Basic Markov Chains

Pierre Brémaud

December 9, 2015

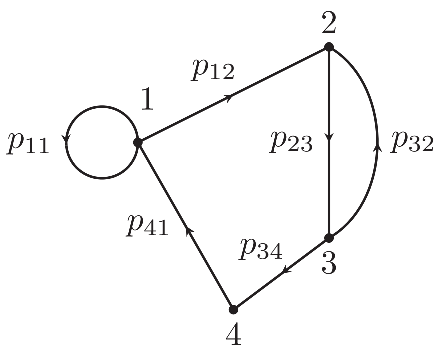
Contents

Chapter 1

The transition matrix

1.1 The distribution of a Markov chain

A particle moves on a denumerable set E. If at time n, the particle is in position $X_n = i$, it will be at time $n+1$ in a position $X_{n+1} = j$ chosen independently of the past trajectory X_{n-1}, X_{n-2} with probability p_{ij}. This can be represented by a labeled directed graph, called the *transition graph*, whose set of vertices is E, and for which there is a directed edge from $i \in E$ to $j \in E$ with label p_{ij} if and only the latter quantity is positive. Note that there may be "self-loops", corresponding to positions i such that $p_{ii} > 0$.



This graphical interpretation of as Markov chain in terms of a "random walk" on a set E is adapted to the study of random walks on graphs. Since the interpretation of a Markov chain in such terms is not always the natural one, we proceed to give a more formal definition.

Recall that a sequence $\{X_n\}_{n\geq 0}$ of random variables with values in a set E is called a *discrete-time stochastic process* with state space E. In this chapter, the state space is countable, and its elements will be denoted by i, j, k,... If $X_n = i$, the process is said to be in state i at time n, or to visit state i at time n.

Definition 1.1.1 *If for all integers $n \geq 0$ and all states $i_0, i_1, \dots, i_{n-1}, i, j$,*

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j \mid X_n = i)\,,$$

this stochastic process is called a Markov chain*, and a* homogeneous *Markov chain* (HMC) *if, in addition, the right-hand side is independent of n.*

The matrix $\mathbf{P} = \{p_{ij}\}_{i,j\in E}$, where

$$p_{ij} = P(X_{n+1} = j \mid X_n = i),$$

is called the *transition matrix* of the HMC. Since the entries are probabilities, and since a transition from any state i must be to some state, it follows that

$$p_{ij} \geq 0, \text{ and } \sum_{k\in E} p_{ik} = 1$$

for all states i, j. A matrix $\mathbf{P}$ indexed by E and satisfying the above properties is called a *stochastic matrix*. The state space may be infinite, and therefore such a matrix is in general not of the kind studied in linear algebra. However, the basic operations of addition and multiplication will be defined by the same formal rules. The notation $x = \{x(i)\}_{i\in E}$ formally represents a column vector, and x^T is the corresponding row vector.

The Markov property easily extends (Exercise 1.5.2) to

$$P(A \mid X_n = i, B) = P(A \mid X_n = i)\,,$$

where

$$A = \{X_{n+1} = j_1, \dots, X_{n+k} = j_k\}, B = \{X_0 = i_0, \dots, X_{n-1} = i_{n-1}\}.$$

This is in turn equivalent to

$$P(A \cap B \mid X_n = i) = P(A \mid X_n = i)P(B \mid X_n = i).$$

That is, A and B are conditionaly independent given $X_n = i$. In other words, the future at time n and the past at time n are conditionally independent given the present state $X_n = i$. In particular, the Markov property is independent of the direction of time.

Notation. We shall from now on abbreviate $P(A \mid X_0 = i)$ as $P_i(A)$. Also, if μ is a probability distribution on E, then $P_\mu(A)$ is the probability of A given that the initial state X_0 is distributed according to μ.

The distribution at time n of the chain is the vector $\nu_n := \{\nu_n(i)\}_{i \in E}$, where

$$\nu_n(i) := P(X_n = i).$$

From the Bayes rule of exclusive and exhaustive causes, $\nu_{n+1}(j) = \sum_{i \in E} \nu_n(i) p_{ij}$, that is, in matrix form, $\nu_{n+1}^T = \nu_n^T \mathbf{P}$. Iteration of this equality yields

$$\nu_n^T = \nu_0^T \mathbf{P}^n. \tag{1.1}$$

The matrix $\mathbf{P}^m$ is called the *m-step transition matrix* because its general term is

$$p_{ij}(m) = P(X_{n+m} = j \mid X_n = i).$$

In fact, by the Bayes sequential rule and the Markov property, the right-hand side equals $\sum_{i_1, \dots, i_{m-1} \in E} p_{ii_1} p_{i_1 i_2} \cdots p_{i_{m-1} j}$, which is the general term of the m-th power of $\mathbf{P}$.

The probability distribution ν_0 of the *initial state* X_0 is called the *initial distribution*. From the Bayes sequential rule and in view of the homogeneous Markov property and the definition of the transition matrix,

$$P(X_0 = i_0, X_1 = i_1, \dots, X_k = i_k) = \nu_0(i_0) p_{i_0 i_1} \cdots p_{i_{k-1} i_k}.$$

Therefore,

Theorem 1.1.1 *The distribution of a discrete-time* HMC *is uniquely determined by its initial distribution and its transition matrix.*

Sample path realization

Many HMC's receive a natural description in terms of a recurrence equation.

Theorem 1.1.2 *Let $\{Z_n\}_{n \geq 1}$ be an* IID *sequence of random variables with values in an arbitrary space F. Let E be a countable space, and $f : E \times F \to E$ be some function. Let X_0 be a random variable with values in E, independent of $\{Z_n\}_{n \geq 1}$. The recurrence equation*

$$X_{n+1} = f(X_n, Z_{n+1}) \tag{1.2}$$

then defines a HMC.

Proof. Iteration of recurrence (1.2) shows that for all $n \geq 1$, there is a function g_n such that $X_n = g_n(X_0, Z_1, \dots, Z_n)$, and therefore $P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(f(i, Z_{n+1}) = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(f(i, Z_{n+1}) = j)$, since the event $\{X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i\}$ is expressible in terms of $X_0, Z_1, \dots, Z_n$ and is therefore independent of Z_{n+1}. Similarly, $P(X_{n+1} = j \mid X_n = i) = P(f(i, Z_{n+1}) = j)$. We therefore have a Markov chain, and it is homogeneous since the right-hand side of the last equality does not depend on n. Explicitly:

$$p_{ij} = P(f(i, Z_1) = j)\,. \tag{1.3}$$

□

EXAMPLE 1.1.1: 1-D RANDOM WALK, TAKE 1. Let X_0 be a random variable with values in $\mathbb{Z}$. Let $\{Z_n\}_{n\geq 1}$ be a sequence of IID random variables, independent of X_0, taking the values $+1$ or -1, and with the probability distribution

$$P(Z_n = +1) = p,$$

where $p \in (0, 1)$. The process $\{X_n\}_{n\geq 1}$ defined by

$$X_{n+1} = X_n + Z_{n+1}$$

is, in view of Theorem 1.1.2, an HMC, called a *random walk* on $\mathbb{Z}$. It is called a "symmetric" random walk if $p = \frac{1}{2}$.

Not all homogeneous Markov chains receive a "natural" description of the type featured in Theorem 1.1.2. However, it is always possible to find a "theoretical" description of the kind. More exactly,

Theorem 1.1.3 *For any transition matrix* $\mathbf{P}$ *on* E*, there exists a homogeneous Markov chain with this transition matrix and with a representation such as in Theorem 1.1.2.*

Proof. Define

$$X_{n+1} := j \text{ if } \sum_{k=0}^{j-1} p_{X_n k} \leq Z_{n+1} < \sum_{k=0}^{j} p_{X_n k}\,,$$

where $\{Z_n\}_{n\geq 1}$ is IID, uniform on $[0, 1]$. By application of Theorem 1.1.2 and of formula (1.3), we check that this HMC has the announced transition matrix. □

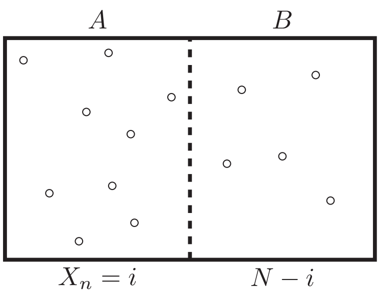
As we already mentioned, not all homogeneous Markov chains are *naturally* described by the model of Theorem 1.1.2. A slight modification of this result considerably enlarges its scope.

Theorem 1.1.4 *Let things be as in Theorem 1.1.2 except for the joint distribution of* $X_0, Z_1, Z_2, \ldots$. *Suppose instead that for all* $n \geq 0$, Z_{n+1} *is conditionally independent of* $Z_n, \ldots, Z_1, X_{n-1}, \ldots, X_0$ *given* X_n, *and that for all* $i, j \in E$, $P(Z_{n+1} = k \mid X_n = i)$ *is independent of* n. *Then* $\{X_n\}_{n\geq 0}$ *is a* HMC, *with transition probabilities*

$$p_{ij} = P(f(i, Z_1) = j \mid X_0 = i).$$

Proof. The proof is quite similar to that of Theorem 1.1.2 (Exercise **??**). □

EXAMPLE 1.1.2: THE EHRENFEST URN, TAKE 1. This idealized model of diffusion through a porous membrane, proposed in 1907 by the Austrian physicists Tatiana and Paul Ehrenfest to describe in terms of statistical mechanics the exchange of heat between two systems at different temperatures, considerably helped understanding the phenomenon of thermodynamic irreversibility (see Example **??**). It features N particles that can be either in compartment A or in compartment B.

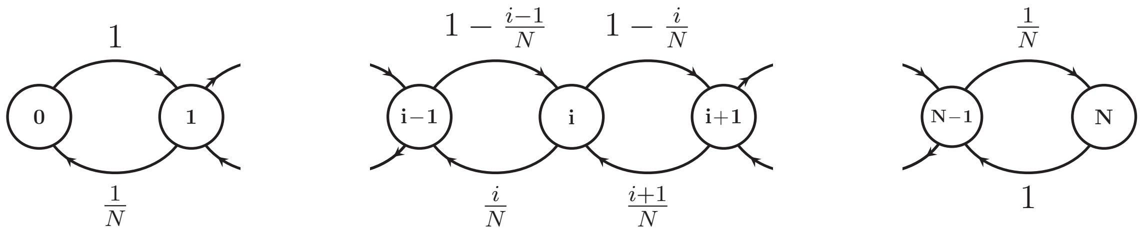


Suppose that at time $n \geq 0$, $X_n = i$ particles are in A. One then chooses a particle at random, and this particle is moved at time $n+1$ from where it is to the other compartment. Thus, the next state X_{n+1} is either $i - 1$ (the displaced particle was found in compartment A) with probability $\frac{i}{N}$, or $i + 1$ (it was found in B) with probability $\frac{N-i}{N}$. This model pertains to Theorem 1.1.4. For all $n \geq 0$,

$$X_{n+1} = X_n + Z_{n+1},$$

where $Z_n \in \{-1, +1\}$ and $P(Z_{n+1} = -1 \mid X_n = i) = \frac{i}{N}$. The nonzero entries of the transition matrix are therefore

$$p_{i,i+1} = \frac{N-i}{N}, \quad p_{i,i-1} = \frac{i}{N}.$$



First-step analysis

Some functionals of homogeneous Markov chains such as probabilities of absorption by a closed set (A is called *closed* if $\sum_{j \in A} p_{ij} = 1$ for all $i \in A$) and average times before absorption can be evaluated by a technique called *first-step analysis*.

EXAMPLE 1.1.3: THE GAMBLER'S RUIN, TAKE 1. Two players A and B play "heads or tails", where heads occur with probability $p \in (0,1)$, and the successive outcomes form an IID sequence. Calling X_n the fortune in dollars of player A at time n, then $X_{n+1} = X_n + Z_{n+1}$, where $Z_{n+1} = +1$ (*resp.*, -1) with probability p (*resp.*, $q := 1-p$), and $\{Z_n\}_{n\geq 1}$ is IID. In other words, A bets \$1 on heads at each toss, and B bets \$1 on tails. The respective initial fortunes of A and B are a and b (positive integers). The game ends when a player is ruined, and therefore the process $\{X_n\}_{n\geq 1}$ is a random walk as described in Example 1.1.1, except that it is restricted to $E = \{0, \dots, a, a+1, \dots, a+b = c\}$. The duration of the game is T, the first time n at which $X_n = 0$ or c, and the probability of winning for A is $u(a) = P(X_T = c \mid X_0 = a)$.

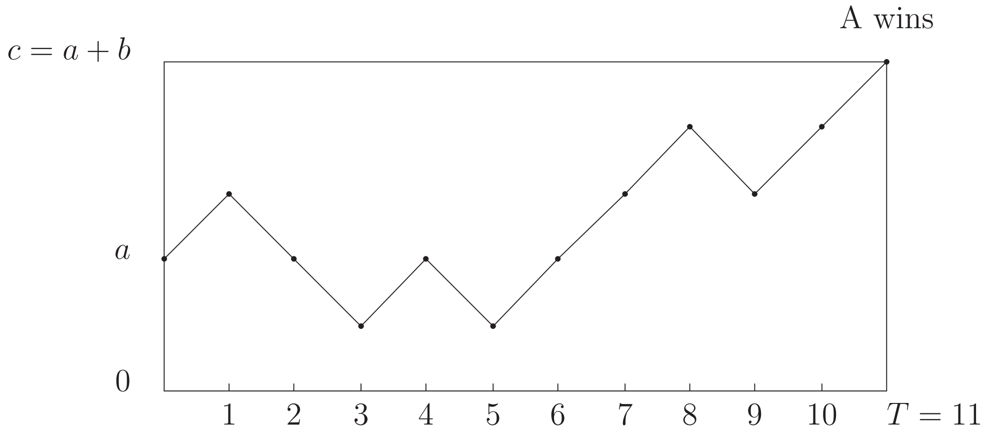
Instead of computing $u(a)$ alone, first-step analysis computes

$$u(i) = P(X_T = c \mid X_0 = i)$$

for all states i, $0 \leq i \leq c$, and for this, it first generates a recurrence equation for $u(i)$ by breaking down event "A wins" according to what can happen after the first step (the first toss) and using the rule of exclusive and exhaustive causes. If $X_0 = i$, $1 \leq i \leq c-1$, then $X_1 = i+1$ (resp., $X_1 = i-1$) with probability p (resp., q), and the probability of winning for A with updated initial fortune $i+1$ (resp., $i-1$) is $u(i+1)$ (resp., $u(i-1)$). Therefore, for i, $1 \leq i \leq c-1$,

$$u(i) = pu(i+1) + qu(i-1),$$

with the boundary conditions $u(0) = 0$, $u(c) = 1$.



The gambler's ruin

The characteristic equation associated with this linear recurrence equation is $pr^2 - r + q = 0$. It has two distinct roots, $r_1 = 1$ and $r_2 = \frac{q}{p}$, if $p \neq \frac{1}{2}$, and a double root, $r_1 = 1$, if $p = \frac{1}{2}$. Therefore, the general solution is $u(i) = \lambda r_1^i + \mu r_2^i = \lambda + \mu \left(\frac{q}{p}\right)^i$ when $p \neq q$, and $u(i) = \lambda r_1^i + \mu i r_1^i = \lambda + \mu i$ when $p = q = \frac{1}{2}$. Taking into account the boundary conditions, one can determine the values of λ and μ. The result is, for $p \neq q$,

$$u(i) = \frac{1 - (\frac{q}{p})^i}{1 - (\frac{q}{p})^c},$$

and for $p = q = \frac{1}{2}$,

$$u(i) = \frac{i}{c}.$$

In the case $p = q = \frac{1}{2}$, the probability $v(i)$ that B wins when the initial fortune of B is $c-i$ is obtained by replacing i by $c-i$ in expression for $u(i)$: $v(i) = \frac{c-i}{c} = 1 - \frac{i}{c}$. One checks that $u(i) + v(i) = 1$, which means in particular that the probability that the game lasts forever is null. The reader is invited to check that the same is true in the case $p \neq q$.

