is irreducible then $I - P + A$ is invertible. *Note that this enables one to compute the invariant distribution by any standard method of inverting a matrix.*

1.8 Convergence to equilibrium

We shall investigate the limiting behaviour of the n-step transition probabilities $p_{ij}^{(n)}$ as $n \to \infty$. As we saw in Theorem 1.7.2, if the state-space is finite and if for some i the limit exists for all j, then it must be an invariant distribution. But, as the following example shows, the limit does not always exist.

Example 1.8.1

Consider the two-state chain with transition matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then $P^2 = I$, so $P^{2n} = I$ and $P^{2n+1} = P$ for all n. Thus $p_{ij}^{(n)}$ fails to converge for all i, j.

Let us call a state i *aperiodic* if $p_{ii}^{(n)} > 0$ for all sufficiently large n. We leave it as an exercise to show that i is aperiodic if and only if the set $\{n \geq 0 : p_{ii}^{(n)} > 0\}$ has no common divisor other than 1. This is also a consequence of Theorem 1.8.4. The behaviour of the chain in Example 1.8.1 is connected with its periodicity.

Lemma 1.8.2. *Suppose P is irreducible and has an aperiodic state i. Then, for all states j and k, $p_{jk}^{(n)} > 0$ for all sufficiently large n. In particular, all states are aperiodic.*

Proof. There exist $r, s \geq 0$ with $p_{ji}^{(r)}, p_{ik}^{(s)} > 0$. Then

$$p_{jk}^{(r+n+s)} \geq p_{ji}^{(r)} p_{ii}^{(n)} p_{ik}^{(s)} > 0$$

for all sufficiently large n. $\square$

Here is the main result of this section. The method of proof, by coupling two Markov chains, is ingenious.

Theorem 1.8.3 (Convergence to equilibrium). *Let P be irreducible and aperiodic, and suppose that P has an invariant distribution π. Let λ be any distribution. Suppose that $(X_n)_{n\geq 0}$ is Markov(λ, P). Then*

$$\mathbb{P}(X_n = j) \to \pi_j \quad \text{as } n \to \infty \text{ for all } j.$$

In particular

$$p_{ij}^{(n)} \to \pi_j \quad \text{as } n \to \infty \text{ for all } i, j.$$

Proof. We use a coupling argument. Let $(Y_n)_{n\geq 0}$ be Markov(π, P) and independent of $(X_n)_{n\geq 0}$. Fix a reference state b and set $T = \inf\{n \geq 1 : X_n = Y_n = b\}$.

Step 1. We show $\mathbb{P}(T < \infty) = 1$. The process $W_n = (X_n, Y_n)$ is a Markov chain on $I \times I$ with transition probabilities

$$\widetilde{p}_{(i,k)(j,l)} = p_{ij} p_{kl}$$

and initial distribution

$$\mu_{(i,k)} = \lambda_i \pi_k .$$

Since P is aperiodic, for all states i, j, k, l we have

$$\widetilde{p}^{(n)}_{(i,k)(j,l)} = p^{(n)}_{ij} p^{(n)}_{kl} > 0$$

for all sufficiently large n; so $\widetilde{P}$ is irreducible. Also, $\widetilde{P}$ has an invariant distribution given by

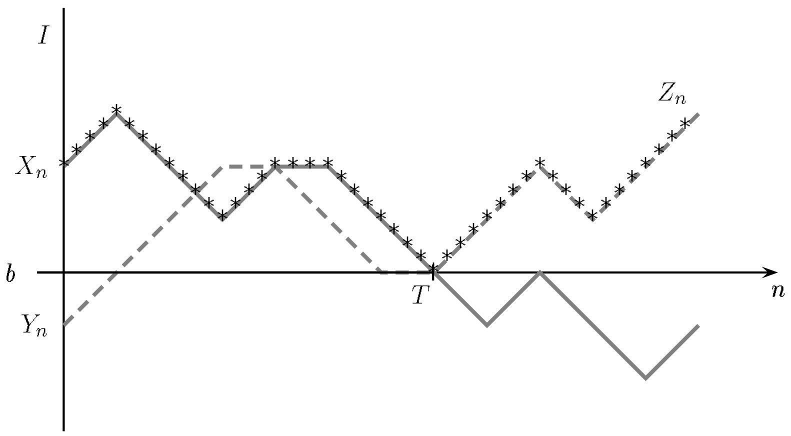
$$\widetilde{\pi}_{(i,k)} = \pi_i \pi_k$$

so, by Theorem 1.7.7, $\widetilde{P}$ is positive recurrent. But T is the first passage time of W_n to (b, b) so $\mathbb{P}(T < \infty) = 1$, by Theorem 1.5.7.

