[Y]|] = \mathbb{E}_\pi[|Y - \mathbb{E}_{X_0}[Y]| \circ \theta_k] = \mathbb{E}_\pi[|Y \circ \theta_k - \mathbb{E}_{X_k}[Y]|] \\ = \mathbb{E}_\pi[|Y - \mathbb{E}_{X_k}[Y]|].\end{aligned}$$

Now, letting $k$ tend to infinity and applying (8.11) yields

$$\mathbb{E}_\pi[|Y - \mathbb{E}_{X_0}[Y]|] = \lim_{k\to\infty} \mathbb{E}_\pi[|Y - \mathbb{E}_{X_k}[Y]|] = 0,$$

which shows that $Y = \mathbb{E}_{X_0}[Y]$, $\mathbb{P}_\pi$-a.s. □

*Proof of Proposition 8.8.* Since $A \in \mathcal{J}$, the random variable $\mathbb{1}_A$ is invariant. Thus, by Lemma 8.10, $\mathbb{P}_\pi$-a.s.,

$$\mathbb{1}_A = \mathbb{E}_{X_0}[\mathbb{1}_A] = \mathbb{P}_{X_0}(A). \tag{8.12}$$

Let $B = \{x \in \mathsf{X} : \mathbb{P}_x(A) = 1\}$; then, the following identities hold $\mathbb{P}_\pi$-a.s. If $\mathbb{1}_B(X_0) = 1$, then, by (8.12), $\mathbb{1}_A = 1$; thus, $\mathbb{1}_B(X_0) \le \mathbb{1}_A$. On the contrary, if $\mathbb{1}_A = 1$, then, again by (8.12), $\mathbb{P}_{X_0}(A) = 1$, which implies that $\mathbb{1}_B(X_0) = 1$. Consequently, $\mathbb{1}_A \le \mathbb{1}_B(X_0)$, implying that $\mathbb{1}_A = \mathbb{1}_B(X_0)$, $\mathbb{P}_\pi$-a.s. As $\mathbb{1}_A$ is invariant, this implies that

$$\mathbb{1}_A = \mathbb{1}_B(X_0) = \mathbb{1}_B(X_1) = \mathbb{1}_B(X_2) = \ldots = \mathbb{1}_B(X_k) = \ldots = \prod_{k=0}^{\infty} \mathbb{1}_B(X_k), \quad \mathbb{P}_\pi\text{-a.s.}$$

These equalities have the following implications. First, for all $D \in \mathcal{X}$,

$$\begin{aligned}\pi_B P(D) = \mathbb{P}_{\pi_B}(X_1 \in D) = \alpha^{-1}\mathbb{P}_\pi(X_0 \in B, X_1 \in D) = \alpha^{-1}\mathbb{P}_\pi(X_1 \in B, X_1 \in D) \\ = \alpha^{-1}\mathbb{P}_\pi(X_0 \in B, X_0 \in D) = \pi_B(D),\end{aligned}$$

which shows that $\pi_B$ is invariant with respect to $P$. Second,

$$1 = \mathbb{P}_{\pi_B}(X_0 \in B) = \alpha^{-1}\mathbb{P}_\pi(X_0 \in B) = \alpha^{-1}\mathbb{P}_\pi(X_k \in B \text{ for all } k \in \mathbb{N}) \\ = \mathbb{P}_{\pi_B}(X_k \in B \text{ for all } k \in \mathbb{N}).$$

Exactly the same arguments apply to $\pi_{B^\complement}$, and the proof is complete.

□

## A central limit theorem for additive state functionals

As before, let $\{X_k : k \in \mathbb{N}\}$ be the canonical Markov chain with Markov kernel $P$ on $\mathsf{X} \times \mathcal{X}$ and let $\pi \in \mathsf{M}_1(\mathcal{X})$ be invariant with respect to $P$. The goal of the following section is to derive a *central limit theorem* (CLT) for Markov chain path averages of increasing length. The main tool will be the following.

**Definition 8.13** (Poisson equation)**.** *Assume that $P$ admits a unique invariant distribution $\pi$. For $h \in L^1(\pi)$, the equation*

$$\hat{h} - P\hat{h} = h - \pi h \tag{8.14}$$

*is called the* Poisson equation *associated to the function $h$. Any $\hat{h} \in \mathsf{F}(\mathcal{X})$ satisfying $P|\hat{h}|(x) < \infty$ for all $x \in \mathsf{X}$ and such that (8.14) holds true is called a* solution of the Poisson equation associated to *$h$.*

**Theorem 8.15.** *Let $P$ be a Markov kernel that admits a unique invariant probability measure $\pi$. Let $h \in L^2(\pi)$ and assume that there exists a solution $\hat{h} \in L^2(\pi)$ of the Poisson equation associated to $h$. Then*

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

*where $Z$ is standard normally distributed and*

$$\sigma_\pi^2(\hat{h}) = \mathbb{E}_\pi\left[\{\hat{h}(X_1) - P\hat{h}(X_0)\}^2\right]. \tag{8.16}$$

We preface the proof of Theorem 8.15 by the following lemma, which is a straightforward consequence of the CLT for stationary martingale difference sequences; see Theorem 8.19. For the invariant distribution $\pi$ in the statement of the theorem, denote $\bar{\pi} = \pi \otimes P$.

