

Theorem

(i) Assume X is Markov(λ, P), irreducible,

$$\forall x \in E, \frac{\sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}}}{n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\lambda\text{-a.s.}} \frac{1}{\mathbb{E}_x[T_x^+]}$$

(ii) Assume in addition that X is recurrent and that ν is a nondegenerate invariant measure for the chain. If $f : E \rightarrow \mathbb{R}, g : E \rightarrow \mathbb{R}_+^*$ are such that $\sum_{x \in E} |f(x)|\nu(x) < \infty, \sum_{x \in E} g(x)\nu(x) < \infty$, then

$$\frac{\sum_{k=0}^{n-1} f(X_k)}{\sum_{k=0}^{n-1} g(X_k)} \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\lambda\text{-a.s.}} \frac{\sum_{x \in E} f(x)\nu(x)}{\sum_{x \in E} g(x)\nu(x)}$$

Corollary

Assume X is Markov(λ, P), irreducible, positive recurrent, with invariant probability π , and assume $f : E \rightarrow \mathbb{R}$ is such that $\sum_{x \in E} |f(x)|\pi(x) < \infty$. Then

$$\frac{\sum_{k=0}^{n-1} f(X_k)}{n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\lambda\text{-a.s.}} \sum_{x \in E} f(x)\pi(x).$$

Proof of the corollary : Simply apply (ii) of the theorem with $g \equiv 1$: we have indeed $\sum_{k=0}^{n-1} g(X_k) = n$, while $\sum_{x \in E} g(x)\pi(x) = 1$ because π is a probability.

Preliminary considerations

When the chain is positive recurrent, (ii) implies (i) : simply take $g \equiv 1$, $f = \mathbb{1}_{\{x\}}$.

When the chain is transient, $\mathbb{E}_x[T_x^+] = \infty$ and \mathbb{P}_λ -a.s., $\mathcal{V}_x < \infty$, thus

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}} \leq \frac{1}{n} \mathcal{V}_x \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\lambda\text{-a.s.}} 0 = \frac{1}{\mathbb{E}_x[T_x^+]}$$

As for the null recurrent case, we shall prove (i) and (ii) in the same argument, later on.

Proof of ergodic theorem (ii)

First observe that if we can prove the assertion under \mathbb{P}_x for any $x \in E$, the general result follows for $\lambda = \sum_{x \in E} \lambda(x) \delta_x$. Also, even if it means writing $f = f^+ - f^-$, we may assume $f : E \rightarrow \mathbb{R}_+$. From now on, we therefore fix some $x \in E$ and work under \mathbb{P}_x , and assume f takes nonnegative values.

Because the chain is irreducible recurrent, by 1.3, the invariant measure ν of the statement has to be a multiple of the measure ν_x defined in that section : there exists $\alpha > 0$ such that $\nu = \alpha \nu_x$ (in fact since $\nu_x(x) = 1$ it must be that $\alpha = \nu(x)$, but this is irrelevant for the rest of our proof).

Let $T_0 = 0, T_1 = T_x^+, \dots, T_{r+1} = \inf\{n > T_r : X_n = x\}, \dots$ denote the successive returns at x , and for any $r \geq 0$, $\epsilon_{r+1} := (X_{T_r}, \dots, X_{T_{r+1}-1})$ the r th excursion away from x . By strong Markov ($\epsilon_r, r \geq 1$) forms a sequence of i.i.d random variables.

Proof of ergodic theorem (ii)

Let $Z_r(f) := \sum_{k=T_{r-1}}^{T_r-1} f(X_k)$ (it is called an *additive functionnal* of ϵ_r), it follows that $(Z_r(f), r \geq 1)$ are also i.i.d. Observe that

$$\mathbb{E}[Z_1(f)] = \sum_{y \in E} \nu_x(y) f(y) = \alpha^{-1} \sum_{y \in E} \nu(y) f(y) < \infty$$

so the strong law of large numbers implies

$$\frac{1}{N} \sum_{r=1}^N Z_r(f) \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\lambda\text{-a.s.}} \alpha^{-1} \sum_{y \in E} \nu(y) f(y).$$

Proof of ergodic theorem (ii)

Let $r(n)$ be such that $T_{r(n)} \leq n < T_{r(n)+1}$. Because x is recurrent it must be that $r(n) \rightarrow \infty$ a.s. as $n \rightarrow \infty$.

