1 Ergodic theorem: proof

Let us first restate the theorem.

**Theorem 1.1** (Ergodic theorem). Let \( (X_n, n \geq 0) \) be an ergodic (i.e., irreducible, aperiodic and positive-recurrent) Markov chain with state space \( S \) and transition matrix \( P \). Then it admits a unique limiting and stationary distribution \( \pi \), i.e., \( \forall \pi(0), \lim_{n \to \infty} \pi^{(n)} = \pi \) and \( \pi = \pi P \).

1.1 Tools for the proof

**Total variation distance between two distributions.**

**Definition 1.2.** Let \( \mu \) and \( \nu \) be two distributions on the same state space \( S \) (i.e. \( 0 \leq \mu_i, \nu_i \leq 1 \), \( \sum_{i \in S} \mu_i = \sum_{i \in S} \nu_i = 1 \)). The total variation between \( \mu \) and \( \nu \) is defined as

\[
||\mu - \nu||_{TV} = \sup_{A \subseteq S} |(\mu(A) - \nu(A))|
\]

where \( \mu(A) = \sum_{i \in A} \mu_i \) and \( \nu(A) = \sum_{i \in A} \nu_i \).

**Properties.** (see exercises for the proof)

- \( 0 \leq ||\mu - \nu||_{TV} \leq 1 \). Moreover, \( ||\mu - \nu||_{TV} = 0 \) iff \( \mu = \nu \), and \( ||\mu - \nu||_{TV} = 1 \) iff \( \mu \) and \( \nu \) have disjoint support (i.e., \( \exists A \subseteq S \) such that \( \mu(A) = 1 \) and \( \nu(A) = 0 \)).
- \( ||\mu - \nu||_{TV} = \frac{1}{2} \sum_{i \in S} |\mu_i - \nu_i| \).
- Triangle inequality: \( ||\mu - \pi||_{TV} \leq ||\mu - \nu||_{TV} + ||\nu - \pi||_{TV} \)

**Coupling between two distributions.**

**Definition 1.3.** Let \( \mu, \nu \) be two distributions on \( S \). A coupling between \( \mu \) and \( \nu \) is a pair of random variables \( (X, Y) \) with a joint distribution on \( S \times S \) such that \( P(X = i) = \mu_i \) and \( P(Y = i) = \nu_i \), for \( i \in S \).

Note that there exist multiple possible couplings for a given pair \( \mu, \nu \).

**Example 1.4.** Consider \( S = \{0, 1\} \) and \( \mu_0 = \mu_1 = \nu_0 = \nu_1 = \frac{1}{2} \).

a) choose \( X, Y \) independent with \( P(X = i, Y = j) = \frac{1}{4}, \forall i, j \in S \) (statistical coupling)

b) choose \( X = Y \) with \( P(X = Y = 0) = P(X = Y = 1) = \frac{1}{2} \) (grand coupling)

**Proposition 1.5.** For every coupling \( (X, Y) \) of \( \mu, \nu \), we have \( ||\mu - \nu||_{TV} \leq P(X \neq Y) \)

**Proof.** Let \( A \) be any subset of \( S \):

\[
\mu(A) = P(X \in A) = P(X \in A, Y \in A) + P(X \in A, Y \in A^c)
\]

and

\[
\nu(A) = P(Y \in A) = P(X \in A, Y \in A) + P(X \in A^c, Y \in A)
\]

so

\[
\mu(A) - \nu(A) = P(X \in A, Y \in A) - P(X \in A^c, Y \in A) \leq P(X \in A, Y \in A^c) \leq P(X \neq Y)
\]

and

\[
\nu(A) - \mu(A) = P(X \in A^c, Y \in A) - P(X \in A, Y \in A^c) \leq P(X \in A^c, Y \in A) \leq P(X \neq Y)
\]

which in turn implies that

\[
||\mu - \nu||_{TV} = \sup_{A \subseteq S} |\mu(A) - \nu(A)| \leq P(X \neq Y)
\]

\( \square \)
Coupling between two Markov chains.