First-step analysis can also be used to compute average times before absorption (Exercise 1.5.5).

1.2 Communication and period

Communication and period are *topological* properties in the sense that they concern only the *naked* transition graph (with only the arrows, without the labels).

Communication and irreducibility

Definition 1.2.1 *State j is said to be accessible from state i if there exists $M \geq 0$ such that $p_{ij}(M) > 0$. States i and j are said to communicate if i is accessible from j and j is accessible from i, and this is denoted by $i \leftrightarrow j$.*

In particular, a state i is always accessible from itself, since $p_{ii}(0) = 1$ ($\mathbf{P}^0 = I$, the identity).

For $M \geq 1$, $p_{ij}(M) = \sum_{i_1,\ldots,i_{M-1}} p_{ii_1} \cdots p_{i_{M-1}j}$, and therefore $p_{ij}(M) > 0$ if and only if there exists at least one path $i, i_1, \ldots, i_{M-1}, j$ from i to j such that

$$p_{ii_1} p_{i_1 i_2} \cdots p_{i_{M-1}j} > 0,$$

or, equivalently, if there is a directed path from i to j in the transition graph G. Clearly,

$$\begin{aligned} i \leftrightarrow i &\qquad \text{(reflexivity)}, \\ i \leftrightarrow j \Rightarrow j \leftrightarrow i &\qquad \text{(symmetry)}, \\ i \leftrightarrow j, j \leftrightarrow k \Rightarrow i \leftrightarrow k &\qquad \text{(transivity)}. \end{aligned}$$

Therefore, the communication relation ($\leftrightarrow$) is an equivalence relation, and it generates a partition of the state space E into disjoint equivalence classes called *communication classes*.

Definition 1.2.2 *A state i such that $p_{ii} = 1$ is called closed. More generally, a set C of states such that for all $i \in C$, $\sum_{j \in C} p_{ij} = 1$ is called closed.*

Definition 1.2.3 *If there exists only one communication class, then the chain, its transition matrix, and its transition graph are said to be irreducible.*

Period

Consider the random walk on $\mathbb{Z}$ (Example 1.1.1). Since $0 < p < 1$, it is irreducible. Observe that $E = C_0 + C_1$, where C_0 and C_1, the set of even and odd relative integers respectively, have the following property. If you start from $i \in C_0$ (resp., C_1), then in one step you can go only to a state $j \in C_1$ (resp., C_0). The chain

$\{X_n\}$ passes alternately from cyclic class to the other. In this sense, the chain has a periodic behavior, corresponding to the period 2. More generally, for any *irreducible* Markov chain, one can find a *unique partition* of E into d classes C_0, C_1, ..., C_{d-1} such that for all $k, i \in C_k$,

$$\sum_{j \in C_{k+1}} p_{ij} = 1,$$

where by convention $C_d = C_0$, and where d is maximal (that is, there is no other such partition C'_0, $C'_1, \dots, C'_{d'-1}$ with $d' > d$). The proof follows directly from Theorem 1.2.2 below.

The number $d \geq 1$ is called the *period* of the chain (resp., of the transition matrix, of the transition graph). The classes $C_0, C_1, \dots, C_{d-1}$ are called the *cyclic classes*.The chain therefore moves from one class to the other at each transition, and this cyclically.

We now give the formal definition of period. It is based on the notion of *greatest common divisor* of a set of positive integers.

Definition 1.2.4 *The period d_i of state $i \in E$ is, by definition,*

$$d_i = \text{GCD}\{n \geq 1\,;\, p_{ii}(n) > 0\},$$

with the convention $d_i = +\infty$ if there is no $n \geq 1$ with $p_{ii}(n) > 0$. If $d_i = 1$, the state i is called aperiodic .

Theorem 1.2.1 *If states i and j communicate they have the same period.*

Proof. As i and j communicate, there exist integers N and M such that $p_{ij}(M) > 0$ and $p_{ji}(N) > 0$. For any $k \geq 1$,

$$p_{ii}(M + nk + N) \geq p_{ij}(M)(p_{jj}(k))^n p_{ji}(N)$$

(indeed, the path $X_0 = i, X_M = j, X_{M+k} = j, \dots, X_{M+nk} = j, X_{M+nk+N} = i$ is just one way of going from i to i in $M + nk + N$ steps). Therefore, for any $k \geq 1$ such that $p_{jj}(k) > 0$, we have $p_{ii}(M + nk + N) > 0$ for all $n \geq 1$. Therefore, d_i divides $M+nk+N$ for all $n \geq 1$, and in particular, d_i divides k. We have therefore shown that d_i divides all k such that $p_{jj}(k) > 0$, and in particular, d_i divides d_j. By symmetry, d_j divides d_i, and therefore, finally, $d_i = d_j$. □

We can therefore speak of the period of a communication class or of an irreducible chain.

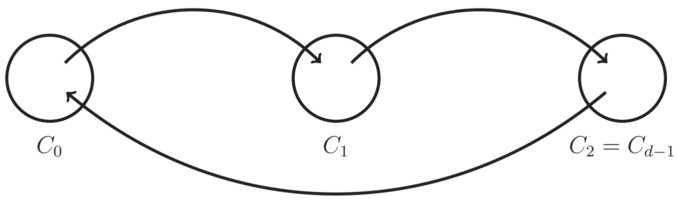
The important result concerning periodicity is the following.

Theorem 1.2.2 *Let* $\mathbf{P}$ *be an irreducible stochastic matrix with period* d*. Then for all states* i, j *there exist* $m \geq 0$ *and* $n_0 \geq 0$ *(*m *and* n_0 *possibly depending on* i, j*) such that*

$$p_{ij}(m + nd) > 0, \quad \textit{for all } n \geq n_0.$$

Proof. It suffices to prove the theorem for $i = j$. Indeed, there exists m such that $p_{ij}(m) > 0$, because j is accessible from i, the chain being irreducible, and therefore, if for some $n_0 \geq 0$ we have $p_{jj}(nd) > 0$ for all $n \geq n_0$, then $p_{ij}(m+nd) \geq p_{ij}(m)p_{jj}(nd) > 0$ for all $n \geq n_0$.

The rest of the proof is an immediate consequence of a classical result of number theory (Theorem A.1.1). Indeed, the GCD of the set $A = \{k \geq 1; p_{jj}(k) > 0\}$ is d, and A is closed under addition. The set A therefore contains all but a finite number of the positive multiples of d. In other words, there exists n_0 such that $n > n_0$ implies $p_{jj}(nd) > 0$. □



Behaviour of a Markov chain with period 3

1.3 Stationarity and reversibility

The central notion of the stability theory of discrete-time HMC's is that of *stationary distribution*.

Definition 1.3.1 *A probability distribution* π *satisfying*

$$\pi^T = \pi^T \mathbf{P} \tag{1.4}$$

is called a stationary distribution of the transition matrix $\mathbf{P}$*, or of the corresponding* HMC.

The *global balance equation* (1.4) says that for all states i,

$$\pi(i) = \sum_{j \in E} \pi(j) p_{ji}.$$

Iteration of (1.4) gives $\pi^T = \pi^T \mathbf{P}^n$ for all $n \geq 0$, and therefore, in view of (1.1), if the initial distribution $\nu = \pi$, then $\nu_n = \pi$ for all $n \geq 0$. Thus, if a chain is started with a stationary distribution, it keeps the same distribution forever. But there is more, because then,

$$\begin{aligned} P(X_n = i_0, X_{n+1} = i_1, \dots, X_{n+k} = i_k) &= P(X_n = i_0) p_{i_0 i_1} \dots p_{i_{k-1} i_k} \\ &= \pi(i_0) p_{i_0 i_1} \dots p_{i_{k-1} i_k} \end{aligned}$$

does not depend on n. In this sense the chain is *stationary*. One also says that the chain is in a *stationary regime*, or in *equilibrium*, or in *steady state*. In summary:

Theorem 1.3.1 *A* HMC *whose initial distribution is a stationary distribution is stationary.*

The balance equation $\pi^T \mathbf{P} = \pi^T$, together with the requirement that π be a probability vector, that is, $\pi^T \mathbf{1} = 1$ (where $\mathbf{1}$ is a column vector with all its entries equal to 1), constitute when E is finite, $|E|+1$ equations for $|E|$ unknown variables. One of the $|E|$ equations in $\pi^T \mathbf{P} = \pi^T$ is superfluous given the constraint $\pi^T \mathbf{1} = 1$. Indeed, summing up all equalities of $\pi^T \mathbf{P} = \pi^T$ yields the equality $\pi^T \mathbf{P}\mathbf{1} = \pi^T \mathbf{1}$, that is, $\pi^T \mathbf{1} = 1$.

EXAMPLE 1.3.1: TWO-STATE MARKOV CHAIN. Take $E = \{1, 2\}$ and define the transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix},$$

where $\alpha, \beta \in (0, 1)$. The global balance equations are

$$\pi(1) = \pi(1)(1-\alpha) + \pi(2)\beta\,, \qquad \pi(2) = \pi(1)\alpha + \pi(2)(1-\beta\,.$$

These two equations are dependent and reduce to the single equation $\pi(1)\alpha = \pi(2)\beta$, to which must be added the constraint $\pi(1) + \pi(2) = 1$ expressing that π is a probability vector. We obtain

$$\pi(1) = \frac{\beta}{\alpha+\beta}, \quad \pi(2) = \frac{\alpha}{\alpha+\beta}\,.$$

EXAMPLE 1.3.2: THE EHRENFEST URN, TAKE 2. The global balance equations are, for $i \in [1, N-1]$,

$$\pi(i) = \pi(i-1)\left(1 - \frac{i-1}{N}\right) + \pi(i+1)\frac{i+1}{N}$$

and, for the boundary states,

$$\pi(0) = \pi(1)\frac{1}{N}, \quad \pi(N) = \pi(N-1)\frac{1}{N}.$$

Leaving $\pi(0)$ undetermined, one can solve the balance equations for $i = 0, 1, \dots, N$ successively, to obtain $\pi(i) = \pi(0)\binom{N}{i}$. The value of $\pi(0)$ is then determined by writing that π is a probability vector: $1 = \sum_{i=0}^{N} \pi(i) = \pi(0) \sum_{i=0}^{N} \binom{N}{i} = \pi(0)2^N$. This gives for π the binomial distribution of size N and parameter $\frac{1}{2}$:

$$\pi(i) = \frac{1}{2^N}\binom{N}{i}.$$

This is the distribution one would obtain by placing independently each particle in the compartments, with probability $\frac{1}{2}$ for each compartment.

Stationary distributions may be many. Take the identity as transition matrix. Then any probability distribution on the state space is a stationary distribution. Also ther may well not exist any stationary distribution. See Exercise 2.5.5.

Reversible chains

The notions of time-reversal and time-reversibility are very productive, as we shall see in several occasions in the sequel.

Let $\{X_n\}_{n\geq 0}$ be an HMC with transition matrix $\mathbf{P}$ and admitting a stationary distribution $\pi > 0$ (meaning $\pi(i) > 0$ for all states i). Define the matrix $\mathbf{Q}$, indexed by E, by

$$\pi(i)q_{ij} = \pi(j)p_{ji}. \tag{1.5}$$

This is a stochastic matrix since

$$\sum_{j\in E} q_{ij} = \sum_{j\in E} \frac{\pi(j)}{\pi(i)} p_{ji} = \frac{1}{\pi(i)} \sum_{j\in E} \pi(j)p_{ji} = \frac{\pi(i)}{\pi(i)} = 1\,,$$

where the third equality uses the global balance equations. Its interpretation is the following: Suppose that the initial distribution of the chain is π, in which case for all $n \geq 0$, all $i \in E$, $P(X_n = i) = \pi(i)$. Then, from Bayes's retrodiction formula,

$$P(X_n = j \mid X_{n+1} = i) = \frac{P(X_{n+1} = i \mid X_n = j)P(X_n = j)}{P(X_{n+1} = i)},$$

that is, in view of (1.5),

$$P(X_n = j \mid X_{n+1} = i) = q_{ji}\,.$$

We see that $\mathbf{Q}$ is the transition matrix of the initial chain when time is reversed.

The following is a very simple observation that will be promoted to the rank of a theorem in view of its usefulness and also for the sake of easy reference.

Theorem 1.3.2 *Let $\mathbf{P}$ be a stochastic matrix indexed by a countable set E, and let π be a probability distribution on E. Define the matrix $\mathbf{Q}$ indexed by E by (1.5). If $\mathbf{Q}$ is a stochastic matrix, then π is a stationary distribution of $\mathbf{P}$.*

Proof. For fixed $i \in E$, sum equalities (1.5) with respect to $j \in E$ to obtain

$$\sum_{j\in E} \pi(i) q_{ij} = \sum_{j\in E} \pi(j) p_{ji}.$$

This is the global balance equation since the left-hand side is equal to $\pi(i) \sum_{j\in E} q_{ij} = \pi(i)$. □

Definition 1.3.2 *One calls* reversible *a stationary Markov chain with initial distribution π (a stationary distribution) if for all $i, j \in E$, we have the so-called* detailed balance equations

$$\pi(i)p_{ij} = \pi(j)p_{ji}. \tag{1.6}$$

We then say: the pair $(\mathbf{P}, \pi)$ is reversible.

In this case, $q_{ij} = p_{ij}$, and therefore the chain and the time-reversed chain are statistically the same, since the distribution of a homogeneous Markov chain is entirely determined by its initial distribution and its transition matrix.

The following is an immediate corollary of Theorem 1.3.2.

Theorem 1.3.3 *Let $\mathbf{P}$ be a transition matrix on the countable state space E, and let π be some probability distribution on E. If for all $i, j \in E$, the detailed balance equations (1.6) are satisfied, then π is a stationary distribution of $\mathbf{P}$.*

EXAMPLE 1.3.3: THE EHRENFEST URN, TAKE 3. The verification of the detailed balance equations $\pi(i)p_{i,i+1} = \pi(i+1)p_{i+1,i}$ is immediate.

1.4 Strong Markov property

The Markov property, that is, the independence of past and future given the present state, extends to the situation where the present time is a *stopping time*, a notion which we now introduce.

Stopping times

Let $\{X_n\}_{n\geq 0}$ be a stochastic process with values in the denumerable set E. For an event A, the notation $A \in \mathcal{X}_0^n$ means that there exists a function $\varphi : E^{n+1} \mapsto \{0,1\}$ such that

$$1_A(\omega) = \varphi(X_0(\omega), \dots, X_n(\omega))\,.$$

In other terms, this event is expressible in terms of $X_0(\omega), \dots, X_n(\omega)$. Let now τ be a random variable with values in $\overline{\overline{\ }}$. It is called a X_0^n-stopping time if for all $m \in \overline{\ }$, $\{\tau = m\} \in X_0^m$. In other words, it is a non-anticipative random time with respect to $\{X_n\}_{n\geq 0}$, since in order to check if $\tau = m$, one needs only observe the process up to time m and not beyond. It is immediate to check that if τ is a X_0^n-stopping time, then so is $\tau + n$ for all $n \geq 1$.