Now, since we assumed f to be nonnegative, we have

$$\frac{\sum_{i=1}^{r(n)} Z_i(f)}{r(n)} \leq \frac{\sum_{k=0}^{n-1} f(X_k)}{r(n)} \leq \frac{\sum_{i=1}^{r(n)+1} Z_i(f)}{r(n)}.$$

Since $r(n) \rightarrow +\infty$, both LHS and RHS tend to $\alpha^{-1} \sum_{y \in E} \nu(y) f(y)$, so does the middle term. By performing the exact same reasoning with g , the desired result follows.

Proof of ergodic theorem (i) : the null recurrent case

Assume x is null recurrent. We have proven, only assuming recurrence, that

$$\frac{\sum_{k=0}^{n-1} f(X_k)}{r(n)} \rightarrow \alpha^{-1} \sum_{y \in E} \nu(y) f(y).$$

On the other hand we have a.s.

$$\frac{T_{r(n)}}{r(n)} \rightarrow \mathbb{E}_x[T_x^+] = +\infty,$$

and also $\frac{T_{r(n)+1}}{r(n)} \rightarrow +\infty$ a.s, because $r(n) \rightarrow \infty$. Since

$$\frac{T_{r(n)}}{r(n)} \leq \frac{n}{r(n)} \leq \frac{T_{r(n)+1}}{r(n)},$$

it follows that $\frac{n}{r(n)} \rightarrow +\infty$ a.s. Thus

$$\frac{\sum_{k=0}^{n-1} f(X_k)}{n} = \frac{\sum_{k=0}^{n-1} f(X_k)}{r(n)} \frac{r(n)}{n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\lambda \text{-a.s.}} 0 = \frac{1}{\mathbb{E}_x[T_x^]}.$$

Aperiodicity (exercice 11)

Definition

Let $\mathfrak{T}(x) := \{n \in \mathbb{N}^* : P^n(x, x) > 0\}$, and $d(x) = \gcd(\mathfrak{T}(x))$

Lemma

If $x \leftrightarrow y$, then $d(x) = d(y)$.

Proof : exercice 11.

Definition

Assume X is irreducible. If for some (hence all) $x \in E$ one has $d(x) = 1$ the chain is called *aperiodic*. If for some (hence all) $x \in E$ one has $d(x) = d \geq 2$ the chain is called *d-periodic*.

Aperiodicity (exercise 11)

Lemma

Assume X irreducible, aperiodic. For any x, y there exists $n_0(x, y) \in \mathbb{N}$ such that

$$\forall n \geq n_0(x, y) \quad P^n(x, y) > 0$$

Proof : When $x = y$, $\gcd([1, m] \cap \mathfrak{T}(x))$ decreases with n and tends to 1 so it eventually reaches value 1, at m_0 , say. Write $\{k_1, \dots, k_r\} := \gcd([1, m] \cap \mathfrak{T}(x))$, by Bezout there exists $(\ell_1, \dots, \ell_r) \in \mathbb{Z}$ such that $\sum_{i=1}^r \ell_i k_i = 1$.

With $n_0(x, x) = k_1 \sum_{i=1}^r |\ell_i| k_i$, any $n \geq n_0(x, x)$ can be written

$$n = k_1 \sum_{i=1}^r |\ell_i| k_i + k_1 a + b \sum_{i=1}^r \ell_i k_i,$$

with $\sum_{i=1}^r |\ell_i| k_i + a$ the quotient of Euclidian division of n by k_1 and $0 \leq b < k_1$, its remainder of Euclidian division of n by k_1 .

This proves the lemma for $x = y$.

For $x \neq y$, by irreducibility there exists $n_1 = n_1(x, y) \in \mathbb{N}$ such that $P^{n_1}(x, y) > 0$ and it suffices to take $n_0(x, y) = n_1 + n_0(x, x)$, so we are done with the proof of the lemma.

In the particular case when E is finite,

$N_0 := \max\{n_0(x, y), (x, y) \in E\}$ is finite. Then

$$\forall n \geq N_0, \forall (x, y) \in E^2, P^n(x, y) > 0,$$

in other words, every entry of the matrix P^n is positive.

When this property is satisfied, we say the chain is *strongly irreducible*. We have proven

Corollary

If E finite and X is irreducible aperiodic, then it is strongly irreducible.

d -periodic chain (exercise 12)

Theorem

Assume X is irreducible, d -periodic for some $d \geq 2$. Then there exists a partition of E into d classes C_0, \dots, C_{d-1} such that for any $r \in \{0, \dots, d-1\}$, $(X_{dk+r}, k \geq 0)$ is an irreducible aperiodic chain with state space C_r .

Proof : exercise 12.

