1.10 Ergodic theorem

Ergodic theorems concern the limiting behaviour of averages over time. We shall prove a theorem which identifies for Markov chains the long-run proportion of time spent in each state. An essential tool is the following ergodic theorem for independent random variables which is a version of the strong law of large numbers.

Theorem 1.10.1 (Strong law of large numbers). *Let $Y_1, Y_2, \dots$ be a sequence of independent, identically distributed, non-negative random variables with $E(Y_1) = \mu$. Then*

$$P\left(\frac{Y_1 + \dots + Y_n}{n} \to \mu \text{ as } n \to \infty\right) = 1.$$

Proof. A proof for the case $\mu < \infty$ may be found, for example, in *Probability with Martingales* by David Williams (Cambridge University Press, 1991). The case where $\mu = \infty$ is a simple deduction. Fix $N < \infty$ and set $Y_n^{(N)} = Y_n \wedge N$. Then

$$\frac{Y_1 + \dots + Y_n}{n} \geq \frac{Y_1^{(N)} + \dots + Y_n^{(N)}}{n} \to E(Y_1 \wedge N) \qquad \text{as } n \to \infty$$

with probability one. As $N \uparrow \infty$ we have $E(Y_1 \wedge N) \uparrow \mu$ by monotone convergence (see Section 6.4). So we must have, with probability 1

$$\frac{Y_1 + \dots + Y_n}{n} \to \infty \qquad \text{as } n \to \infty. \qquad \square$$

We denote by $V_i(n)$ the *number of visits to i before n*:

$$V_i(n) = \sum_{k=0}^{n-1} 1_{\{X_k = i\}}.$$

Then $V_i(n)/n$ is the proportion of time before n spent in state i. The following result gives the long-run proportion of time spent by a Markov chain in each state.

Theorem 1.10.2 (Ergodic theorem). *Let P be irreducible and let λ be any distribution. If $(X_n)_{n\geq 0}$ is Markov(λ, P) then*

$$P\left(\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ as } n \to \infty\right) = 1$$

where $m_i = E_i(T_i)$ is the expected return time to state i. Moreover, in the positive recurrent case, for any bounded function $f : I \to R$ we have

$$P\left(\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \overline{f} \text{ as } n \to \infty\right) = 1$$

where

$$\overline{f} = \sum_{i \in I} \pi_i f_i$$

and where $(\pi_i : i \in I)$ *is the unique invariant distribution.*

Proof. If P is transient, then, with probability 1, the total number V_i of visits to i is finite, so

$$\frac{V_i(n)}{n} \leq \frac{V_i}{n} \to 0 = \frac{1}{m_i}.$$

Suppose then that P is recurrent and fix a state i. For $T = T_i$ we have $P(T < \infty) = 1$ by Theorem 1.5.7 and $(X_{T+n})_{n \geq 0}$ is Markov(δ_i, P) and independent of $X_0, X_1, \ldots, X_T$ by the strong Markov property. The long-run proportion of time spent in i is the same for $(X_{T+n})_{n \geq 0}$ and $(X_n)_{n \geq 0}$, so it suffices to consider the case $\lambda = \delta_i$.

Write $S_i^{(r)}$ for the length of the rth excursion to i, as in Section 1.5. By Lemma 1.5.1, the non-negative random variables $S_i^{(1)}, S_i^{(2)}, \ldots$ are independent and identically distributed with $E_i(S_i^{(r)}) = m_i$. Now

$$S_i^{(1)} + \ldots + S_i^{(V_i(n)-1)} \leq n - 1,$$

the left-hand side being the time of the last visit to i before n. Also

$$S_i^{(1)} + \ldots + S_i^{(V_i(n))} \geq n,$$

the left-hand side being the time of the first visit to i after $n - 1$. Hence

$$\frac{S_i^{(1)} + \ldots + S_i^{(V_i(n)-1)}}{V_i(n)} \leq \frac{n}{V_i(n)} \leq \frac{S_i^{(1)} + \ldots + S_i^{(V_i(n))}}{V_i(n)}. \tag{1.8}$$

By the strong law of large numbers

$$P\left(\frac{S_i^{(1)} + \ldots + S_i^{(n)}}{n} \to m_i \text{ as } n \to \infty\right) = 1$$

and, since P is recurrent

$$P(V_i(n) \to \infty \text{ as } n \to \infty) = 1.$$

So, letting $n \to \infty$ in (1.8), we get

$$P\left(\frac{n}{V_i(n)} \to m_i \text{ as } n \to \infty\right) = 1,$$

which implies

$$P\left(\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ as } n \to \infty\right) = 1.$$