Let \((X_n, n \geq 0), (Y_n, n \geq 0)\) be two Markov chains on the same state set \(S\) and with the same transition matrix \(P\), but with initial distributions \(\mu\) and \(\nu\) respectively. As seen before, the distributions of these two Markov chains at time \(n\) are given by:

\[
P(X_n = i) = (\mu P^n) i, \quad \text{and} \quad P(Y_n = i) = (\nu P^n) i, \quad \text{for} \quad i \in S
\]

In order to couple \(X\) and \(Y\), we need to specify their joint distribution. One possibility is the following. Let \((Z_n = (X_n, Y_n), n \geq 0)\) be the process defined on the state space \(S \times S\) as:

- \(P(Z_0 = (i, k)) = \mu_i \nu_k, \forall i, k \in S\)
- Let \(X, Y\) evolve independently according to \(P\) (following the rules for their own chain) as long as \(X_n \neq Y_n\) (statistical coupling).
- As soon as \(X_n = Y_n\), the process coalesces, i.e., \(X_m = Y_m, \forall m \geq n\), and they evolve together according to \(P\) (grand coupling).

You should think of two people starting from two different random positions and walking randomly in town; when they meet by chance, they continue walking randomly, but together.

**Definition 1.6.** The **coupling time** of the chains \(X\) and \(Y\) is defined as \(\tau_c = \inf\{n \geq 1: X_n = Y_n\}\).

**Lemma 1.7.** For any \(n \geq 0\), it holds that:

\[
\|\mu P^n - \nu P^n\|_{TV} \leq P(\tau_c > n)
\]

**Proof.** The proof is a simple consequence of Proposition 1.5: for a given \(n \geq 0\), \(\mu P^n, \nu P^n\) are distributions on \(S\), and \((X_n, Y_n)\) is a coupling of these two distributions, so

\[
\|\mu P^n - \nu P^n\|_{TV} \leq P(X_n \neq Y_n) = P(\tau_c > n)
\]

\(\square\)

**1.2 Proof of the ergodic theorem**

Because the chain \((X_n, n \geq 0)\) is assumed to be irreducible and positive-recurrent, we know from the first theorem of last week that the chain admits a unique stationary distribution \(\pi\). What remains therefore to be proven is that for any initial distribution \(\pi^{(0)}\),

\[
\lim_{n \to \infty} \mathbb{P}(X_n = i) = \lim_{n \to \infty} \pi_i^{(n)} = \pi_i, \quad \forall i \in S
\]

We will actually prove something slightly stronger below, namely that for any \(\pi^{(0)}\),

\[
\lim_{n \to \infty} ||\pi^{(n)} - \pi||_{TV} = 0
\]

(this is equivalent to the above statement if \(S\) is finite and stronger if \(S\) is infinite). Let \(X\) (resp. \(Y\)) be the Markov chain with transition matrix \(P\) and initial distribution \(\pi^{(0)}\) (resp. \(\pi\)). We moreover assume that \(X\) and \(Y\) are coupled as described in the previous section. Then for all \(i \in S\), we have:

\[
\mathbb{P}(X_n = i) = (\pi^{(0)} P^n)_i = \pi_i^{(n)} \quad \text{and} \quad \mathbb{P}(Y_n = i) = (\pi P^n)_i = \pi_i
\]

and Lemma 1.7 asserts that

\[
||\pi^{(n)} - \pi||_{TV} = ||\pi^{(0)} P^n - \pi P^n||_{TV} \leq \mathbb{P}(X_n \neq Y_n) = \mathbb{P}(\tau_c > n)
\]
What remains therefore to be shown is that \( \lim_{n \to \infty} P(\tau_c > n) = 0 \).

**Remark.** Before we move on, let us observe the following: it is in general not true that at some time \( n \), the distribution \( \pi^{(n)} \) of the chain \( X_n \) becomes exactly equal to \( \pi \): this happens only for exceptional chains. The above coupling argument just proves that the total variation distance between \( \pi^{(n)} \) and \( \pi \) converges to 0 as \( n \) gets large.

Now, because

\[
\lim_{n \to \infty} P(\tau_c > n) = P(\tau_c > n, \forall n \geq 1) = P(\tau_c = +\infty) = 1 - P(\tau_c < +\infty)
\]

we obtain that the limit is equal to 0 iff \( P(\tau_c < +\infty) = 1 \).