Example 1.4.1: Return time. Let $\{X_n\}_{n\geq 0}$ be an HMC with state space E. Define for $i \in E$ the *return time* to i by

$$T_i := \inf\{n \geq 1\,;\, X_n = i\}$$

using the convention $\inf \varnothing = \infty$ for the empty set of $\overline{\ }$. This is a X_0^n-stopping time since for all $m \in \overline{\ }$,

$$\{T_i = m\} = \{X_1 \neq i, X_2 \neq i, \dots, X_{m-1} \neq i, X_m = i\}\,.$$

Note that $T_i \geq 1$. It is a "return" time, not to be confused with the closely related "hitting" time of i, defined as $S_i := \inf\{n \geq 0\,;\, X_n = i\}$, which is also a X_0^n-stopping time, equal to T_i if and only if $X_0 \neq i$.

Example 1.4.2: Successive return times. This continues the previous example. Let us fix a state, conventionally labeled 0, and let T_0 be the return time to 0. We define the successive return times to 0, τ_k, $k \geq 1$ by $\tau_1 = T_0$ and for $k \geq 1$,

$$\tau_{k+1} := \inf\{n \geq \tau_k + 1\,;\, X_n = 0\}$$

with the above convention that $\inf \varnothing = \infty$. In particular, if $\tau_k = \infty$ for some k, then $\tau_{k+\ell} = \infty$ for all $\ell \geq 1$. The identity

$$\{\tau_k = m\} \equiv \left\{\sum_{n=1}^{m-1} 1_{\{X_n=0\}} = k-1\,,\ X_m = 0\right\}$$

for $m \geq 1$ shows that τ_k is a X_0^n-stopping time.

Let $\{X_n\}_{n\geq 0}$ be a stochastic process with values in the countable set E and let τ be a random time taking its values in $\overline{} := \ \cup\{+\infty\}$. In order to define X_τ when $\tau = \infty$, one must decide how to define X_∞. This is done by taking some arbitrary element Δ not in E, and setting

$$X_\infty = \Delta.$$

By definition, the "process after τ" is the stochastic process

$$\{S_\tau X_n\}_{n\geq 0} := \{X_{n+\tau}\}_{n\geq 0}.$$

The "process before τ," or the "process stopped at τ," is the process

$$\{X_n^\tau\}_{n\geq 0} := \{X_{n\wedge\tau}\}_{n\geq 0},$$

which freezes at time τ at the value X_τ.

Theorem 1.4.1 *Let $\{X_n\}_{n\geq 0}$ be an* HMC *with state space E and transition matrix $\mathbf{P}$. Let τ be a X_0^n-stopping time. Then for any state $i \in E$,*

(α) Given that $X_\tau = i$, the process after τ and the process before τ are independent.

(β) Given that $X_\tau = i$, the process after τ is an HMC *with transition matrix $\mathbf{P}$.*

Proof. (α) We have to show that for all times $k \geq 1, n \geq 0$, and all states $i_0, \dots, i_n, i, j_1, \dots, j_k$,

$$P(X_{\tau+1} = j_1, \dots, X_{\tau+k} = j_k \,|\, X_\tau = i, X_{\tau\wedge 0} = i_0, \dots, X_{\tau\wedge n} = i_n)$$
$$= P(X_{\tau+1} = j_1, \dots, X_{\tau+k} = j_k \,|\, X_\tau = i).$$

We shall prove a simplified version of the above equality, namely

$$P(X_{\tau+k} = j \,|\, X_\tau = i, X_{\tau\wedge n} = i_n) = P(X_{\tau+k} = j \,|\, X_\tau = i)\,. \qquad (\star)$$

The general case is obtained by the same arguments. The left-hand side of $(\star)$ equals

$$\frac{P(X_{\tau+k} = j, X_\tau = i, X_{\tau\wedge n} = i_n)}{P(X_\tau = i, X_{\tau\wedge n} = i_n)}.$$

The numerator of the above expression can be developed as

$$\sum_{r\in} P(\tau = r, X_{r+k} = j, X_r = i, X_{r\wedge n} = i_n)\,. \qquad (\star\star)$$

(The sum is over because $X_\tau = i \neq \Delta$ implies that $\tau < \infty$.) But $P(\tau = r, X_{r+k} = j, X_r = i, X_{r\wedge n} = i_n) = P(X_{r+k} = j \,|\, X_r = i,\ X_{r\wedge n} = i_n, \tau = r)$

$P(\tau = r, X_{r\wedge n} = i_n, X_r = i)$, and since $r \wedge n \le r$ and $\{\tau = r\} \in X_0^r$, the event $B := \{X_{r\wedge n} = i_n, \tau = r\}$ is in X_0^r. Therefore, by the Markov property, $P(X_{r+k} = j \mid X_r = i, X_{r\wedge n} = i_n, \tau = r\} = P(X_{r+k} = j \mid X_r = i) = p_{ij}(k)$. Finally, expression $(\star\star)$ reduces to

$$\sum_{r\in} p_{ij}(k)P(\tau = r, X_{r\wedge n} = i_n, X_r = i) = p_{ij}(k)P(X_{\tau=i}, X_{\tau\wedge n} = i_n).$$

Therefore, the left-hand side of $(\star)$ is just $p_{ij}(k)$. Similar computations show that the right-hand side of $(\star)$ is also $p_{ij}(k)$, so that (α) is proven.

(β) We must show that for all states $i, j, k, i_{n-1}, \dots, i_1$,

$$P(X_{\tau+n+1} = k \mid X_{\tau+n} = j, X_{\tau+n-1} = i_{n-1}, \dots, X_\tau = i) \\ = P(X_{\tau+n+1} = k \mid X_{\tau+n} = j) = p_{jk}.$$

But the first equality follows from the fact proven in (α) that for the stopping time $\tau' = \tau + n$, the processes before and after τ' are independent given $X_{\tau'} = j$. The second equality is obtained by the same calculations as in the proof of (α). □

The cycle independence property

Consider a Markov chain with a state conventionally denoted by 0 such that $P_0(T_0 < \infty) = 1$. In view of the strong Markov property, the chain starting from state 0 will return infinitely often to this state. Let $\tau_1 = T_0, \tau_2, \dots$ be the successive return times to 0, and set $\tau_0 \equiv 0$.

By the strong Markov property, for any $k \ge 1$, the process after τ_k is independent of the process before τ_k (observe that condition $X_{\tau_k} = 0$ is always satisfied), and the process after τ_k is a Markov chain with the same transition matrix as the original chain, and with initial state 0, by construction. Therefore, the successive times of visit to 0, the pieces of trajectory

$$\{X_{\tau_k}, X_{\tau_k+1}, \dots, X_{\tau_{k+1}-1}\}, k \ge 0,$$

are independent and identically distributed. Such pieces are called the *regenerative cycles* of the chain between visits to state 0. Each random time τ_k is a *regeneration time*, in the sense that $\{X_{\tau_k+n}\}_{n\ge 0}$ is independent of the past $X_0, \dots, X_{\tau_k-1}$ and has the same distribution as $\{X_n\}_{n\ge 0}$. In particular, the sequence $\{\tau_k - \tau_{k-1}\}_{k\ge 1}$ is IID.

1.5 Exercises

Exercise 1.5.1. A COUNTEREXAMPLE.
The Markov property does not imply that the past and the future are independent given *any* information concerning the present. Find a simple example of an HMC $\{X_n\}_{n\geq 0}$ with state space $E = \{1, 2, 3, 4, 5, 6\}$ such that

$$P(X_2 = 6 \,|\, X_1 \in \{3, 4\}, X_0 = 2) \neq P(X_2 = 6 \,|\, X_1 \in \{3, 4\}).$$

Exercise 1.5.2. PAST, PRESENT, FUTURE.
For an HMC $\{X_n\}_{n\geq 0}$ with state space E, prove that for all $n \in \mathbb{N}$, and all states $i_0, i_1, \dots, i_{n-1}, i, j_1, j_2, \dots, j_k \in E$,

$$P(X_{n+1} = j_1, \dots, X_{n+k} = j_k \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0)$$
$$= P(X_{n+1} = j_1, \dots, X_{n+k} = j_k \mid X_n = i).$$

Exercise 1.5.3.
Let $\{X_n\}_{n\geq 0}$ be a HMC with state space E and transition matrix $\mathbf{P}$. Show that for all $n \geq 1$, all $k \geq 2$, X_n is conditionally independent of $X_0, \dots, X_{n-2}, X_{n+2}, \dots, X_{n+k}$ given X_{n-1}, X_{n+1} and compute the conditional distribution of X_n given X_{n-1}, X_{n+1}.

Exercise 1.5.4. STREETGANGS.
Three characters, A, B, and C, armed with guns, suddenly meet at the corner of a Washington D.C. street, whereupon they naturally start shooting at one another. Each street-gang kid shoots every tenth second, as long as he is still alive. The probability of a hit for A, B, and C are α, β, and γ respectively. A is the most hated, and therefore, as long as he is alive, B and C ignore each other and shoot at A. For historical reasons not developed here, A cannot stand B, and therefore he shoots only at B while the latter is still alive. Lucky C is shot at if and only if he is in the presence of A alone or B alone. What are the survival probabilities of A, B, and C, respectively?

Exercise 1.5.5. THE GAMBLER'S RUIN.
(This exercise continues Example 1.1.3.) Compute the average duration of the game when $p = \frac{1}{2}$.

Exercise 1.5.6. RECORDS.
Let $\{Z_n\}_{n\geq 1}$ be an IID sequence of geometric random variables: For $k \geq 0$, $P(Z_n = k) = (1-p)^k p$, where $p \in (0, 1)$. Let $X_n = \max(Z_1, \dots, Z_n)$ be the *record value*

at time n, and suppose X_0 is an $\mathbb{N}$-valued random variable independent of the sequence $\{Z_n\}_{n\geq 1}$. Show that $\{X_n\}_{n\geq 0}$ is an HMC and give its transition matrix.

Exercise 1.5.7. AGGREGATION OF STATES.
Let $\{X_n\}_{n\geq 0}$ be a HMC with state space E and transition matrix $\mathbf{P}$, and let $(A_k, k \geq 1)$ be a countable partition of E. Define the process $\{\hat{X}_n\}_{n\geq 0}$ with state space $\hat{E} = \{\hat{1}, \hat{2}, \dots\}$ by $\hat{X}_n = \hat{k}$ if and only if $X_n \in A_k$. Show that if $\sum_{j\in A_\ell} p_{ij}$ is independent of $i \in A_k$ for all k, ℓ, $\{\hat{X}_n\}_{n\geq 0}$ is a HMC with transition probabilities $\hat{p}_{\hat{k}\hat{\ell}} = \sum_{j\in A_\ell} p_{ij}$ (any $i \in A_k$).

Chapter 2

Recurrence

2.1 The potential matrix criterion

The potential matrix criterion

The distribution given $X_0 = j$ of $N_i = \sum_{n\geq 1} 1_{\{X_n=i\}}$, the number of visits to state i strictly after time 0, is

$$P_j(N_i = r) = f_{ji} f_{ii}^{r-1}(1 - f_{ii})\,(r \geq 1)$$
$$P_j(N_i = 0) = 1 - f_{ji},$$

where $f_{ji} = P_j(T_i < \infty)$ and T_i is the return time to i.

Proof. We first go from j to i (probability f_{ji}) and then, $r-1$ times in succession, from i to i (each time with probability f_{ii}), and the last time, that is the $r+1$-st time, we leave i never to return to it (probability $1-f_{ii}$). By the cycle independence property, all these "cycles" are independent, so that the successive probabilities multiplicate. □

The distribution of N_i given $X_0 = j$ and given $N_i \geq 1$ is geometric. This has two main consequences. Firstly, $P_i(T_i < \infty) = 1 \iff P_i(N_i = \infty) = 1$. In words: if starting from i the chain almost surely returns to i, and will then visit i infinitely often. Secondly,

$$E_i[N_i] = \sum_{r=1}^{\infty} r P_i(N_i = r) = \sum_{r-1}^{\infty} r f_{ii}^r (1 - f_{ii}) = \frac{f_{ii}}{1 - f_{ii}}.$$

In particular, $P_i(T_i < \infty) < 1 \iff E_i[N_i] < \infty$.

We collect these results for future reference. For any state $i \in E$,

$$P_i(T_i < \infty) = 1 \iff P_i(N_i = \infty) = 1$$

and

$$P_i(T_i < \infty) < 1 \iff P_i(N_i = \infty) = 0 \iff E_i[N_i] < \infty. \tag{2.1}$$

In particular, the event $\{N_i = \infty\}$ has P_i-probability 0 or 1.

The *potential matrix* $\mathbf{G}$ associated with the transition matrix $\mathbf{P}$ is defined by

$$\mathbf{G} = \sum_{n \geq 0} \mathbf{P}^n.$$

Its general term

$$g_{ij} = \sum_{n=0}^{\infty} p_{ij}(n) = \sum_{n=0}^{\infty} P_i(X_n = j) = \sum_{n=0}^{\infty} E_i[1_{\{X_n=j\}}] = E_i\left[\sum_{n=0}^{\infty} 1_{\{X_n=j\}}\right]$$

is the average number of visits to state j, given that the chain starts from state i.

Recall that T_i denotes the *return* time to state i.

Definition 2.1.1 *State $i \in E$ is called* recurrent *if*

$$P_i(T_i < \infty) = 1,$$

and otherwise it is called transient. *A recurrent state $i \in E$ such that*

$$E_i[T_i] < \infty,$$

is called positive *recurrent , and otherwise it is called* null *recurrent.*

Although the next criterion of recurrence is of theoretical rather than practical interest, it can be helpful in a few situations, for instance in the study of recurrence of random walks (see the examples below).

Theorem 2.1.1 *State $i \in E$ is recurrent if and only if*

$$\sum_{n=0}^{\infty} p_{ii}(n) = \infty.$$

Proof. This merely rephrases Eqn. (2.1). □

EXAMPLE 2.1.1: 1-D RANDOM WALK. The state space of this Markov chain is $E :=$ and the non-null terms of its transition matrix are $p_{i,i+1} = p$, $p_{i,i-1} = 1-p$, where $p \in (0, 1)$. Since this chain is irreducible, it suffices to elucidate the nature (recurrent or transient) of any one of its states, say, 0. We have $p_{00}(2n + 1) = 0$ and

$$p_{00}(2n) = \frac{(2n)!}{n!n!} p^n (1-p)^n.$$

By Stirling's equivalence formula $n! \sim (n/e)^n \sqrt{2\pi n}$, the above quantity is equivalent to

$$\frac{[4p(1-p)]^n}{\sqrt{\pi n}} \tag{$\star$}$$

and the nature of the series $\sum_{n=0}^{\infty} p_{00}(n)$ (convergent or divergent) is that of the series with general term ($\star$). If $p \neq \frac{1}{2}$, in which case $4p(1-p) < 1$, the latter series converges, and if $p = \frac{1}{2}$, in which case $4p(1-p) = 1$, it diverges. In summary, the states of the 1-D random walk are transient if $p \neq \frac{1}{2}$, recurrent if $p = \frac{1}{2}$.