Theorem

Assume X is irreducible, aperiodic, positive recurrent, and let π denote the unique invariant probability of X , and λ any starting distribution. Then

$$\forall x \in E \quad \mathbb{P}_\lambda(X_n = x) \xrightarrow{n \rightarrow \infty} \pi(x).$$

In other words, X_n converges in law towards $X_\infty \sim \pi$.

Proof of convergence theorem : preliminaries

In the case when E is finite, there are direct and easier proofs (using Perron-Frobenius theorem) than the Doeblin's proof we are going to present.

More generally, if the operator P possesses a spectral gap (it is automatically the case under the assumptions of the theorem when E is finite), not only do we have convergence, but in that case the convergence is exponentially fast. We'll come back to this when speaking of mixing times.

Doeblin's proof, on the other hand, relies on a *coupling* argument : when coupling occurs, the chain has reached equilibrium. The proof is general enough that it can not give any additional information on the speed at which this coupling occurs. Before getting into the proof, we shall digress and precise what we mean by *coupling*.

Definition

Assume μ and ν are two probability distributions. A *coupling* of μ and ν is any couple of random variables (X, Y) such that $X \sim \mu$ and $Y \sim \nu$.

For example, a coupling of $\mathbb{P}_{\lambda_1}, \mathbb{P}_{\lambda_2}$ (we'll also say a coupling of Markov chains started at λ_1, λ_2) is any pair of chains (X, Y) such that X is Markov (λ_1, P) and Y is Markov (λ_2, P) . If (X, Y) remains itself Markov we speak of a *Markovian coupling*.

Coupling of Markov chains

Recall we can use a mapping representation for the chain X , so that for some application $\phi : E \times [0, 1] \rightarrow E$ and $(U_n, n \geq 0)$ i.i.d $\sim \text{Unif}[0, 1]$, one has $X_{n+1} = \phi(X_n, U_n)$. Introducing $(V_n, n \geq 0)$ i.d. and independent of $(U_n, n \geq 0)$ allows to define very natural Markovian couplings : $X_0 \sim \lambda_1, Y_0 \sim \lambda_2$, and

- *independent coupling* :

$$X_{n+1} = \phi(X_n, U_n), \quad Y_{n+1} = \phi(Y_n, V_n)$$

- *independent/coalescing coupling* :

$$X_{n+1} = \phi(X_n, U_n), \quad Y_{n+1} = \begin{cases} \phi(Y_n, V_n) & \text{if } X_n \neq Y_n \\ \phi(Y_n, U_n) & \text{if } X_n = Y_n \end{cases}.$$

- *free coupling*

$$X_{n+1} = \phi(X_n, U_n), \quad Y_{n+1} = \phi(Y_n, U_n)$$

Coupling of Markov chains

The last two (independent/coalescent, free) are *coalescent couplings*, meaning that when the trajectories meet, from then on they remain together :

$$X_n = Y_n \Rightarrow X_{n+1} = Y_{n+1}$$

In general, independent coupling is *not* coalescing. However, as Doeblin observed, it has many interesting properties. First, the kernel Q of (X, Y) on $E \times E$ and its powers are easy to write down :

$$Q^n((x, y), (x', y')) = P^n(x, x')P^n(y, y') \quad \forall n \in \mathbb{N}.$$

Hence, if X is irreducible aperiodic, so is the independent coupling (X, Y) . Indeed by the lemma on aperiodicity, for any $n \geq \max(n_0(x, x'), n_0(y, y'))$, $Q^n((x, y), (x', y')) > 0$. Moreover if π is an invariant probability for X , then $\pi \otimes \pi$ is an invariant probability for (X, Y) . Thus if X is irreducible, aperiodic, positive recurrent, so is (X, Y) .

Doebelin's proof of convergence theorem

Recall X is assumed irreducible, aperiodic, positive recurrent. Consider the independent coupling (X, Y) and the independent/coalescing coupling (X, Z) with X started at λ and Y and Z both started at π . Let

$$T = \inf\{n \geq 0 : X_n = Y_n\} = \inf\{n \geq 0 : X_n = Z_n\}.$$

Now T is the hitting time of any point of the diagonal, so it is less than $T_{(x,x)}$ for some $x \in E$. It follows, by positive recurrence of (X, Y) , that $\mathbb{E}[T] \leq \mathbb{E}[T_{(x,x)}] < \infty$, in particular this implies that $T < \infty$ a.s.

Then for any $y \in E$,

$$|\mathbb{P}(X_n = y) - \pi(y)| = |\mathbb{P}(X_n = y) - \mathbb{P}(Z_n = y)| \leq \mathbb{P}(T > n) \rightarrow 0,$$

and we are done.