**Lemma 8.17.** *Let $G \in L^2(\bar{\pi})$ and assume that $\int G(x,y)\,P(x,\mathrm{d}y) = 0$, for all $x \in \mathsf{X}$. Then*

$$n^{-1/2}\sum_{k=0}^{n-1} G(X_k, X_{k+1}) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \bar{\pi}(G^2)Z,$$

*where $Z$ is standard normally distributed.*

*Proof.* By the Markov property and by assumption, for all $k \in \mathbb{N}$, $\mathbb{P}_\pi$-a.s.,

$$\mathbb{E}_\pi[G(X_k, X_{k+1}) \mid \mathcal{F}_k] = \mathbb{E}_{X_k}[G(X_0, X_1)] = 0,$$

which means, as $G \in L^2(\bar{\pi})$ by assumption, that $\{G(X_k, X_{k+1}) : k \in \mathbb{N}\}$ is a square integrable martingale difference sequence under $\mathbb{P}_\pi$. Thus, Theorem 8.19 applies if the conditions (8.20) and (8.21) hold true. On the other hand, by the Markov property and Theorem 8.6, $\mathbb{P}_\pi$-a.s.,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{E}_\pi\left[G^2(X_k, X_{k+1}) \mid \mathcal{F}_k\right] &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{E}_{X_k}\left[G^2(X_0, X_1)\right] \\ &= \int \mathbb{E}_x\left[G^2(X_0, X_1)\right]\, \pi(\mathrm{d}x) = \bar{\pi}(G^2),\end{aligned}$$

where the convergence holds $\mathbb{P}_\pi$-a.s. and hence in $\mathbb{P}_\pi$-probability. This shows (8.20). In addition, pick $\delta > 0$; then for all $\varepsilon > 0$, by Markov's inequality and stationarity,

$$\begin{aligned}&\mathbb{P}_\pi\left(\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{E}_\pi\left[G^2(X_k, X_{k+1})\mathbb{1}_{\{G^2(X_k,X_{k+1})\geq \delta\sqrt{n}\}} \mid \mathcal{F}_k\right] \geq \varepsilon\right) \\ &\qquad \leq \frac{1}{n\varepsilon}\sum_{k=0}^{n-1} \mathbb{E}_\pi\left[\mathbb{E}_{X_k}\left[G^2(X_0, X_1)\mathbb{1}_{\{G^2(X_0,X_1)\geq \delta\sqrt{n}\}}\right]\right] \\ &\qquad \leq \frac{1}{\varepsilon}\mathbb{E}_\pi\left[G^2(X_0, X_1)\mathbb{1}_{\{G^2(X_0,X_1)\geq \delta\sqrt{n}\}}\right],\end{aligned}$$

where the right hand side tends to zero as $n$ tends to infinity by monotone convergence and the fact that $G \in L^2(\bar{\pi})$ by assumption. This shows (8.21) and hence completes the proof. □

*Proof of Theorem 8.15.* Without loss of generality, assume that $\pi h = 0$. Then, by the Poisson equation, we have the decomposition

$$\sum_{k=0}^{n-1} h(X_k) = M_n + \hat{h}(X_0) - \hat{h}(X_n),$$

where

$$M_n = \sum_{k=0}^{n-1} \left(\hat{h}(X_{k+1}) - P\hat{h}(X_k)\right)$$

and $G(x, y) = \hat{h}(y) - P\hat{h}(x)$, $(x, y) \in \mathsf{X}^2$, satisfies the assumptions of Lemma 8.17; indeed, in that case, for all $x \in \mathsf{X}$,

$$\mathbb{E}_x\left[G(X_0, X_1)\right] = \mathbb{E}_x\left[\hat{h}(X_1)\right] - P\hat{h}(x) = 0.$$

As a consequence, by Lemma 8.17,

$$n^{-1/2} M_n \overset{\mathbb{P}_\pi}{\Longrightarrow} \sigma_\pi(\bar{h}) Z,$$

where $\sigma_\pi(\hat{h})$ is given by (8.16). Moreover, since by stationarity,

$$\mathbb{E}_\pi \left[ |\hat{h}(X_0) - \hat{h}(X_n)| \right] \leq 2\pi(|\hat{h}|),$$

implying that

$$n^{-1/2} \left( \hat{h}(X_0) - \hat{h}(X_n) \right) \overset{\mathbb{P}_\pi}{\Longrightarrow} 0,$$

the result follows.

□

## A Martingales

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

**Theorem 8.18.** *Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\{\mathcal{F}_k : k \in \mathbb{N}\}$ be a filtration of $\mathcal{F}$. Then the sequence $\{\mathbb{E}[X \mid \mathcal{F}_k] : k \in \mathbb{N}\}$ converges $\mathbb{P}$-a.s. and in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathbb{E}[X \mid \mathcal{F}_\infty]$, where $\mathcal{F}_\infty = \sigma(\cup_{k=0}^{\infty} \mathcal{F}_k)$.*

**Theorem 8.19.** *Let $\{(Z_k, \mathcal{F}_k) : k \in \mathbb{N}\}$ be square integrable martingale difference sequence. Assume that there exists $\sigma > 0$ such that as $n$ tends to infinity,*

$$n^{-1} \sum_{k=0}^{n-1} \mathbb{E}\left[ Z_{k+1}^2 \mid \mathcal{F}_k \right] \xrightarrow{\mathbb{P}} \sigma^2, \tag{8.20}$$

$$n^{-1} \sum_{k=0}^{n-1} \mathbb{E}\left[ Z_{k+1}^2 \mathbb{1}_{\{|Z_{k+1}| > \delta\sqrt{n}\}} \mid \mathcal{F}_k \right] \xrightarrow{\mathbb{P}} 0, \tag{8.21}$$

*for all $\delta > 0$. Then*

$$n^{-1/2} \sum_{k=0}^{n-1} Z_k \overset{\mathbb{P}}{\Longrightarrow} \sigma V,$$

*as $n$ tends to infinity, where $V$ is standard normally distributed.*