EXAMPLE 2.1.2: 3-D RANDOM WALK. The state space of this HMC is $E = \mathbb{Z}^3$. Denoting by e_1, e_2, and e_3 the canonical basis vectors of $\mathbb{R}^3$ (respectively $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$), the nonnull terms of the transition matrix of the 3-D symmetric random walk are given by

$$p_{x,x\pm e_i} = \frac{1}{6}.$$

We elucidate the nature of state, say, $0 = (0, 0, 0)$. Clearly, $p_{00}(2n + 1) = 0$ for all $n \geq 0$, and (exercise)

$$p_{00}(2n) = \sum_{0 \leq i+j \leq n} \frac{(2n)!}{(i!j!(n-i-j)!)^2} \left(\frac{1}{6}\right)^{2n}.$$

This can be rewritten as

$$p_{00}(2n) = \sum_{0 \leq i+j \leq n} \frac{1}{2^{2n}} \binom{2n}{n} \left(\frac{n!}{i!j!(n-i-j)!}\right)^2 \left(\frac{1}{3}\right)^{2n}.$$

Using the *trinomial formula*

$$\sum_{0 \leq i+j \leq n} \frac{n!}{i!j!(n-i-j)!} \left(\frac{1}{3}\right)^n = 1,$$

we obtain the bound

$$p_{00}(2n) \le K_n \frac{1}{2^{2n}} \binom{2n}{n} \left(\frac{1}{3}\right)^n ,$$

where

$$K_n = \max_{0 \le i+j \le n} \frac{n!}{i!j!(n-i-j)!}.$$

For large values of n, K_n is bounded as follows. Let i_0 and j_0 be the values of i, j that maximize $n!/(i!j!(n+-i-j)!)$ in the domain of interest $0 \le i+j \le n$. From the definition of i_0 and j_0, the quantities

$$\frac{n!}{(i_0-1)!j_0!(n-i_0-j_0+1)!,}$$

$$\frac{n!}{(i_0+1)!j_0!(n-i_0-j_0-1)!,}$$

$$\frac{n!}{i_0!(j_0-1)!(n-i_0-j_0+1)!,}$$

$$\frac{n!}{i_0!(j_0+1)!(n-i_0-j_0-1)!},$$

are bounded by

$$\frac{n!}{i_0!j_0!(n-i_0-j_0)!}.$$

The corresponding inequalities reduce to

$$n - i_0 - 1 \le 2j_0 \le n - i_0 + 1 \text{ and } n - j_0 - 1 \le 2i_0 \le n - j_0 + 1,$$

and this shows that for large n, $i_0 \sim n/3$ and $j_0 \sim n/3$. Therefore, for large n,

$$p_{00}(2n) \sim \frac{n!}{(n/3)!(n/3)!2^{2n}e^n} \binom{2n}{n}.$$

By Stirling's equivalence formula, the right-hand side of the latter equivalence is in turn equivalent to

$$\frac{3\sqrt{3}}{2(\pi n)^{3/2}},$$

the general term of a convergent series. State 0 is therefore transient.

One might wonder at this point about the symmetric random walk on ^{2}, which moves at each step northward, southward, eastward and westward equiprobably. Exercise **??** asks you to show that it is null recurrent. Exercise **??** asks you to prove that the symmetric random walk on ^{p}, $p \ge 4$ are transient.

A theoretical application of the potential matrix criterion is to the proof that recurrence is a (communication) class property.

Theorem 2.1.2 *If i and j communicate, they are either both recurrent or both transient.*

Proof. By definition, i and j communicate if and only if there exist integers M and N such that $p_{ij}(M) > 0$ and $p_{ji}(N) > 0$. Going from i to j in M steps, then from j to j in n steps, then from j to i in N steps, is just one way of going from i back to i in $M + n + N$ steps. Therefore, $p_{ii}(M + n + N) \geq p_{ij}(M) \times p_{jj}(n) \times p_{ji}(N)$. Similarly, $p_{jj}(N + n + M) \geq p_{ji}(N) \times p_{ii}(n) \times p_{ij}(M)$. Therefore, with $\alpha := p_{ij}(M)\, p_{ji}(N)$ (a strictly positive quantity), we have $p_{ii}(M + N + n) \geq \alpha\, p_{jj}(n)$ and $p_{jj}(M + N + n) \geq \alpha\, p_{ii}(n)$. This implies that the series $\sum_{n=0}^{\infty} p_{ii}(n)$ and $\sum_{n=0}^{\infty} p_{jj}(n)$ either both converge or both diverge. The potential matrix criterion concludes the proof. □

2.2 Stationary distribution criterion

Invariant measure

The notion of invariant measure plays an important technical role in the recurrence theory of Markov chains. It extends the notion of stationary distribution.

Definition 2.2.1 *A non-trivial (that is, non-null) vector x (indexed by E) of non-negative real numbers (notation: $0 \leq x < \infty$) is called an* invariant measure *of the stochastic matrix $\mathbf{P}$ (indexed by E) if*

$$x^T = x^T \mathbf{P} . \tag{2.2}$$

Theorem 2.2.1 *Let $\mathbf{P}$ be the transition matrix of an irreducible recurrent* HMC *$\{X_n\}_{n\geq 0}$. Let 0 be an arbitrary state and let T_0 be the return time to 0. Define for all $i \in E$*

$$x_i = E_0 \left[\sum_{n=1}^{T_0} 1_{\{X_n = i\}} \right] . \tag{2.3}$$

(For $i \neq 0$, x_i is the expected number of visits to state i before returning to 0). Then, $0 < x < \infty$ and x is an invariant measure of $\mathbf{P}$.

Proof. We make three preliminary observations. First, it will be convenient to rewrite (2.3) as

$$x_i = E_0\left[\sum_{n\geq 1} 1_{\{X_n=i\}}1_{\{n\leq T_0\}}\right].$$

Next, when $1 \leq n \leq T_0$, $X_n = 0$ if and only if $n = T_0$. Therefore,

$$x_0 = 1.$$

Also, $\sum_{i\in E}\sum_{n\geq 1} 1_{\{X_n=i\}}1_{\{n\leq T_0\}} = \sum_{n\geq 1}\left(\sum_{i\in E} 1_{\{X_n=i\}}\right)1_{\{n\leq T_0\}} = \sum_{n\geq 1} 1_{\{n\leq T_0\}} = T_0$, and therefore

$$\sum_{i\in E} x_i = E_0[T_0]. \tag{2.4}$$

We introduce the quantity

$${}_0p_{0i}(n) := E_0[1_{\{X_n=i\}}1_{\{n\leq T_0\}}] = P_0(X_1 \neq 0, \cdots, X_{n-1} \neq 0, X_n = i).$$

This is the probability, starting from state 0, of visiting i at time n before returning to 0. From the definition of x,

$$x_i = \sum_{n\geq 1} {}_0p_{0i}(n)\,. \tag{†}$$

We first prove (2.2). Observe that ${}_0p_{0i}(1) = p_{0i}$, and, by first-step analysis, for all $n \geq 2$, ${}_0p_{0i}(n) = \sum_{j\neq 0} {}_0p_{0j}(n-1)p_{ji}$. Summing up all the above equalities, and taking (†) into account, we obtain

$$x_i = p_{0i} + \sum_{j\neq 0} x_j p_{ji},$$

that is, (2.2), since $x_0 = 1$.

Next we show that $x_i > 0$ for all $i \in E$. Indeed, iterating (2.2), we find $x^T = x^T\mathbf{P}^n$, that is, since $x_0 = 1$,

$$x_i = \sum_{j\in E} x_j p_{ji}(n) = p_{0i}(n) + \sum_{j\neq 0} x_j p_{ji}(n).$$

If x_i were null for some $i \in E$, $i \neq 0$, the latter equality would imply that $p_{0i}(n) = 0$ for all $n \geq 0$, which means that 0 and i do not communicate, in contradiction to the irreducibility assumption.

It remains to show that $x_i < \infty$ for all $i \in E$. As before, we find that

$$1 = x_0 = \sum_{j\in E} x_j p_{j0}(n)$$

for all $n \geq 1$, and therefore if $x_i = \infty$ for some i, necessarily $p_{i0}(n) = 0$ for all $n \geq 1$, and this also contradicts irreducibility. □

Theorem 2.2.2 *The invariant measure of an irreducible recurrent* HMC *is unique up to a multiplicative factor.*

Proof. In the proof of Theorem 2.2.1, we showed that for an invariant measure y of an irreducible chain, $y_i > 0$ for all $i \in E$, and therefore, one can define, for all $i, j \in E$, the matrix $\mathbf{Q}$ by

$$q_{ji} = \frac{y_i}{y_j} p_{ij} \,. \qquad (\star)$$

It is a transition matrix, since $\sum_{i \in E} q_{ji} = \frac{1}{y_j} \sum_{i \in E} y_i p_{ij} = \frac{y_j}{y_j} = 1$. The general term of $\mathbf{Q}^n$ is

$$q_{ji}(n) = \frac{y_i}{y_j} p_{ij}(n) \,. \qquad (\star\star)$$

Indeed, supposing $(\star\star)$ true for n,

$$\begin{aligned} q_{ji}(n+1) &= \sum_{k \in E} q_{jk} q_{ki}(n) = \sum_{k \in E} \frac{y_k}{y_j} p_{kj} \frac{y_i}{y_k} p_{ik}(n) \\ &= \frac{y_i}{y_j} \sum_{k \in E} p_{ik}(n) p_{kj} = \frac{y_i}{y_j} p_{ij}(n+1), \end{aligned}$$

and $(\star\star)$ follows by induction.

Clearly, $\mathbf{Q}$ is irreducible, since $\mathbf{P}$ is irreducible (just observe that $q_{ji}(n) > 0$ if and only if $p_{ij}(n) > 0$ in view of $(\star\star)$). Also, $p_{ii}(n) = q_{ii}(n)$, and therefore $\sum_{n \geq 0} q_{ii}(n) = \sum_{n \geq 0} p_{ii}(n)$, and therefore $\mathbf{Q}$ is recurrent by the potential matrix criterion. Call $g_{ji}(n)$ the probability, relative to the chain governed by the transition matrix $\mathbf{Q}$, of returning to state i for the first time at step n when starting from j. First-step analysis gives

$$g_{i0}(n+1) = \sum_{j \neq 0} q_{ij} g_{j0}(n) \,,$$

that is, using $(\star)$,

$$y_i g_{i0}(n+1) = \sum_{j \neq 0} (y_j g_{j0}(n)) p_{ji}.$$

Recall that ${}_0p_{0i}(n+1) = \sum_{j \neq 0} {}_0p_{0j}(n) p_{ji}$, or, equivalently,

$$y_0 \, {}_0p_{0i}(n+1) = \sum_{j \neq 0} (y_0 \, {}_0p_{0j}(n)) p_{ji}.$$

We therefore see that the sequences $\{y_0\ {}_0p_{0i}(n)\}$ and $\{y_i g_{i0}(n)\}$ satisfy the same recurrence equation. Their first terms ($n = 1$), respectively $y_0\ {}_0p_{0i}(1) = y_0 p_{0i}$ and $y_i g_{i0}(1) = y_i q_{i0}$, are equal in view of ($\star$). Therefore, for all $n \geq 1$,

$$ {}_0p_{0i}(n) = \frac{y_i}{y_0} g_{i0}(n). $$

Summing up with respect to $n \geq 1$ and using $\sum_{n\geq 1} g_{i0}(n) = 1$ ($\mathbf{Q}$ is recurrent), we obtain that $x_i = \frac{y_i}{y_0}$. □

Equality (2.4) and the definition of positive recurrence give the following.

Theorem 2.2.3 *An irreducible recurrent* HMC *is positive recurrent if and only if its invariant measures x satisfy*

$$ \sum_{i\in E} x_i < \infty\,. $$

Stationary distribution criterion of positive recurrence

An HMC may well be irreducible and possess an invariant measure, and yet not be recurrent. The simplest example is the 1-D non-symmetric random walk, which was shown to be transient and yet admits $x_i \equiv 1$ for invariant measure. It turns out, however, that the existence of a stationary probability distribution is necessary and sufficient for an irreducible chain (not a priori assumed recurrent) to be recurrent positive.

Theorem 2.2.4 *An irreducible* HMC *is positive recurrent if and only if there exists a stationary distribution. Moreover, the stationary distribution π is, when it exists, unique, and $\pi > 0$.*

Proof. The direct part follows from Theorems 2.2.1 and 2.2.3. For the converse part, assume the existence of a stationary distribution π. Iterating $\pi^T = \pi^T\mathbf{P}$, we obtain $\pi^T = \pi^T\mathbf{P}^n$, that is, for all $i \in E$, $\pi(i) = \sum_{j\in E} \pi(j)p_{ji}(n)$. If the chain were transient, then, for all states i, j,

$$ \lim_{n\uparrow\infty} p_{ji}(n) = 0\,. $$

The following is a formal proof[1]:

$$\begin{aligned}\sum_{n\geq 1} p_{ji}(n) &= \sum_{n\geq 1}\sum_{k\geq 1} P_j(T_i = k)p_{ii}(n-k)\\ &= \sum_{k\geq 1} P_j(T_i = k)\sum_{n\geq 1} p_{ii}(n-k)\\ &\leq \left(\sum_{k\geq 1} P_j(T_i = k)\right)\left(\sum_{n\geq 1} p_{ii}(n)\right)\\ &= P_j(T_i < \infty)\left(\sum_{n\geq 1} p_{ii}(n)\right) \leq \sum_{n\geq 1} p_{ii}(n) < \infty\,.\end{aligned}$$

In particular, $\lim_n p_{ji}(n) = 0$. Since $p_{ji}(n)$ is bounded uniformly in j and n by 1 , by dominated convergence (Theorem A.2.1):

$$\pi(i) = \lim_{n\uparrow\infty}\sum_{j\in E}\pi(j)p_{ji}(n) = \sum_{j\in E}\pi(j)\left(\lim_{n\uparrow\infty} p_{ji}(n)\right) = 0.$$

This contradicts the assumption that π is a stationary distribution ($\sum_{i\in E}\pi(i) = 1$). The chain must therefore be recurrent, and by Theorem 2.2.3, it is positive recurrent.

The stationary distribution π of an irreducible positive recurrent chain is unique (use Theorem 2.2.2 and the fact that there is no choice for a multiplicative factor but 1). Also recall that $\pi(i) > 0$ for all $i \in E$ (see Theorem 2.2.1). □

Theorem 2.2.5 *Let π be the unique stationary distribution of an irreducible positive recurrent* HMC*, and let T_i be the return time to state i . Then*

$$\pi(i)E_i[T_i] = 1. \tag{2.5}$$

Proof. This equality is a direct consequence of expression (2.3) for the invariant measure. Indeed, π is obtained by normalization of x: for all $i \in E$,

$$\pi(i) = \frac{x_i}{\sum_{j\in E} x_j},$$

and in particular, for $i = 0$, recalling that $x_0 = 1$ and using (2.4),

$$\pi(0) = \frac{1}{E_0[T_0]}\,.$$

[1]Rather awkward, but using only the elementary tools available.

Since state 0 does not play a special role in the analysis, (2.5) is true for all $i \in E$. □

The situation is extremely simple when the state space is finite.

Theorem 2.2.6 *An irreducible* HMC *with* finite *state space is positive recurrent.*

Proof. We first show recurrence. We have

$$\sum_{j \in E} p_{ij}(n) = 1,$$

and in particular, the limit of the left hand side is 1. If the chain were transient, then, as we saw in the proof of Theorem 2.2.4, for all $i, j \in E$,

$$\lim_{n \uparrow \infty} p_{ij}(n) = 0,$$

and therefore, since the state space is finite

$$\lim_{n \uparrow \infty} \sum_{j \in E} p_{ij}(n) = 0\,,$$

a contradiction. Therefore, the chain is recurrent. By Theorem 2.2.1 it has an invariant measure x. Since E is finite, $\sum_{i \in E} x_i < \infty$, and therefore the chain is positive recurrent, by Theorem 2.2.3. □

2.3 Foster's theorem

The stationary distribution criterion of positive recurrence of an irreducible chain requires solving the balance equations, and this is not always feasible. Therefore one needs more efficient conditions guaranteeing positive recurrence. The following result (*Foster's theorem*) gives a sufficient condition of positive recurrence.