Consider the chain \( (Z_n = (X_n, Y_n), n \geq 0) \) before coalescence. First, observe that it is a Markov chain on the state space \( S \times S \) with transition probabilities

\[
P(Z_{n+1} = (j, l)|Z_n = (i, k)) = p_{ij} p_{kl} = (P \otimes P)_{ik,jl}
\]

where \( P \otimes P \) denotes the tensor product of \( P \) with itself. It is here just a notation for the transition matrix of the chain \( Z \) with state space \( S \times S \).

Second, observe that the chain \( Z \) is itself irreducible and aperiodic. Indeed, it holds for an irreducible and aperiodic chain (like \( X \) and \( Y \)),

\[
\forall i, j \in S, \ \exists N(i, j) \text{ such that } \forall n \geq N(i, j), \ p_{ij}(n) > 0
\]

Thus, for the chain \( Z \), we have:

\[
\forall (i, k), (j, l) \in S \times S, \ \exists N(ik, jl) = \max(N(i, j), N(k, l)) \text{ such that } \forall n \geq N(ik, jl), \ P(Z_n = (jl)|Z_0 = (ik)) = p_{ik}(n) p_{kl}(n) > 0
\]

So the chain \( Z \) is irreducible and aperiodic.

Third, \( Z \) admits a stationary distribution. Indeed, consider the distribution \( \Pi = \pi \otimes \pi \), i.e. \( \Pi_{ik} = \pi_i \pi_k \). We have:

\[
((\pi \otimes \pi)(P \otimes P))_{jl} = \sum_{ik \in S} (\pi \otimes \pi)_{ik}(P \otimes P)_{ik,jl} = \sum_{i, k \in S} \pi_i \pi_k p_{ij} p_{kl} = \sum_{i \in S} \pi_i p_{ij} \sum_{i \in S} \pi_k p_{kl} = \pi_j \pi_l = (\pi \otimes \pi)_{jl}
\]

So far, we have shown that the Markov chain \( Z \), which is a coupling of our original Markov chain \( X \) and the Markov chain \( Y \) starting with the stationary distribution \( \pi \) as initial distribution, is irreducible, aperiodic and admits a stationary distribution. So by the first theorem of last week, \( Z \) is positive-recurrent. This will allow us to prove that \( P(\tau_c < +\infty) = 1 \).

For \((ik) \in S \times S\), define the first time \( Z \) reach state \((ik)\):

\[
T_{(ik)} = \inf\{n \geq 1 : Z_n = (ik)\}
\]

Since \( Z \) is positive-recurrent, we have:

\[
P(T_{(ik)} < +\infty|Z_0 = (ik)) = 1
\]

Considering then \( n \geq 1 \) such that \( p_{ik,jl}(n) > 0 \) (such an \( n \) is guaranteed to exist because \( Z \) is irreducible), we deduce that

\[
0 = P(T_{(ik)} = +\infty|Z_0 = (ik)) \geq P(T_{(ik)} = +\infty, Z_n = (jl)|Z_0 = (ik)) = P(T_{(ik)} = +\infty|Z_n = (jl), Z_0 = (ik)) \cdot p_{ik,jl}(n)
\]
Using $p_{ik, jl}(n) > 0$, as well as the Markov property and the time homogeneity, we obtain

$$
P(T_{(ik)} = +\infty | Z_0 = (jl)) = 0
$$

or equivalently:

$$
P(T_{(ik)} < +\infty | Z_0 = (jl)) = 1
$$

Compared to the positive-recurrent property, this says that $Z$ will reach state $(ik)$ in finite time with probability 1 not only starting from state $(ik)$, but from any other state $(jl)$ also.

Consider now any $i = k \in S$, and $j, l \in S$. We have

$$
P(T_{(ii)} < +\infty | Z_0 = (jl)) = 1
$$

Observing that $\tau_c \leq T_{(ii)}$ for any $i$ (as for a given $i \in S$, $T_{(ii)}$ is a just possible coupling time), we finally obtain that for any $j, l \in S$,

$$
P(\tau_c < +\infty | Z_0 = (jl)) = 1
$$

which completes the proof. \(\square\)

**Note:** A last formal step would be needed here to deduce that for any initial distribution of $Z$ on $S \times S$, we have $P(\tau_c < +\infty) = 1$. We indeed only showed here that $P(\tau_c < +\infty) = 1$ starting from any initial state $(jl)$. In case of a finite $S$, these two statements are clearly equivalent. In the infinite setting, this requires a proof.