Markov Chains and Algorithmic Applications: WEEK 4

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 π, 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 μ and ν 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 μ and ν is defined as

$$||\mu - \nu||_{\mathrm{TV}} = \sup_{A \subset S} |\mu(A) - \nu(A)|$$

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

Properties. (see exercises for the proof)

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

Coupling between two distributions.

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

Note that there exist multiple possible couplings for a given pair μ, ν.

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 $\mathbb{P}(X = i, Y = j) = \frac{1}{4}$, $\forall i, j \in S$ (**statistical coupling**)

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

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

Proof. Let A be any subset of S:

$$\mu(A) = \mathbb{P}(X \in A) = \mathbb{P}(X \in A, Y \in A) + \mathbb{P}(X \in A, Y \in A^c)$$

and

$$\nu(A) = \mathbb{P}(Y \in A) = \mathbb{P}(X \in A, Y \in A) + \mathbb{P}(X \in A^c, Y \in A)$$

so

$$\mu(A) - \nu(A) = \mathbb{P}(X \in A, Y \in A^c) - \mathbb{P}(X \in A^c, Y \in A) \leq \mathbb{P}(X \in A, Y \in A^c) \leq \mathbb{P}(X \neq Y)$$

and

$$\nu(A) - \mu(A) = \mathbb{P}(X \in A^c, Y \in A) - \mathbb{P}(X \in A, Y \in A^c) \leq \mathbb{P}(X \in A^c, Y \in A) \leq \mathbb{P}(X \neq Y)$$

which in turn implies that

$$||\mu - \nu||_{\mathrm{TV}} = \sup_{A\subset S} |\mu(A) - \nu(A)| \leq \mathbb{P}(X \neq Y)$$

□

Coupling between two Markov chains.

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

$$\mathbb{P}(X_n = i) = (\mu P^n)_i \quad \text{and} \quad \mathbb{P}(Y_n = i) = (\nu P^n)_i \quad \text{for } i \in S$$

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:

- $\mathbb{P}(Z_0 = (i,k)) = \mu_i \nu_k$, $\forall i, k \in S$
- 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 $\tau_c = \inf\{n \geq 1\colon\ X_n = Y_n\}$.

Lemma 1.7. For any $n \geq 0$, it holds that:

$$|| \underbrace{\mu P^n}_{\substack{\text{distribution of X} \\ \text{at time n}}} - \underbrace{\nu P^n}_{\substack{\text{distribution of Y} \\ \text{at time n}}} ||_{\text{TV}} \leq \mathbb{P}(\tau_c > n)$$

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

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

□

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 π. 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 $\pi^{(0)}$,

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

(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. π). 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

$$||\pi^{(n)} - \pi||_{\text{TV}} = ||\pi^{(0)} P^n - \pi P^n||_{\text{TV}} \leq \mathbb{P}(X_n \neq Y_n) = \mathbb{P}(\tau_c > n)$$

What remains therefore to be shown is that $\lim_{n\to\infty} \mathbb{P}(\tau_c > n) = 0$.

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 π: this happens only for exceptional chains. The above coupling argument just proves that the total variation distance between $\pi^{(n)}$ and π converges to 0 as n gets large.

Now, because

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

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

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

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

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),

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

Thus, for the chain Z, we have:

$$\begin{aligned}
&\forall (i,k), (j,l) \in S \times S, \quad \exists N(ik, jl) = \max(N(i,j), N(k,l)) \quad \text{such that} \\
&\forall n \geq N(ik, jl), \quad \mathbb{P}(Z_n = (jl) | Z_0 = (ik)) = p_{ij}(n)\, p_{kl}(n) > 0
\end{aligned}$$

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:

$$\begin{aligned}
((\pi \otimes \pi)(P \otimes P))_{jl} &= \sum_{ik \in S} (\pi \otimes \pi)_{ik} (P \otimes P)_{ik,jl} \\
&= \sum_{i,k \in S} \pi_i \pi_k P_{ij} P_{kl} = \sum_{i \in S} \pi_i P_{ij} \sum_{i \in S} \pi_k P_{kl} = \pi_j \pi_l = (\pi \otimes \pi)_{jl}
\end{aligned}$$

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 π 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 $\mathbb{P}(\tau_c < +\infty) = 1$.

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:

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

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

$$\begin{aligned}
0 &= \mathbb{P}(T_{(ik)} = +\infty | Z_0 = (ik)) \\
&\geq \mathbb{P}(T_{(ik)} = +\infty, Z_n = (jl) | Z_0 = (ik)) \\
&= \mathbb{P}(T_{(ik)} = +\infty | Z_n = (jl), Z_0 = (ik)) \cdot p_{ik,jl}(n)
\end{aligned}$$

Using $p_{ik,jl}(n) > 0$, as well as the Markov property and the time homogeneity, we obtain

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

or equivalently:

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

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

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

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 $j, l \in S$,

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

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 $\mathbb{P}(\tau_c < +\infty) = 1$. We indeed only showed here that $\mathbb{P}(\tau_c < +\infty) = 1$ 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.