Step 2. Set

$$Z_n = \begin{cases} X_n & \text{if } n < T \\ Y_n & \text{if } n \geq T. \end{cases}$$

The diagram below illustrates the idea. We show that $(Z_n)_{n \geq 0}$ is Markov(λ, P).



The strong Markov property applies to $(W_n)_{n \geq 0}$ at time T, so $(X_{T+n}, Y_{T+n})_{n \geq 0}$ is Markov$(\delta_{(b,b)}, \widetilde{P})$ and independent of $(X_0, Y_0), (X_1, Y_1), \dots, (X_T, Y_T)$. By symmetry, we can replace the process $(X_{T+n}, Y_{T+n})_{n \geq 0}$ by $(Y_{T+n}, X_{T+n})_{n \geq 0}$ which is also Markov$(\delta_{(b,b)}, \widetilde{P})$ and remains independent of $(X_0, Y_0), (X_1, Y_1), \dots, (X_T, Y_T)$. Hence $W'_n = (Z_n, Z'_n)$ is Markov$(\mu, \widetilde{P})$ where

$$Z'_n = \begin{cases} Y_n & \text{if } n < T \\ X_n & \text{if } n \geq T. \end{cases}$$

In particular, $(Z_n)_{n\geq 0}$ is Markov(λ, P).

Step 3. We have

$$\mathbb{P}(Z_n = j) = \mathbb{P}(X_n = j \text{ and } n < T) + \mathbb{P}(Y_n = j \text{ and } n \geq T)$$

so

$$\begin{aligned}|\mathbb{P}(X_n = j) - \pi_j| &= |\mathbb{P}(Z_n = j) - \mathbb{P}(Y_n = j)|\\ &= |\mathbb{P}(X_n = j \text{ and } n < T) - \mathbb{P}(Y_n = j \text{ and } n < T)|\\ &\leq \mathbb{P}(n < T)\end{aligned}$$

and $\mathbb{P}(n < T) \to 0$ as $n \to \infty$. □

To understand this proof one should see what goes wrong when P is not aperiodic. Consider the two-state chain of Example 1.8.1 which has $(1/2, 1/2)$ as its unique invariant distribution. We start $(X_n)_{n\geq 0}$ from 0 and $(Y_n)_{n\geq 0}$ with equal probability from 0 or 1. However, if $Y_0 = 1$, then, because of periodicity, $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ will never meet, and the proof fails. We move on now to the cases that were excluded in the last theorem, where $(X_t)_{\geq 0}$ is periodic or transient or null recurrent. The remainder of this section might be omitted on a first reading.

Theorem 1.8.4. *Let P be irreducible. There is an integer $d \geq 1$ and a partition*

$$I = C_0 \cup C_1 \cup \ldots \cup C_{d-1}$$

such that (setting $C_{nd+r} = C_r$)

(i) $p_{ij}^{(n)} > 0$ *only if $i \in C_r$ and $j \in C_{r+n}$ for some r;*

(ii) $p_{ij}^{(nd)} > 0$ *for all sufficiently large n, for all $i, j \in C_r$, for all r.*

Proof. Fix a state k and consider $S = \{n \geq 0 : p_{kk}^{(n)} > 0\}$. Choose $n_1, n_2 \in S$ with $n_1 < n_2$ and such that $d := n_2 - n_1$ is as small as possible. (Here and throughout we use the symbol $:=$ to mean 'defined to equal'.) Define for $r = 0, \ldots, d-1$

$$C_r = \{i \in I : p_{ki}^{(nd+r)} > 0 \text{ for some } n \geq 0\}.$$

Then $C_0 \cup \ldots \cup C_{d-1} = I$, by irreducibility. Moreover, if $p_{ki}^{(nd+r)} > 0$ and $p_{ki}^{(nd+s)} > 0$ for some $r, s \in \{0, 1, \ldots, d-1\}$, then, choosing $m \geq 0$ so that $p_{ik}^{(m)} > 0$, we have $p_{kk}^{(nd+r+m)} > 0$ and $p_{kk}^{(nd+s+m)} > 0$ so $r = s$ by minimality of d. Hence we have a partition.