Theorem 2.3.1 *Let the transition matrix* $\mathbf{P}$ *on the countable state space* E *be irreducible and suppose that there exists a function* $h : E \to$ *such that* $\inf_i h(i) > -\infty$ *and*

$$\sum_{k \in E} p_{ik} h(k) < \infty \text{ for all } i \in F, \tag{2.6}$$

$$\sum_{k \in E} p_{ik} h(k) \leq h(i) - \epsilon \text{ for all } i \notin F, \tag{2.7}$$

for some finite *set* F *and some* $\epsilon > 0$. *Then the corresponding* HMC *is positive recurrent.*

Proof. Since $\inf_i h(i) > -\infty$, one may assume without loss of generality that $h \geq 0$, by adding a constant if necessary. Call τ the return time to F, and define $Y_n = h(X_n)1_{\{n<\tau\}}$. Equality (2.7) is just $E[h(X_{n+1}) \mid X_n = i] \leq h(i) - \epsilon$ for all $i \notin F$. For $i \notin F$,

$$\begin{aligned} E_i[Y_{n+1} \mid X_0^n] &= E_i[Y_{n+1}1_{\{n<\tau\}} \mid X_0^n] + E_i(Y_{n+1}1_{\{n\geq\tau\}} \mid X_0^n] \\ &= E_i[Y_{n+1}1_{\{n<\tau\}} \mid X_0^n] \leq E_i[h(X_{n+1})1_{\{n<\tau\}} \mid X_0^n] \\ &= 1_{\{n<\tau\}}E_i[h(X_{n+1}) \mid X_0^n] = 1_{\{n<\tau\}}E_i[h(X_{n+1}) \mid X_n] \\ &\leq 1_{\{n<\tau\}}h(X_n) - \epsilon 1_{\{n<\tau\}}, \end{aligned}$$

where the third *equality* comes from the fact that $1_{\{n<\tau\}}$ is a function of X_0^n, the fourth *equality* is the Markov property, and the last *inequality* is true because P_i-a.s., $X_n \notin F$ on $n < \tau$. Therefore, P_i-a.s., $E_i[Y_{n+1} \mid X_0^n] \leq Y_n - \epsilon 1_{\{n<\tau\}}$, and taking expectations,

$$E_i[Y_{n+1}] \leq E_i[Y_n] - \epsilon P_i(\tau > n).$$

Iterating the above equality, and observing that Y_n is non-negative, we obtain

$$0 \leq E_i[Y_{n+1}] \leq E_i[Y_0] - \epsilon \sum_{k=0}^{n} P_i(\tau > k).$$

But $Y_0 = h(i)$, P_i-a.s., and $\sum_{k=0}^{\infty} P_i(\tau > k) = E_i[\tau]$. Therefore, for all $i \notin F$,

$$E_i[\tau] \leq \epsilon^{-1}h(i).$$

For $j \in F$, first-step analysis yields

$$E_j[\tau] = 1 + \sum_{i\notin F} p_{ji}E_i[\tau].$$

Thus $E_j[\tau] \leq 1+\epsilon^{-1}\sum_{i\notin F} p_{ji}h(i)$, and this quantity is finite in view of assumption (2.6). Therefore, the return time to F starting anywhere in F has finite expectation. Since F is a finite set, this implies positive recurrence in view of the following lemma. □

Lemma 2.3.1 *Let $\{X_n\}_{n\geq0}$ be an irreducible* HMC*, let F be a finite subset of the state space E, and let $\tau(F)$ be the return time to F. If $E_j[\tau(F)] < \infty$ for all $j \in F$, the chain is positive recurrent.*

Proof. Select $i \in F$, and let T_i be the return time of $\{X_n\}$ to i. Let $\tau_1 = \tau(F), \tau_2, \tau_3, \ldots$ be the successive return times to F. It follows from the strong Markov property that $\{Y_n\}_{n\geq0}$ defined by $Y_0 = X_0 = i$ and $Y_n = X_{\tau_n}$ for $n \geq 1$ is an HMC with state space F. Since $\{X_n\}$ is irreducible, so is $\{Y_n\}$. Since F is finite, $\{Y_n\}$ is positive recurrent, and in particular, $E_i[\tilde{T}_i] < \infty$, where $\tilde{T}_i$ is the return time to i of $\{Y_n\}$. Defining $S_0 = \tau_1$ and $S_k = \tau_{k+1} - \tau_k$ for $k \geq 1$, we have

$$T_i = \sum_{k=0}^{\infty} S_k 1_{\{k<\tilde{T}_i\}}\,,$$

and therefore

$$E_i[T_i] = \sum_{k=0}^{\infty} E_i[S_k 1_{\{k<\tilde{T}_i\}}].$$

Now,

$$E_i[S_k 1_{\{k<\tilde{T}_i\}}] = \sum_{\ell \in F} E_i[S_k 1_{\{k<\tilde{T}_i\}} 1_{\{X_{\tau_k}=\ell\}}]\,,$$

and by the strong Markov property applied to $\{X_n\}_{n\geq 0}$ and the stopping time τ_k, and the fact that the event $\{k < \tilde{T}_i\}$ belongs to the past of $\{X_n\}_{n\geq 0}$ at time τ_k,

$$\begin{aligned} E_i[S_k 1_{\{k<\tilde{T}_i\}} 1_{\{X_{\tau_k}=\ell\}}] &= E_i[S_k \mid k < \tilde{T}_i, X_{\tau_k} = \ell] P_i(k < \tilde{T}_i, X_{\tau_k} = \ell) \\ &= E_i[S_k \mid X_{\tau_k} = \ell] P_i(k < \tilde{T}_i, X_{\tau_k} = \ell)\,. \end{aligned}$$

Observing that $E_i[S_k \mid X_{\tau_k} = \ell] = E_\ell[\tau(F)]$, we see that the latter expression is bounded by $(\max_{\ell \in F} E_\ell[\tau(F)])\, P_i(k < \tilde{T}_i, X_{\tau_k} = \ell)$, and therefore

$$E_i[T_i] \leq \left(\max_{\ell\in F} E_\ell(\tau(F))\right) \sum_{k=0}^{\infty} P_i(\tilde{T}_i > k) = \left(\max_{\ell\in F} E_\ell(\tau(F))\right) E_i[\tilde{T}_i] < \infty.$$

□

The function h in Foster's theorem is called a *Lyapunov function* because it plays a role similar to the Lyapunov functions in the stability theory of ordinary differential equations. The corollary below is refered to as *Pakes's lemma*.

Corollary 2.3.1 *Let $\{X_n\}_{n\geq 0}$ be an irreducible* HMC *on $E =$ such that for all $n \geq 0$ and all $i \in E$,*

$$E[X_{n+1} - X_n \mid X_n = i] < \infty$$

and

$$\limsup_{i\uparrow\infty} E[X_{n+1} - X_n \mid X_n = i] < 0. \tag{2.8}$$

Such an HMC *is positive recurrent.*

Proof. Let -2ϵ be the left-hand side of (2.8). In particular, $\epsilon > 0$. By (2.8), for i sufficiently large, say $i > i_0$, $E[X_{n+1} - X_n \mid X_n = i] < -\epsilon$. We are therefore in the conditions of Foster's theorem with $h(i) = i$ and $F = \{i; i \leq i_0\}$. □

Example 2.3.1: A random walk on $\mathbb{N}$. Let $\{Z_n\}_{n\geq 1}$ be an IID sequence of integrable random variables with values in $\mathbb{Z}$ such that

$$E[Z_1] < 0,$$

and define $\{X_n\}_{n\geq 0}$, an HMC with state space $E = \mathbb{N}$, by

$$X_{n+1} = (X_n + Z_{n+1})^+,$$

where X_0 is independent of $\{Z_n\}_{n\geq 1}$. Assume irreducibility (the industrious reader will find the necessary and sufficient condition for this). Here

$$\begin{aligned} E[X_{n+1} - i \mid X_n = i] &= E[(i + Z_{n+1})^+ - i] \\ &= E[-i 1_{\{Z_{n+1} \leq -i\}} + Z_{n+1} 1_{\{Z_{n+1} > -i\}}] \leq E[Z_1 1_{\{Z_1 > -i\}}]. \end{aligned}$$

By dominated convergence, the limit of $E[Z_1 1_{\{Z_1 > -i\}}]$ as i tends to ∞ is $E[Z_1] < 0$ and therefore, by Pakes's lemma, the HMC is positive recurrent.

The following is a Foster-type theorem, only with a negative conclusion.

Theorem 2.3.2 *Let the transition matrix* $\mathbf{P}$ *on the countable state space* E *be irreducible and suppose that there exists a* finite *set* F *and a function* $h : E \to {}_+$ *such that*

$$\text{there exists } j \notin F \text{ such that } h(j) > \max_{i\in F} h(i) \tag{2.9}$$

$$\sup_{i\in E} \sum_{k\in E} p_{ik} |h(k) - h(i)| < \infty, \tag{2.10}$$

$$\sum_{k\in E} p_{ik}(h(k) - h(i)) \leq 0 \text{ for all } i \notin F. \tag{2.11}$$

Then the corresponding HMC *cannot be positive recurrent.*

Proof. Let τ be the return time to F. Observe that

$$h(X_\tau) 1_{\{\tau<\infty\}} = h(X_0) + \sum_{n=0}^{\infty} (h(X_{n+1}) - h(X_n)) \, 1_{\{\tau>n\}}.$$

Now, with $j \notin F$,

$$\begin{aligned} &\sum_{n=0}^{\infty} E_j \left[|h(X_{n+1}) - h(X_n)| \, 1_{\{\tau>n\}} \right] \\ &\quad = \sum_{n=0}^{\infty} E_j \left[E_j \left[|h(X_{n+1}) - h(X_n)| \, |X_0^n \right] 1_{\{\tau>n\}} \right] \\ &\quad = \sum_{n=0}^{\infty} E_j \left[E_j \left[|h(X_{n+1}) - h(X_n)| \, |X_n \right] 1_{\{\tau>n\}} \right] \\ &\quad \leq K \sum_{n=0}^{\infty} P_j(\tau > n) \end{aligned}$$

for some finite positive constant K by (2.10). Therefore, if the chain is positive recurrent, the latter bound is $KE_j[\tau] < \infty$. Therefore

$$\begin{aligned} E_j[h(X_\tau)] &= E_j\left[h(X_\tau)1_{\{\tau<\infty\}}\right] \\ &= h(j) + \sum_{n=0}^{\infty} E_j\left[\left(h(X_{n+1}) - h(X_n)\right)1_{\{\tau>n\}}\right] > h(j), \end{aligned}$$

by (2.11). In view of assumption (2.9), we have $h(j) > \max_{i\in F} h(i) \geq E_j[h(X_\tau)]$, hence a contradiction. The chain therefore cannot be positive recurrent. □

2.4 Examples

Birth-and-death Markov chain

We first define the birth-and-death process with a bounded population. The state space of such a chain is $E = \{0, 1, \ldots, N\}$ and its transition matrix is

$$\mathbf{P} = \begin{pmatrix} r_0 & p_0 & & & & & & \\ q_1 & r_1 & p_1 & & & & & \\ & q_2 & r_2 & p_2 & & & & \\ & & \ddots & & & & & \\ & & & q_i & r_i & p_i & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & q_{N-1} & r_{N-1} & p_{N-1} \\ & & & & & & p_N & r_N \end{pmatrix},$$

where $p_i > 0$ for all $i \in E\backslash\{N\}$, $q_i > 0$ for all $i \in E\backslash\{0\}$, $r_i \geq 0$ for all $i \in E$, and $p_i + q_i + r_i = 1$ for all $i \in E$. The positivity conditions placed on the p_i's and q_i's guarantee that the chain is irreducible. Since the state space is finite, it is positive recurrent (Theorem 2.2.6), and it has a unique stationary distribution. Motivated by the Ehrenfest HMC which is reversible in the stationary state, we make the educated guess that the birth and death process considered has the same property. This will be the case if and only if there exists a probability distribution π on E satisfying the detailed balance equations, that is, such that for all $1 \leq i \leq N$, $\pi(i-1)p_{i-1} = \pi(i)q_i$. Letting $w_0 = 1$ and for all $1 \leq i \leq N$,

$$w_i = \prod_{k=1}^{i} \frac{p_{k-1}}{q_k}$$

we find that

$$\pi(i) = \frac{w_i}{\sum_{j=0}^{N} w_j} \tag{2.12}$$

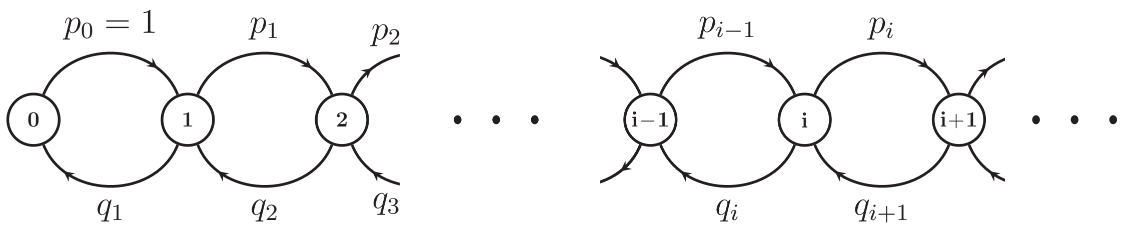
indeed satisfies the detailed balance equations and is therefore the (unique) stationary distribution of the chain.

We now consider the unbounded birth-and-death process. This chain has the state space $E = \mathbb{N}$ and its transition matrix is as in the previous example (only, it is unbounded on the right). In particular, we assume that the p_i's and q_i's are positive in order to guarantee irreducibility. The same reversibility argument as above applies with a little difference. In fact we can show that the w_i's defined above satisfy the detailed balance equations and therefore the global balance equations. Therefore the vector $\{w_i\}_{i\in E}$ the unique, up to a multiplicative factor, invariant measure of the chain. It can be normalized to a probability distribution if and only if

$$\sum_{j=0}^{\infty} w_j < \infty .$$

Therefore, in this case and only in this case there exists a (unique) stationary distribution, also given by (2.12).

Note that the stationary distribution, when it exists, does not depend on the r_i's. The recurrence properties of the above unbounded birth-and-death process are therefore the same as those of the chain below, which is however not aperiodic. For aperiodicity, it suffices to suppose at least one of the r_i's to be positive.



We now compute for the (bounded or unbounded) irreducible birth-and death process, the average time it takes to reach a state b from a state $a < b$. In fact, we shall prove that

$$E_a\left[T_b\right] = \sum_{k=a+1}^{b} \frac{1}{q_k w_k} \sum_{j=0}^{k-1} w_j . \tag{2.13}$$

Since obviously $E_a\left[T_b\right] = \sum_{k=a+1}^{b} E_{k-1}\left[T_k\right]$, it suffices to prove that

$$E_{k-1}\left[T_k\right] = \frac{1}{q_k w_k} \sum_{j=0}^{k-1} w_j . \tag{$\star$}$$

For this, consider for any given $k \in \{0, 1, \dots, N\}$ the truncated chain, which moves on the state space $\{0, 1, \dots, k\}$ as the original chain, except in state k where it moves one step down with probability q_k and stays still with probability $p_k + r_k$. Write $\widetilde{E}$ for expectations of the modified chain. The unique stationary distribution of this chain is given by

$$\widetilde{\pi}_\ell = \frac{w_\ell}{\sum_{j=0}^{k} w_\ell}$$

for all $0 \le \ell \le k$. First-step analysis shows that $\widetilde{E}_k[T_k] = (r_k + p_k) \times 1 + q_k\left(1 + \widetilde{E}_{k-1}[T_k]\right)$, that is

$$\widetilde{E}_k[T_k] = 1 + q_k \widetilde{E}_{k-1}[T_k] \, .$$

Also

$$\widetilde{E}_k[T_k] = \frac{1}{\widetilde{\pi}_k} = \frac{1}{w_k} \sum_{j=0}^{k} w_j \, ,$$

and therefore, since $\widetilde{E}_{k-1}[T_k] = E_{k-1}[T_k]$, we have $(\star)$.