Assume now that $(X_n)_{n\geq 0}$ has an invariant distribution $(\pi_i : i \in I)$. Let $f : I \to R$ be a bounded function and assume without loss of generality that $|f| \leq 1$. For any $J \subseteq I$ we have

$$\begin{aligned}
\left|\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \overline{f}\right| &= \left|\sum_{i\in I}\left(\frac{V_i(n)}{n} - \pi_i\right) f_i\right| \\
&\leq \sum_{i\in J}\left|\frac{V_i(n)}{n} - \pi_i\right| + \sum_{i\notin J}\left|\frac{V_i(n)}{n} - \pi_i\right| \\
&\leq \sum_{i\in J}\left|\frac{V_i(n)}{n} - \pi_i\right| + \sum_{i\notin J}\left(\frac{V_i(n)}{n} + \pi_i\right) \\
&\leq 2\sum_{i\in J}\left|\frac{V_i(n)}{n} - \pi_i\right| + 2\sum_{i\notin J}\pi_i.
\end{aligned}$$

We proved above that

$$P\left(\frac{V_i(n)}{n} \to \pi_i \text{ as } n \to \infty \text{ for all } i\right) = 1.$$

Given $\varepsilon > 0$, choose J finite so that

$$\sum_{i\notin J}\pi_i < \varepsilon/4$$

and then $N = N(\omega)$ so that, for $n \geq N(\omega)$

$$\sum_{i\in J}\left|\frac{V_i(n)}{n} - \pi_i\right| < \varepsilon/4.$$

Then, for $n \geq N(\omega)$, we have

$$\left|\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \overline{f}\right| < \varepsilon,$$

which establishes the desired convergence. □

We consider now the statistical problem of estimating an unknown transition matrix P on the basis of observations of the corresponding Markov chain. Consider, to begin, the case where we have $N+1$ observations $(X_n)_{0\leq n\leq N}$. The log-likelihood function is given by

$$l(P) = \log(\lambda_{X_0} p_{X_0X_1} \dots p_{X_{N-1}X_N}) = \sum_{i,j\in I} N_{ij}\log p_{ij}$$

up to a constant independent of P, where N_{ij} is the number of transitions from i to j. A standard statistical procedure is to find the *maximum likelihood estimate* $\widehat{P}$, which is the choice of P maximizing $l(P)$. Since P must satisfy the linear constraint $\sum_j p_{ij} = 1$ for each i, we first try to maximize

$$l(P) + \sum_{i,j \in I} \mu_i p_{ij}$$

and then choose $(\mu_i : i \in I)$ to fit the constraints. This is the method of Lagrange multipliers. Thus we find

$$\widehat{p}_{ij} = \sum_{n=0}^{N-1} 1_{\{X_n = i, X_{n+1} = j\}} \Big/ \sum_{n=0}^{N-1} 1_{\{X_n = i\}}$$

which is the proportion of jumps from i which go to j.

We now turn to consider the *consistency* of this sort of estimate, that is to say whether $\widehat{p}_{ij} \to p_{ij}$ with probability 1 as $N \to \infty$. Since this is clearly false when i is transient, we shall slightly modify our approach. Note that to find $\widehat{p}_{ij}$ we simply have to maximize

$$\sum_{j \in I} N_{ij} \log p_{ij}$$

subject to $\sum_j p_{ij} = 1$: the other terms and constraints are irrelevant. Suppose then that instead of $N + 1$ observations we make enough observations to ensure the chain leaves state i a total of N times. In the transient case this may involve restarting the chain several times. Denote again by N_{ij} the number of transitions from i to j.

To maximize the likelihood for $(p_{ij} : j \in I)$ we still maximize

$$\sum_{j \in I} N_{ij} \log p_{ij}$$

subject to $\sum_j p_{ij} = 1$, which leads to the maximum likelihood estimate

$$\widehat{p}_{ij} = N_{ij}/N.$$

But $N_{ij} = Y_1 + \ldots + Y_N$, where $Y_n = 1$ if the nth transition from i is to j, and $Y_n = 0$ otherwise. By the strong Markov property $Y_1, \ldots, Y_N$ are independent and identically distributed random variables with mean p_{ij}. So, by the strong law of large numbers

$$P(\widehat{p}_{ij} \to p_{ij} \text{ as } N \to \infty) = 1,$$

which shows that $\widehat{p}_{ij}$ is consistent.

Exercises

1.10.1 Prove the claim (d) made in example (v) of the Introduction.

1.10.2 A professor has N umbrellas. He walks to the office in the morning and walks home in the evening. If it is raining he likes to carry an umbrella and if it is fine he does not. Suppose that it rains on each journey with probability p, independently of past weather. What is the long-run proportion of journeys on which the professor gets wet?

1.10.3 Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain on I having an invariant distribution π. For $J \subseteq I$ let $(Y_m)_{m\geq 0}$ be the Markov chain on J obtained by observing $(X_n)_{n\geq 0}$ whilst in J. (See Example 1.4.4.) Show that $(Y_m)_{m\geq 0}$ is positive recurrent and find its invariant distribution.

1.10.4 An opera singer is due to perform a long series of concerts. Having a fine artistic temperament, she is liable to pull out each night with probability 1/2. Once this has happened she will not sing again until the promoter convinces her of his high regard. This he does by sending flowers every day until she returns. Flowers costing x thousand pounds, $0 \leq x \leq 1$, bring about a reconciliation with probability $\sqrt{x}$. The promoter stands to make £750 from each successful concert. How much should he spend on flowers?