To prove (i) suppose $p_{ij}^{(n)} > 0$ and $i \in C_r$. Choose m so that $p_{ki}^{(md+r)} > 0$, then $p_{kj}^{(md+r+k)} > 0$ so $j \in C_{r+n}$ as required. By taking $i = j = k$ we now see that d must divide every element of S, in particular n_1.

Now for $nd \geq n_1^2$, we can write $nd = qn_1 + r$ for integers $q \geq n_1$ and $0 \leq r \leq n_1 - 1$. Since d divides n_1 we then have $r = md$ for some integer m and then $nd = (q-m)n_1 + mn_2$. Hence

$$p_{kk}^{(nd)} \geq (p_{kk}^{(n_1)})^{q-m}(p_{kk}^{(n_2)})^m > 0$$

and hence $nd \in S$. To prove (ii) for $i, j \in C_r$ choose m_1 and m_2 so that $p_{ik}^{(m_1)} > 0$ and $p_{kj}^{(m_2)} > 0$, then

$$p_{ij}^{(m_1+nd+m_2)} \geq p_{ik}^{(m_1)} p_{kk}^{(nd)} p_{kj}^{(m_2)} > 0$$

whenever $nd \geq n_1^2$. Since $m_1 + m_2$ is then necessarily a multiple of d, we are done. □

We call d the *period* of P. The theorem just proved shows in particular for all $i \in I$ that d is the greatest common divisor of the set $\{n \geq 0 : p_{ii}^{(n)} > 0\}$. This is sometimes useful in identifying d.

Finally, here is a complete description of limiting behaviour for irreducible chains. This generalizes Theorem 1.8.3 in two respects since we require neither aperiodicity nor the existence of an invariant distribution. The argument we use for the null recurrent case was discovered recently by B. Fristedt and L. Gray.

Theorem 1.8.5. *Let P be irreducible of period d and let $C_0, C_1, \dots, C_{d-1}$ be the partition obtained in Theorem 1.8.4. Let λ be a distribution with $\sum_{i \in C_0} \lambda_i = 1$. Suppose that $(X_n)_{n\geq 0}$ is Markov(λ, P). Then for $r = 0, 1, \dots, d-1$ and $j \in C_r$ we have*

$$\mathbb{P}(X_{nd+r} = j) \to d/m_j \quad \text{as } n \to \infty$$

where m_j is the expected return time to j. In particular, for $i \in C_0$ and $j \in C_r$ we have

$$p_{ij}^{(nd+r)} \to d/m_j \quad \text{as } n \to \infty.$$

Proof.

Step 1. We reduce to the aperiodic case. Set $\nu = \lambda P^r$, then by Theorem 1.8.4 we have

$$\sum_{i \in C_r} \nu_i = 1.$$

Set $Y_n = X_{nd+r}$, then $(Y_n)_{n\geq 0}$ is Markov(ν, P^d) and, by Theorem 1.8.4, P^d is irreducible and aperiodic on C_r. For $j \in C_r$ the expected return time of $(Y_n)_{n\geq 0}$ to j is m_j/d. So if the theorem holds in the aperiodic case, then

$$\mathbb{P}(X_{nd+r} = j) = \mathbb{P}(Y_n = j) \to d/m_j \quad \text{as } n \to \infty$$

so the theorem holds in general.

Step 2. Assume that P is aperiodic. If P is positive recurrent then $1/m_j = \pi_j$, where π is the unique invariant distribution, so the result follows from Theorem 1.8.3. Otherwise $m_j = \infty$ and we have to show that

$$\mathbb{P}(X_n = j) \to 0 \quad \text{as } n \to \infty.$$

If P is transient this is easy and we are left with the null recurrent case.

Step 3. Assume that P is aperiodic and null recurrent. Then

$$\sum_{k=0}^{\infty} \mathbb{P}_j(T_j > k) = \mathbb{E}_j(T_j) = \infty.$$

Given $\varepsilon > 0$ choose K so that

$$\sum_{k=0}^{K-1} \mathbb{P}_j(T_j > k) \geq \frac{2}{\varepsilon}\,.$$

Then, for $n \geq K-1$

$$\begin{aligned}
1 &\geq \sum_{k=n-K+1}^{n} \mathbb{P}(X_k = j \text{ and } X_m \neq j \text{ for } m = k+1, \dots, n) \\
&= \sum_{k=n-K+1}^{n} \mathbb{P}(X_k = j)\mathbb{P}_j(T_j > n-k) \\
&= \sum_{k=0}^{K-1} \mathbb{P}(X_{n-k} = j)\mathbb{P}_j(T_j > k)
\end{aligned}$$

so we must have $\mathbb{P}(X_{n-k} = j) \leq \varepsilon/2$ for some $k \in \{0, 1, \dots, K-1\}$.