In the special case where $(p_j, q_j, r_j) = (p, q, r)$ for all $j \ne 0, N$, $(p_0, q_0, r_0) = (p, q+r, 0)$ and $(p_N, q_N, r_N) = (0, p+r, q)$, we have $w_i = \left(\frac{p}{q}\right)^i$, and for $1 \le k \le N$,

$$E_{k-1}[T_k] = \frac{1}{q\left(\frac{p}{q}\right)^k} \sum_{j=0}^{k-1} \left(\frac{p}{q}\right)^j = \frac{1}{p-q}\left(1 - \left(\frac{q}{p}\right)^k\right) .$$

In the further particularization where $p = q$, $w_i = 1$ for all i and

$$E_{k-1}[T_k] = \frac{k}{p} \, .$$

The repair shop

During day n, Z_{n+1} machines break down, and they enter the repair shop on day $n+1$. Every day one machine among those waiting for service is repaired. Therefore, denoting by X_n the number of machines in the shop on day n,

$$X_{n+1} = (X_n - 1)^+ + Z_{n+1}, \tag{2.14}$$

where $a^+ = \max(a, 0)$. The sequence $\{Z_n\}_{n \ge 1}$ is assumed to be an IID sequence, independent of the initial state X_0, with common probability distribution

$$P(Z_1 = k) = a_k, \; k \ge 0$$

of generating function g_Z. The stochastic process $\{X_n\}_{n\geq 0}$ is a HMC of transition matrix

$$\mathbf{P} = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

Indeed, by formula (1.3), $p_{ij} = P((i-1)^+ + Z_1 = j) = P(Z_1 = j - (i-1)^+)$. The repair shop model may also be interpreted in terms of communications. It describes a communications link in which time is divided into successive intervals (the "slots") of equal length, conventionally taken to be equal to 1. In slot n (extending from time n included to time $n+1$ excluded), there arrive Z_{n+1} messages requiring transmission. Since the link can transmit at most one message in a given slot, the messages may have to be buffered, and X_n represents the number of messages in the buffer (supposed of infinite capacity) at time n. The dynamics of the buffer content are therefore those of Eqn. (2.14).

A necessary and sufficient condition of irreducibility of this chain is that $P(Z_1 = 0) > 0$ *and* $P(Z_1 \geq 2) > 0$ as we now prove formally. Looking at (2.14), we make the following observations. If $P(Z_{n+1} = 0) = 0$, then $X_{n+1} \geq X_n$ a.s. and there is no way of going from i to $i-1$. If $P(Z_{n+1} \leq 1) = 1$, then $X_{n+1} \leq X_n$ and there is no way of going from i to $i+1$. Therefore, the two conditions $P(Z_1 = 0) > 0$ and $P(Z_2 \geq 2) > 0$ are *necessary* for irreducibility. They are also *sufficient*. Indeed if there exists $k \geq 2$ such that $P(Z_{n+1} = k) > 0$, then one can go form any $i > 0$ to $i + k - 1 > i$ or from $i = 0$ to $k > 0$ with positive probability. Also if $P(Z_{n+1} = 0) > 0$, one can go from $i > 0$ to $i - 1$ with positive probability. In particular, one can go from i to $j < i$ with positive probability. Therefore, to go from i to $j \geq i$, one can take several successive steps of height at least $k-1$, and reach a state $l \geq i$, and then, in the case of $l > i$, go down one by one from l to i. All this with positive probability.

Assuming irreducibility, we now seek a necessary and sufficient condition for positive recurrence. For any complex number z with modulus not larger than 1, it follows from the recurrence equation (2.14) that

$$z^{X_{n+1}+1} = \left(z^{(X_n-1)^+ + 1}\right) z^{Z_{n+1}} = \left(z^{X_n} - 1_{\{X_n=0\}} + z 1_{\{X_n=0\}}\right) z^{Z_{n+1}},$$

and therefore $zz^{X_{n+1}} - z^{X_n} z^{Z_{n+1}} = (z-1)1_{\{X_n=0\}} z^{Z_{n+1}}$. From the independence of X_n and Z_{n+1}, $E[z^{X_n} z^{Z_{n+1}}] = E[z^{X_n}]g_Z(z)$, and $E[1_{\{X_n=0\}} z^{Z_{n+1}}] = \pi(0)g_Z(z)$, where $\pi(0) = P(X_n = 0)$. Therefore, $zE[z^{X_{n+1}}] - g_Z(z)E[z^{X_n}] = (z-1)\pi(0)g_Z(z)$.

But in steady state, $E[z^{X_{n+1}}] = E[z^{X_n}] = g_X(z)$, and therefore

$$g_X(z)\,(z - g_Z(z)) = \pi(0)(z-1)g_Z(z). \tag{2.15}$$

This gives the generating function $g_X(z) = \sum_{i=0}^{\infty} \pi(i)z^i$, as long as $\pi(0)$ is available. To obtain $\pi(0)$, differentiate (2.15): $g'_X(z)\,(z - g_Z(z)) + g_X(z)\,(1 - g'_Z(z)) = \pi(0)\,(g_Z(z) + (z-1)g'_Z(z))$, and let $z = 1$, to obtain, taking into account the equalities $g_X(1) = g_Z(1) = 1$ and $g'_Z(1) = E[Z]$,

$$\pi(0) = 1 - E[Z]. \tag{2.16}$$

But the stationary distribution of an irreducible HMC is positive, hence the necessary condition of positive recurrence:

$$E[Z_1] < 1.$$

We now show this condition is also sufficient for positive recurrence. This follows immediately from Pakes's lemma, since for $i \geq 1$, $E[X_{n+1} - X_n \mid X_n = i] = E[Z] - 1 < 0$.

From (2.15) and (2.16), we have the generating function of the stationary distribution:

$$\sum_{i=0}^{\infty} \pi(i)z^i = (1 - E[Z])\frac{(z-1)g_Z(z)}{z - g_Z(z)}. \tag{2.17}$$

If $E[Z_1] > 1$, the chain is transient, as a simple argument based on the strong law of large numbers shows. In fact, $X_n = X_0 + \sum_{k=1}^{n} Z_k - n + \sum_{k=1}^{n} 1_{\{X_k=0\}}$, and therefore

$$X_n \geq \sum_{k=1}^{n} Z_k - n,$$

which tends to ∞ because, by the strong law of large numbers,

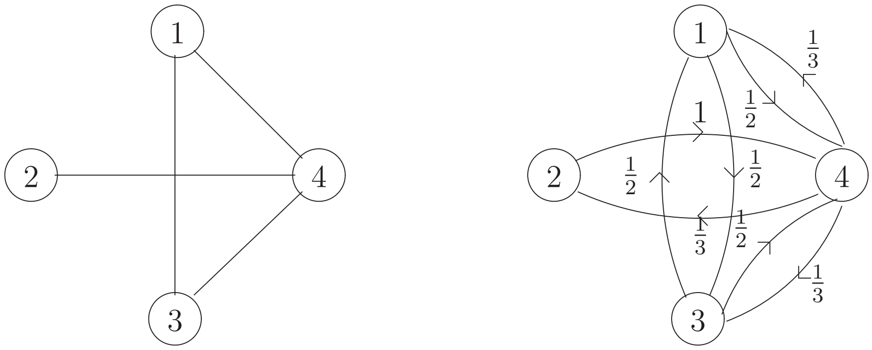
$$\frac{\sum_{k=1}^{n} Z_k - n}{n} \to E[Z] - 1 > 0.$$

This is of course incompatible with recurrence.

We finally examine the case $E[Z_1] = 1$, for which there are only two possibilities left: transient or null recurrent. It turns out that the chain is null recurrent in this case.

The pure random walk on a graph

Consider a finite non-directed connected graph $G = (V, \mathcal{E})$ where V is the set of vertices, or nodes, and $\mathcal{E}$ is the set of edges. Let d_i be the *index* of vertex i (the number of edges "adjacent" to vertex i). Since there is no isolated nodes (a consequence of the connectedness assumption), $d_i > 0$ for all $i \in V$. Transform this graph into a directed graph by splitting each edge into two directed edges of opposite directions, and make it a transition graph by associating to the directed edge from i to j the transition probability $\frac{1}{d_i}$ (see the figure below). Note that $\sum_{i \in V} d_i = 2|\mathcal{E}|$.



A random walk on a graph

The corresponding HMC with state space $E \equiv V$ is irreducible (G is connected). It therefore admits a unique stationary distribution π, that we attempt to find a stationary distribution via Theorem 1.3.3. Let i and j be connected by an edge, and therefore $p_{ij} = \frac{1}{d_i}$ and $p_{ji} = \frac{1}{d_j}$, so that the detailed balance equation between these two states is

$$\pi(i)\frac{1}{d_i} = \pi(j)\frac{1}{d_j}.$$

This gives $\pi(i) = K d_i$, where K is obtained by normalization: $K = \left(\sum_{j \in E} d_j\right)^{-1} = (2|\mathcal{E}|)^{-1}$. Therefore

$$\pi(i) = \frac{d_i}{2|\mathcal{E}|}.$$

EXAMPLE 2.4.1: RANDOM WALK ON THE HYPERCUBE, TAKE 1. The random walk on the (n-dimensional) hypercube is the random walk on the graph with set of vertices $E = \{0, 1\}^n$ and edges between vertices x and y that differ in just one coordinate. For instance, in three dimensions, the only possible motions of a particle performing the random walk on the cube is along its edges in both directions. Clearly, whatever be the dimension $n \geq 2$, $d_i = \frac{1}{n}$ and the stationary distribution is the uniform distribution.

The *lazy random walk* on the graph is, by definition, the Markov chain on V with the transition probabilities $p_{ii} = \frac{1}{2}$ and for $i, j \in V$ such that i and j are connected by an edge of the graph, $p_{i,i} = \frac{1}{2d_i}$. This modified chain admits the same stationary distribution as the original random walk. The difference is that the lazy version is always aperiodic, whereas the original version maybe periodic.

2.5 Exercises

Exercise 2.5.1. Truncated hmc.
Let $\mathbf{P}$ be a transition matrix on the countable state space E, with the positive stationary distribution π. Let A be a subset of the state space, and define the truncation of $\mathbf{P}$ on A to be the transition matrix $\mathbf{Q}$ indexed by A and given by

$$\begin{aligned} q_{ij} &= p_{ij} \text{ if } i, j \in A,\ i \neq j, \\ q_{ii} &= p_{ii} + \sum_{k \in \bar{A}} p_{ik}. \end{aligned}$$

Show that if $(\mathbf{P}, \pi)$ is reversible, then so is $(\mathbf{Q}, \frac{\pi}{\pi(A)})$.

Exercise 2.5.2. Extension to negative times.
Let $\{X_n\}_{n\geq 0}$ be a hmc with state space E, transition matrix $\mathbf{P}$, and suppose that there exists a stationary distribution $\pi > 0$. Suppose moreover that the initial distribution is π. Define the matrix $\mathbf{Q} = \{q_{ij}\}_{i,j\in E}$ by (1.5). Construct $\{X_{-n}\}_{n\geq 1}$, independent of $\{X_n\}_{n\geq 1}$ given X_0, as follows:

$$\begin{aligned} &P(X_{-1} = i_1, X_{-2} = i_2, \ldots, X_{-k} = i_k \mid X_0 = i, X_1 = j_1, \ldots, X_n = j_n) \\ &\quad = P(X_{-1} = i_1, X_{-2} = i_2, \ldots, X_{-k} = i_k \mid X_0 = i) = q_{ii_1} q_{i_1 i_2} \cdots q_{i_{k-1} i_k} \end{aligned}$$

for all $k \geq 1, n \geq 1, i, i_1, \ldots, i_k, j_1, \ldots, j_n \in E$. Prove that $\{X_n\}_{n\in\mathbb{Z}}$ is a hmc with transition matrix $\mathbf{P}$ and $P(X_n = i) = \pi(i)$, for all $i \in E$, all $n \in \mathbb{Z}$.

Exercise 2.5.3. Moving stones.
Stones $S_1, \ldots, S_M$ are placed in line. At each time n a stone is selected at random, and this stone and the one ahead of it in the line exchange positions. If the selected stone is at the head of the line, nothing is changed. For instance, with $M = 5$: Let the current configuration be $S_2S_3S_1S_5S_4$ (S_2 is at the head of the line). If S_5 is selected, the new situation is $S_2S_3S_5S_1S_4$, whereas If S_2 is selected,

the configuration is not altered. At each step, stone S_i is selected with probability $\alpha_i > 0$. Call X_n the situation at time n, for instance $X_n = S_{i_1} \cdots S_{i_M}$, meaning that stone S_{i_j} is in the jth position. Show that $\{X_n\}_{n \geq 0}$ is an irreducible HMC and that it has a stationary distribution given by the formula

$$\pi(S_{i_1} \cdots S_{i_M}) = C\alpha_{i_1}^M \alpha_{i_2}^{M-1} \cdots \alpha_{i_M},$$

for some normalizing constant C.

Exercise 2.5.4. Aperiodicity.
a. Show that an irreducible transition matrix $\mathbf{P}$ with at least one state $i \in E$ such that $p_{ii} > 0$ is aperiodic.

b. Let $\mathbf{P}$ be an irreducible transition matrix on the *finite* state space E. Show that a necessary and sufficient condition for $\mathbf{P}$ to be aperiodic is the existence of an integer m such that $\mathbf{P}^m$ has all its entries positive.

c. Consider a HMC that is irreducible with period $d \geq 2$. Show that the restriction of the transition matrix to any cyclic class is irreducible. Show that the restriction of $\mathbf{P}^d$ to any cyclic class is aperiodic.

Exercise 2.5.5. No stationary distribution.
Show that the symmetric random walk on $\mathbb{Z}$ cannot have a stationary distribution.

Exercise 2.5.6. An interpretation of invariant measure.
A countable number of particles move independently in the countable space E, each according to a Markov chain with the transition matrix $\mathbf{P}$. Let $A_n(i)$ be the number of particles in state $i \in E$ at time $n \geq 0$, and suppose that the random variables $A_0(i)$, $i \in E$, are independent Poisson random variables with respective means $\mu(i), i \in E$, where $\mu = \{\mu(i)\}_{i \in E}$ is an invariant measure of $\mathbf{P}$. Show that for all $n \geq 1$, the random variables $A_n(i)$, $i \in E$, are independent Poisson random variables with respective means $\mu(i), i \in E$.

Exercise 2.5.7. Doubly stochastic transition matrix.
A stochastic matrix $\mathbf{P}$ on the state space E is called *doubly stochastic* if for all states i, $\sum_{j \in E} p_{ji} = 1$. Suppose in addition that $\mathbf{P}$ is irreducible, and that E is *infinite*. Find the invariant measure of $\mathbf{P}$. Show that $\mathbf{P}$ cannot be positive recurrent.

Exercise 2.5.8. Return time to the initial state.
Let τ be the first return time to inital state of an irreducible positive recurrent HMC $\{X_n\}_{n \geq 0}$, that is,

$$\tau = \inf\{n \geq 1; X_n = X_0\},$$

with $\tau = +\infty$ if $X_n \neq X_0$ for all $n \geq 1$. Compute the expectation of τ when the initial distribution is the stationary distribution π. Conclude that it is finite if and only if E is finite. When E is infinite, is this in contradiction to positive recurrence?

Chapter 3

Long-run behaviour

3.1 Ergodic theorem

An important application of the strong law of large numbers is to the ergodic theorem for Markov chains. This theorem gives conditions guaranteeing that empirical averages of the type

$$\frac{1}{N}\sum_{k=1}^{N} g(X_k, \dots, X_{k+L})$$

converge to probabilistic averages. As a matter of fact, if the chain is irreducible positive recurrent with the stationary distribtion π, the above empirical average converges P_μ-almost-surely to $E_\pi[g(X_0, \dots, X_L)]$ for any initial distribution μ (Corollary 3.1.2), at least if $E_\pi[|g(X_0, \dots, X_L)|] < \infty$.