Return now to the coupling argument used in Theorem 1.8.3, only now let $(Y_n)_{n\geq 0}$ be Markov(μ, P), where μ is to be chosen later. Set $W_n = (X_n, Y_n)$. As before, aperiodicity of $(X_n)_{n\geq 0}$ ensures irreducibility of $(W_n)_{n\geq 0}$. If $(W_n)_{n\geq 0}$ is transient then, on taking $\mu = \lambda$, we obtain

$$\mathbb{P}(X_n = j)^2 = \mathbb{P}\big(W_n = (j,j)\big) \to 0$$

as required. Assume then that $(W_n)_{n\geq 0}$ is recurrent. Then, in the notation of Theorem 1.8.3, we have $\mathbb{P}(T < \infty) = 1$ and the coupling argument shows that

$$|\mathbb{P}(X_n = j) - \mathbb{P}(Y_n = j)| \to 0 \qquad \text{as } n \to \infty.$$

We exploit this convergence by taking $\mu = \lambda P^k$ for $k = 1, \dots, K-1$, so that $\mathbb{P}(Y_n = j) = \mathbb{P}(X_{n+k} = j)$. We can find N such that for $n \geq N$ and $k = 1, \dots, K-1$,

$$|\mathbb{P}(X_n = j) - \mathbb{P}(X_{n+k} = j)| \leq \frac{\varepsilon}{2}\,.$$

But for any n we can find $k \in \{0, 1, \dots, K-1\}$ such that $\mathbb{P}(X_{n+k} = j) \leq \varepsilon/2$. Hence, for $n \geq N$

$$\mathbb{P}(X_n = j) \leq \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this shows that $\mathbb{P}(X_n = j) \to 0$ as $n \to \infty$, as required. $\square$

Exercises

1.8.1 Prove the claims (e), (f) and (g) made in example (v) of the Introduction.

1.8.2 Find the invariant distributions of the transition matrices in Exercise 1.1.7, parts (a), (b) and (c), and compare them with your answers there.

1.8.3 A fair die is thrown repeatedly. Let X_n denote the sum of the first n throws. Find

$$\lim_{n\to\infty} \mathbb{P}(X_n \text{ is a multiple of } 13)$$

quoting carefully any general theorems that you use.

1.8.4 Each morning a student takes one of the three books he owns from his shelf. The probability that he chooses book i is α_i, where $0 < \alpha_i < 1$ for $i = 1, 2, 3$, and choices on successive days are independent. In the evening he replaces the book at the left-hand end of the shelf. If p_n denotes the probability that on day n the student finds the books in the order 1,2,3, from left to right, show that, irrespective of the initial arrangement of the books, p_n converges as $n \to \infty$, and determine the limit.

1.8.5 (Renewal theorem). Let $Y_1, Y_2, \dots$ be independent, identically distributed random variables with values in $\{1, 2, \dots\}$. Suppose that the set of integers

$$\{n : \mathbb{P}(Y_1 = n) > 1\}$$

has greatest common divisor 1. Set $\mu = \mathbb{E}(Y_1)$. Show that the following process is a Markov chain:

$$X_n = \inf\{m \geq n : m = Y_1 + \dots + Y_k \text{ for some } k \geq 0\} - n.$$

Determine

$$\lim_{n\to\infty} \mathbb{P}(X_n = 0)$$

and hence show that as $n \to \infty$

$$\mathbb{P}(n = Y_1 + \dots + Y_k \text{ for some } k \geq 0) \to 1/\mu.$$

(*Think of* $Y_1, Y_2, \dots$ *as light-bulb lifetimes. A bulb is replaced when it fails. Thus the limiting probability that a bulb is replaced at time* n *is* $1/\mu$. *Although this appears to be a very special case of convergence to equilibrium, one can actually recover the full result by applying the renewal theorem to the excursion lengths* $S_i^{(1)}, S_i^{(2)}, \dots$ *from state* i.)

1.9 Time reversal

For Markov chains, the past and future are independent given the present. This property is symmetrical in time and suggests looking at Markov chains with time