We shall obtain this result as a corollary of the following proposition concerning irreducible recurrent (not necessarily positive recurrent) HMC's.

Let $\{X_n\}_{n\geq 0}$ be an irreducible recurrent HMC, and let x denote the canonical invariant measure associated with state $0 \in E$,

$$x_i = E_0\left[\sum_{n\geq 1} 1_{\{X_n=i\}} 1_{\{n\leq T_0\}}\right], \tag{3.1}$$

where T_0 is the return time to 0. Define for $n \geq 1$, $\nu(n) := \sum_{k=1}^{n} 1_{\{X_k=0\}}$.

Theorem 3.1.1 *Let $f : E \to \mathbb{R}$ be such that*

$$\sum_{i \in E} |f(i)| x_i < \infty. \tag{3.2}$$

Then, for any initial distribution μ, P_μ-a.s.,

$$\lim_{N \uparrow \infty} \frac{1}{\nu(N)} \sum_{k=1}^{N} f(X_k) = \sum_{i \in E} f(i) x_i. \tag{3.3}$$

Proof. Let $T_0 = \tau_1, \tau_2, \tau_3, \ldots$ be the successive return times to state 0, and define

$$U_p = \sum_{n=\tau_p+1}^{\tau_{p+1}} f(X_n).$$

By the independence property of the regenerative cycles, $\{U_p\}_{p \geq 1}$ is an IID sequence. Moreover, assuming $f \geq 0$ and using the strong Markov property,

$$\begin{aligned} E[U_1] &= E_0\left[\sum_{n=1}^{T_0} f(X_n)\right] \\ &= E_0\left[\sum_{n=1}^{T_0} \sum_{i \in E} f(i) 1_{\{X_n = i\}}\right] = \sum_{i \in E} f(i) E_0\left[\sum_{n=1}^{T_0} 1_{\{X_n = i\}}\right] \\ &= \sum_{i \in E} f(i) x_i. \end{aligned}$$

By hypothesis, this quantity is finite, and threfore the strong law of large numbers applies, to give

$$\lim_{n \uparrow \infty} \frac{1}{n} \sum_{p=1}^{n} U_p = \sum_{i \in E} f(i) x_i,$$

that is,

$$\lim_{n \uparrow \infty} \frac{1}{n} \sum_{k=T_0+1}^{\tau_{n+1}} f(X_k) = \sum_{i \in E} f(i) x_i. \tag{3.4}$$

Observing that

$$\tau_{\nu(n)} \leq n < \tau_{\nu(n)+1},$$

we have

$$\frac{\sum_{k=1}^{\tau_{\nu(n)}} f(X_k)}{\nu(n)} \leq \frac{\sum_{k=1}^{n} f(X_k)}{\nu(n)} \leq \frac{\sum_{k=1}^{\tau_{\nu(n)+1}} f(X_i)}{\nu(n)}.$$

Since the chain is recurrent, $\lim_{n\uparrow\infty} \nu(n) = \infty$, and therefore, from (3.4), the extreme terms of the above chain of inequality tend to $\sum_{i\in E} f(i)x_i$ as n goes to ∞, and this implies (3.3). The case of a function f of arbitrary sign is obtained by considering (3.3) written separately for $f^+ = \max(0, f)$ and $f^- = \max(0, -f)$, and then taking the difference of the two equalities obtained this way. The difference is not an undetermined form $\infty - \infty$ due to hypothesis (3.2). □

Corollary 3.1.1 *Let $\{X_n\}_{n\geq 0}$ be an irreducible positive recurrent Markov chain with the stationary distribution π, and let $f : E \to \mathbb{R}$ be such that*

$$\sum_{i\in E} |f(i)|\pi(i) < \infty. \tag{3.5}$$

Then for any initial distribution μ, P_μ-a.s.,

$$\lim_{n\uparrow\infty} \frac{1}{N} \sum_{k=1}^{N} f(X_k) = \sum_{i\in E} f(i)\pi(i). \tag{3.6}$$

Proof. Apply Theorem 3.1.1 to $f \equiv 1$. Condition (3.2) is satisfied, since in the positive recurrent case, $\sum_{i\in E} x_i = E_0[T_0] < \infty$. Therefore, P_μ-a.s.,

$$\lim_{N\uparrow\infty} \frac{N}{\nu(N)} = \sum_{j\in E} x_j.$$

Now, f satisfying (3.5) also satisfies (3.2), since x and π are proportional, and therefore, P_μ-a.s.,

$$\lim_{N\uparrow\infty} \frac{1}{\nu(N)} \sum_{k=1}^{N} f(X_k) = \sum_{i\in E} f(i)x_i.$$

Combination of the above equalities gives, P_μ-a.s.,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f(X_k) = \lim_{N\to\infty} \frac{\nu(N)}{N} \frac{1}{\nu(N)} \sum_{k=1}^{N} f(X_k) = \frac{\sum_{i\in E} f(i)x_i}{\sum_{j\in E} x_j},$$

from which (3.6) follows, since π is obtained by normalization of x. □

Corollary 3.1.2 *Let $\{X_n\}_{n\geq 1}$ be an irreducible positive recurrent Markov chain with the stationary distribution π, and let $g : E^{L+1} \to \mathbb{R}$ be such that*

$$\sum_{i_0,\dots,i_L} |g(i_0, \dots, i_L)|\pi(i_0)p_{i_0 i_1} \cdots p_{i_{L-1} i_L} < \infty$$

(see Example 3.5.8) Then for all initial distributions μ, P_μ-a.s.

$$\lim \frac{1}{N}\sum_{k=1}^{N} g(X_k, X_{k+1}, \dots, X_{k+L}) = \sum_{i_0,i_1,\dots,i_L} g(i_0, i_1, \dots, i_L)\pi(i_0)p_{i_0 i_1}\cdots p_{i_{L-1}i_L}.$$

Proof. Apply Corollary 3.1.1 to the snake chain $\{(X_n, X_{n+1}, \dots, X_{n+L})\}_{n\geq 0}$, which is irreducible recurrent and admits the stationary distribution

$$\pi(i_0)p_{i_0 i_1}\cdots p_{i_{L-1}i_L}.$$

□

Note that

$$\sum_{i_0,i_1,\dots,i_L} g(i_0, i_1, \dots, i_L)\pi(i_0)p_{i_0 i_1}\cdots p_{i_{L-1}i_L} = E_\pi[g(X_0, \dots, X_L)]$$

3.2 Convergence in variation

The purpose is to bound the “distance” between two probability distributions, this distance being the variation distance. One interest of this, among others, is to replace a random element by another for which computations may be easier.

Definition 3.2.1 *Let E be a countable space. The* distance in variation *between two probability distributions α and β on E is the quantity*

$$d_V(\alpha, \beta) := \frac{1}{2}\sum_{i\in E} |\alpha(i) - \beta(i)|. \tag{3.7}$$

That d_V is indeed a distance is clear.

Lemma 3.2.1 *Let α and β be two probability distributions on the same countable space E. Then*

$$\begin{aligned} d_V(\alpha, \beta) &= \sup_{A\subseteq E}\{\alpha(A) - \beta(A)\} \\ &= \sup_{A\subseteq E}\{|\alpha(A) - \beta(A)|\}. \end{aligned}$$

Proof. For the second equality observe that for each subset A there is a subset B such that $|\alpha(A) - \beta(A)| = \alpha(B) - \beta(B)$ (take $B = A$ or $\bar{A}$). For the first equality, write

$$\alpha(A) - \beta(A) = \sum_{i \in E} 1_A(i)\{\alpha(i) - \beta(i)\}$$

and observe that the right-hand side is maximal for $A = \{i \in E;\ \alpha(i) > \beta(i)\}$. Therefore, with $g(i) = \alpha(i) - \beta(i)$,

$$\sup_{A \subseteq E}\{\alpha(A) - \beta(A)\} = \sum_{i \in E} g^+(i) = \frac{1}{2}\sum_{i \in E} |g(i)|\ ,$$

where the equality $\sum_{i \in E} g(i) = 0$ was taken into account. □

The distance in variation *between two random variables* X and Y with values in E is the distance in variation between their probability distributions, and it is denoted (with a slight abuse of notation) by $d_V(X, Y)$. Therefore

$$d_V(X, Y) := \frac{1}{2}\sum_{i \in E} |P(X = i) - P(Y = i)|\,.$$

The distance in variation *between a random variable X with values in E and a probability distribution α on E* denoted (again with a slight abuse of notation) by $d_V(X, \alpha)$ is defined by

$$d_V(X, \alpha) := \frac{1}{2}\sum_{i \in E} |P(X = i) - \alpha(i)|\,.$$

The coupling inequality

Coupling two discrete probability distributions π' on E' and π'' on E'' consists in the construction of a probability distribution π on $E := E' \times E''$ such that the marginal distributions of π on E' and E'' respectively are π' and π'', that is

$$\sum_{j \in E''} \pi(i, j) = \pi'(i) \text{ and } \sum_{i \in E'} \pi(i, j) = \pi''(j)\,.$$

For two probability distributions α and β on the countable set E, let $\mathcal{D}(\alpha, \beta)$ be the collection of pairs of random variables (X, Y) taking their values in $E \times E$, and with marginal distributions α and β, that is,

$$P(X = i) = \alpha(i),\ P(Y = i) = \beta(i)\,. \tag{3.8}$$

Theorem 3.2.1 *For any pair $(X,Y) \in \mathcal{D}(\alpha,\beta)$, we have the fundamental* coupling inequality

$$d_V(\alpha,\beta) \le P(X \neq Y), \tag{3.9}$$

and equality is attained by some pair $(X,Y) \in \mathcal{D}(\alpha,\beta)$, which is then said to realize maximal coincidence.

Proof. For arbitrary $A \subset E$,

$$P(X \neq Y) \ge P(X \in A, Y \in \bar{A}) = P(X \in A) - P(X \in A, Y \in A) \ge P(X \in A) - P(Y \in A),$$

and therefore

$$P(X \neq Y) \ge \sup_{A \subset E}\{P(X \in A) - P(Y \in A)\} = d_V(\alpha,\beta).$$

We now construct $(X,Y) \in \mathcal{D}(\alpha,\beta)$ realizing equality. Let U, Z, V, and W be independent random variables; U takes its values in $\{0,1\}$, and Z, V, W take their values in E. The distributions of these random variables are given by

$$\begin{aligned}
P(U=1) &= 1 - d_V(\alpha,\beta), \\
P(Z=i) &= (\alpha(i) \wedge \beta(i)) / (1 - d_V(\alpha,\beta)), \\
P(V=i) &= (\alpha(i) - \beta(i))^+ / d_V(\alpha,\beta), \\
P(W=i) &= (\beta(i) - \alpha(i))^+ / d_V(\alpha,\beta).
\end{aligned}$$

Observe that $P(V=W)=0$. Defining

$$\begin{aligned}
(X,Y) &= (Z,Z) \text{ if } U=1 \\
&= (V,W) \text{ if } U=0,
\end{aligned}$$

we have

$$\begin{aligned}
P(X=i) &= P(U=1, Z=i) + P(U=0, V=i) \\
&= P(U=1)P(Z=i) + P(U=0)P(V=i) \\
&= \alpha(i) \wedge \beta(i) + (\alpha(i) - \beta(i))^+ = \alpha(i),
\end{aligned}$$

and similarly, $P(Y=i) = \beta(i)$. Therefore, $(X,Y) \in \mathcal{D}(\alpha,\beta)$. Also, $P(X=Y) = P(U=1) = 1 - d_V(\alpha,\beta)$, that is $P(X \neq Y) = d_V(\alpha,\beta)$. □

A sequence $\{X_n\}_{n \ge 1}$ of discrete random variables with values in E is said to converge in distribution to the probability distribution π on E if for all $i \in E$, $\lim_{n \uparrow \infty} P(X_n = i) = \pi(i)$. It is said to converge in variation to this distribution if

$$\lim_{n \uparrow \infty} \sum_{i \in E} |P(X_n = i) - \pi(i)| = 0.$$

Observe that Definition 3.2.3 concerns only the marginal distributions of the stochastic process, not the stochastic process itself. Therefore, if there exists another stochastic process $\{X'_n\}_{n\geq 0}$ such that $X_n \overset{\mathcal{D}}{\sim} X'_n$ for all $n \geq 0$, and if there exists a third one $\{X''_n\}_{n\geq 0}$ such that $X''_n \overset{\mathcal{D}}{\sim} \pi$ for all $n \geq 0$, then (3.13) follows from

$$\lim_{n\uparrow\infty} d_V(X'_n, X''_n) = 0. \tag{3.10}$$

This trivial observation is useful because of the resulting freedom in the choice of $\{X'_n\}$ and $\{X''_n\}$. An interesting situation occurs when there exists a finite random time τ such that $X'_n = X''_n$ for all $n \geq \tau$.

Definition 3.2.2 *Two stochastic processes $\{X'_n\}_{n\geq 0}$ and $\{X''_n\}_{n\geq 0}$ taking their values in the same state space E are said to couple if there exists an almost surely* finite *random time τ such that*

$$n \geq \tau \Rightarrow X'_n = X''_n. \tag{3.11}$$

The random variable τ is called a coupling time of the two processes.

Theorem 3.2.2 *For any coupling time τ of $\{X'_n\}_{n\geq 0}$ and $\{X''_n\}_{n\geq 0}$, we have the coupling inequality*

$$d_V(X'_n, X''_n) \leq P(\tau > n)\,. \tag{3.12}$$

Proof. For all $A \subseteq E$,

$$\begin{aligned}
P(X'_n \in A) - P(X''_n \in A) &= P(X'_n \in A,\ \tau \leq n) + P(X'_n \in A,\ \tau > n) \\
&- P(X''_n \in A,\ \tau \leq n) - P(X''_n \in A, \tau > n) \\
&= P(X'_n \in A,\ \tau > n) - P(X''_n \in A,\ \tau > n) \\
&\leq P(X'_n \in A,\ \tau > n) \leq P(\tau > n).
\end{aligned}$$

Inequality (3.12) then follows from Lemma 3.2.1. □

Therefore, if the coupling time is P-a.s. *finite*, that is $\lim_{n\uparrow\infty} P(\tau > n) = 0$,

$$\lim_{n\uparrow\infty} d_V(X_n, \pi) = \lim_{n\uparrow\infty} d_V(X'_n, X''_n) = 0\,.$$

Definition 3.2.3 *(A) A sequence $\{\alpha_n\}_{n\geq 0}$ of probability distributions on E is said to converge in variation to the probability distribution β on E if*

$$\lim_{n\uparrow\infty} d_V(\alpha_n, \beta) = 0\,.$$

(B) An E-valued random sequence $\{X_n\}_{n\geq 0}$ *such that for some probability distribution* π *on* E,

$$\lim_{n\uparrow\infty} d_V(X_n, \pi) = 0, \tag{3.13}$$

is said to converge in variation to π.

Kolmogorov's HMC convergence theorem

Consider a HMC that is irreducible and positive recurrent. If its initial distribution is the stationary distribution, it keeps the same distribution at all times. The chain is then said to be in the *stationary regime*, or in *equilibrium*, or in *steady state*. A question arises naturally: What is the long-run behavior of the chain when the initial distribution μ is *arbitrary*? For instance, will it *converge to equilibrium*? in what sense? The classical form of the result is that for arbitrary states i and j,

$$\lim_{n\uparrow\infty} p_{ij}(n) = \pi(j)\,, \tag{3.14}$$

if the chain is *ergodic*, according to the following definition:

Definition 3.2.4 *An irreducible positive recurrent and* aperiodic HMC *is called* ergodic.

If the state space is finite, computation of the n-th iterate of the transition matrix $\mathcal{P}$ is all that we need, in principle, to prove (3.14). Such computation requires some knowledge of the eigenstructure of $\mathbf{P}$, and there is a famous result of linear algebra, the Perron–Fröbenius theorem, that does the work. We shall give the details in Subsection 3.2. However, in the case of infinite state space, linear algebra fails to provide the answer, and recourse to other methods is necessary.

In fact, (3.14) can be drastically improved:

Theorem 3.2.3 *Let* $\{X_n\}_{n\geq 0}$ *be an ergodic* HMC *on the countable state space* E *with transition matrix* $\mathbf{P}$ *and stationary distribution* π*, and let* μ *be an arbitrary initial distribution. Then*

$$\lim_{n\uparrow\infty} \sum_{i\in E} |P_\mu(X_n = i) - \pi(i)| = 0,$$

and in particular, for all $j \in E$,

$$\lim_{n\uparrow\infty} \sum_{i\in E} |p_{ji}(n) - \pi(i)| = 0.$$

The proof will be given in Section 3.2.

The coupling proof

The proof of Theorem 3.2.3 will be given via the coupling method.

Proof. We prove that, for all probability distributions μ and ν on E,

$$\lim_{n\uparrow\infty} d_V(\mu^T \mathbf{P}^n, \nu^T \mathbf{P}^n) = 0.$$

The announced results correspond to the particular case where ν is the stationary distribution π, and particularizing further, $\mu = \delta_j$. From the discussion preceding Definition 3.2.2, it suffices to construct two coupling chains with initial distributions μ and ν, respectively. This is done in the next lemma. □

Lemma 3.2.2 *Let* $\{X_n^{(1)}\}_{n\geq 0}$ *and* $\{X_n^{(2)}\}_{n\geq 0}$ *be two independent* ergodic HMC*s with the same transition matrix* $\mathbf{P}$ *and initial distributions* μ *and* ν*, respectively. Let* $\tau = \inf\{n \geq 0;\ X_n^{(1)} = X_n^{(2)}\}$*, with* $\tau = \infty$ *if the chains never intersect. Then* τ *is, in fact, almost surely finite. Moreover, the process* $\{X'_n\}_{n\geq 0}$ *defined by*

$$X'_n = \begin{cases} X_n^{(1)} & \text{if } n \leq \tau, \\ X_n^{(2)} & \text{if } n \geq \tau \end{cases} \tag{3.15}$$

is an HMC *with transition matrix* $\mathbf{P}$ *(see the figure below).* **FIGURE?**

Proof. Step 1. Consider the product HMC $\{Z_n\}_{n\geq 0}$ defined by $Z_n = (X_n^{(1)}, X_n^{(2)})$. It takes values in $E \times E$, and the probability of transition from (i,k) to (j,ℓ) in n steps is $p_{ij}(n)p_{k\ell}(n)$. We first show that this chain is irreducible. The probability of transition from (i,k) to (j,ℓ) in n steps is $p_{ij}(n)p_{k\ell}(n)$. Since $\mathbf{P}$ is irreducible and *aperiodic*, by Theorem 1.2.2, there exists m such that for all pairs (i,j) and (k,ℓ), $n \geq m$ implies $p_{ij}(n)p_{k\ell}(n) > 0$. This implies irreducibility. (Note the essential role of aperiodicity. A simple counterexample is that of the the symmetric random walk on , which is irreducible but of period 2. The product of two indepencent such HMC's is the symmetric random walk on ^{2} which has two communications classes.)

Step 2. Next we show that the two independent chains meet in finite time. Clearly, the distribution $\tilde{\sigma}$ defined by $\tilde{\sigma}(i,j) := \pi(i)\pi(j)$ is a stationary distribution for the product chain, where π is the stationary distribution of $\mathbf{P}$. Therefore, by the stationary distribution criterion, the product chain is positive recurrent. In particular, it reaches the diagonal of E^2 in finite time, and consequently, $P(\tau < \infty) = 1$.

It remains to show that $\{X'_n\}_{n\geq 0}$ given by (3.15) is an HMC with transition matrix $\mathbf{P}$. For this we use the following lemma.

Lemma 3.2.3 *Let $X_0^1, X_0^2, Z_n^1, Z_n^2$ $(n \geq 1)$, be independent random variables, and suppose moreover that Z_n^1, Z_n^2 $(n \geq 1)$ are identically distributed. Let τ be a non-negative integer-valued random variable such that for all $m \in$, the event $\{\tau = m\}$ is expressible in terms of $X_0^1, X_0^2, Z_n^1, Z_n^2$ $(n \leq m)$. Define the sequence $\{Z_n\}_{n\geq 1}$ by*

$$\begin{aligned} Z_n &= Z_n^1 \text{ if } n \leq \tau \\ &= Z_n^2 \text{ if } n > \tau \end{aligned}$$

Then, $\{Z_n\}_{n\geq 1}$ has the same distribution as $\{Z_n^1\}_{n\geq 1}$ and is independent of X_0^1, X_0^2.

Proof. For any sets sets C_1, C_2, A_1, $\dots$, A_k in the appropriate spaces,

$$\begin{aligned} &P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell \in A_\ell, 1 \leq \ell \leq k) \\ &\quad= \sum_{m=0}^{k} P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell \in A_\ell, 1 \leq \ell \leq k, \tau = m) \\ &\qquad+ P(X_0^1 \in C_1, X_0^2 \in C_2, Z_1 \in A_1, \dots, Z_k \in A_k, \tau > k) \\ &\quad= \sum_{m=0}^{k} P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell^1 \in A_\ell, 1 \leq \ell \leq m, \tau = m, Z_r^2 \in A_r, m+1 \leq r \leq k) \\ &\qquad+ P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell^1 \in A_\ell, 1 \leq \ell \leq k, \tau > k)\,. \end{aligned}$$

Since the event $\{\tau = m\}$ is independent of $Z_{m+1}^2 \in A_{m+1}, \dots, Z_k^2 \in A_k$ $(k \geq m)$,

$$\begin{aligned} &= \sum_{m=0}^{k} P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell^1 \in A_\ell, 1 \leq \ell \leq m, \tau = m) P(Z_r^2 \in A_r, m+1 \leq r \leq k) \\ &\qquad+ P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell^1 \in A_\ell, 1 \leq \ell \leq k, \tau > k) \\ &= \sum_{m=0}^{k} P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell^1 \in A_\ell, 1 \leq \ell \leq m, \tau = m, Z_r^1 \in A_r, m+1 \leq r \leq k) \\ &\qquad+ P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell^1 \in A_\ell, 1 \leq \ell \leq k, \tau > k) \\ &= P(X_0^1 \in C_1, X_0^2 \in C_2, Z_1^1 \in A_1, \dots, Z_k^1 \in A_k)\,. \end{aligned}$$

□

Step 3. We now complete the proof. The statement of the theorem concerns only the distributions of $\{X_n^1\}_{n\geq 0}$ and $\{X_n^2\}_{n\geq 0}$, and therefore we can assume a representation

$$X_{n+1}^\ell = f(X_n^\ell, Z_{n+1}^\ell) \quad (\ell = 1, 2)\,,$$

where $X_0^1, X_0^2, Z_n^1, Z_n^2$ $(n \geq 1)$ satisfy the conditions stated in Lemma 3.2.3. The random time τ satisfies the condition of Lemma 3.2.3. Defining $\{Z_n\}_{n\geq 1}$ in the same manner as in this lemma, we therefore have

$$X_{n+1} = f(X_n, Z_{n+1}),$$

which proves the announced result. □

Null recurrent case

Theorem 3.2.3 concerns the positive recurrent case. In the null recurrent case we have *Orey's theorem*:

Theorem 3.2.4 *Let* $\mathbf{P}$ *be an irreducible null recurrent transition matrix on* E. *Then for all* $i, j \in E$,

$$\lim_{n\uparrow\infty} p_{ij}(n) = 0. \tag{3.16}$$

Perron–Frobenius

When the state space of a HMC is finite, we can rely on the standard results of linear algebra to study the asymptotic behavior of the n-step transition matrix $\mathbf{P}^n$, which depends on the eigenstructure of $\mathbf{P}$. The Perron–Frobenius theorem detailing the eigenstructure of non-negative matrices is therefore all that is needed, at least in theory.

The main result of Perron and Frobenius is that convergence to steady state of an ergodic finite state space HMC is geometric, with relative speed equal to the second-largest eigenvalue modulus (SLEM). Even if there are a few interesting models, especially in biology, where the eigenstructure of the transition matrix can be extracted, this situation remains nevertheless exceptional. It is therefore important to find estimates of the SLEM.

From the basic results of the theory of matrices relative to eigenvalues and eigenvectors we quote the following one, relative to a square matrix A of dimension r with *distinct* eigenvalues denoted $\lambda_1, \ldots, \lambda_r$. Let $u_1, \ldots, u_r$ and $v_1, \ldots, v_r$ be the associated sequences of left and right eigenvectors, respectively. Then, $u_1, \ldots, u_r$ form an independent collection of vectors, and so do $v_1, \ldots, v_r$. Also, $u_i^T v_j = 0$ if $i \neq j$. Since eigenvectors are determined up to multiplication by an arbitrary non-null scalar, one can choose them in such a way that $u_i^T v_i = 1$ for all i, $1 \leq i \leq r$. We then have the spectral decomposition

$$A^n = \sum_{i=1}^{r} \lambda_i^n v_i u_i^T. \tag{3.17}$$

EXAMPLE 3.2.1: TWO-STATE CHAIN. Consider the transition matrix on $E = \{1, 2\}$

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix},$$

where α, $\beta \in (0, 1)$. Its characteristic polynomial $(1-\alpha-\lambda)(1-\beta-\lambda) - \alpha\beta$ admits the roots $\lambda_1 = 1$ and

$$\lambda_2 = 1 - \alpha - \beta.$$

Observe at this point that $\lambda = 1$ is always an eigenvalue of a stochastic $r \times r$ matrix $\mathbf{P}$, associated with the right eigenvector $v = \mathbf{1}$ with all entries equal to 1, since $\mathbf{P1} = \mathbf{1}$. Also, the stationary distribution $\pi^T = \frac{1}{\alpha+\beta}(\beta, \alpha)$ is the left eigenvector corresponding to the eigenvalue 1. In this example, the representation (3.17) takes the form

$$\mathbf{P}^n = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & +\beta \end{pmatrix},$$

and therefore, since $|1-\alpha-\beta| < 1$,

$$\lim_{n\uparrow\infty} \mathbf{P}^n = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix}.$$

In particular, the result of convergence to steady state,

$$\lim_{n\uparrow\infty} \mathbf{P}^n = \mathbf{1}\pi^T = \mathbf{P}^\infty,$$

is obtained for this special case in a purely algebraic way. In addition, this algebraic method gives the convergence speed, which is exponential and determined by the second-largest eigenvalue absolute value. This is a general fact, which follows from the Perron–Frobenius theory of non-negative matrices below.

A matrix $A = \{a_{ij}\}_{1\le i,j\le r}$ with real coefficients is called *non-negative* (resp., *positive*) if all its entries are non-negative (resp., positive). A non-negative matrix A is called *stochastic* if $\sum_{j=1}^r a_{ij} = 1$ for all i, and *substochastic* if $\sum_{j=1}^r a_{ij} \le 1$ for all i, with strict inequality for at least one i.

Non-negativity (resp., positivity) of A will be denoted by $A \ge 0$ (resp., $A > 0$). If A and B are two matrices of the same dimensions with real coefficients, the notation $A \ge B$ (resp., $A > B$) means that $A - B \ge 0$ (resp., $A - B > 0$).

The *communication graph* of a square non-negative matrix A is the directed graph with the state space $E = \{1, \ldots, r\}$ as its set of vertices and an directed edge from vertex i to vertex j if and only if $a_{ij} > 0$.

A non-negative square matrix A is called *irreducible* (resp., *irreducible aperiodic*) if it has the same communication graph as an irreducible (resp., irreducible aperiodic) stochastic matrix. It is called *primitive* if there exists an integer k such that $A^k > 0$.

EXAMPLE 3.2.2: A non-negative matrix is primitive if and only if it is irreducible and aperiodic (Exercise **??**).

Let A be a *non-negative primitive* $r \times r$ matrix. There exists a real eigenvalue λ_1 with algebraic as well as geometric multiplicity one such that $\lambda_1 > 0$, and $\lambda_1 > |\lambda_j|$ for any other eigenvalue λ_j. Moreover, the left eigenvector u_1 and the right eigenvector v_1 associated with λ_1 can be chosen positive and such that $u_1^T v_1 = 1$.

Let $\lambda_2, \lambda_3, \dots, \lambda_r$ be the eigenvalues of A other than λ_1 ordered in such a way that

$$\lambda_1 > |\lambda_2| \geq \cdots \geq |\lambda_r| \tag{3.18}$$

We may always order the eigenvalues in such a way that if $|\lambda_2| = |\lambda_j|$ for some $j \geq 3$, then $m_2 \geq m_j$, where m_j is the algebraic multiplicity of λ_j. Then

$$A^n = \lambda_1^n v_1 u_1^T + O(n^{m_2 - 1}|\lambda_2|^n). \tag{3.19}$$

If in addition, A is stochastic (resp., substochastic), then $\lambda_1 = 1$ (resp., $\lambda_1 < 1$).

If A is stochastic and irreducible with period $d > 1$, then there are exactly d distinct eigenvalues of modulus 1, namely the d-th roots of unity, and all other eigenvalues have modulus strictly less than 1.

EXAMPLE 3.2.3: CONVERGENCE RATES VIA PERRON–FROBENIUS. If $\mathbf{P}$ is a transition matrix on $E = \{1, \dots, r\}$ that is irreducible and aperiodic, and therefore primitive, then

$$v_1 = \mathbf{1}, \; u_1 = \pi,$$

where π is the unique stationary distribution. Therefore

$$\mathbf{P}^n = \mathbf{1}\pi^T + O(n^{m_2 - 1}|\lambda_2|^n), \tag{3.20}$$

which generalizes the observation in Example 3.2.1.

3.3 Monte Carlo

Recall the method of the inverse in order to generate a discrete random variable Z with distribution $P(Z = a_i) = p_i \quad (0 \le i \le K)$. A crude algorithm based on this method would perform successively the tests $U \le p_0$?, $U \le p_0 + p_1$?, ..., until the answer is positive. Although very simple in principle, the inverse method has the following drawbacks when the size r of the state space E is large.

(a) Problems arise that are due to the small size of the intervals partitioning $[0, 1]$ and to the cost of precision in computing.

(b) Another situation is that in which the probability density π is known only up to a normalizing factor, that is, $\pi(i) = K\tilde{\pi}(i)$, and when the sum $\sum_{i \in E} \pi(i) = K^{-1}$ that gives the normalizing factor is prohibitively difficult to compute. In physics, this is a frequent case.

Approximate sampling

The quest for a random generator without these ailments is at the origin of the Monte Carlo Markov chain (MCMC) sampling methodology. The basic principle is the following. One constructs an irreducible aperiodic HMC $\{X_n\}_{n \ge 0}$ with state space E and stationary distribution π. Since the state space is finite, the chain is ergodic, and therefore, by Theorem 3.2.3, for any initial distribution μ and all $i \in E$,

$$\lim_{n \to \infty} P_\mu(X_n = i) = \pi(i) \,. \tag{3.21}$$

Therefore, when n is "large," we can consider that X_n has a distribution close to π.

The first task is that of designing the MCMC algorithm. One must find an ergodic transition matrix $\mathbf{P}$ on E, the stationary distribution of which π. In the Monte Carlo context, the transition mechanism of the chain is called a *sampling algorithm*, and the asymptotic distribution π is called the *target distribution*, or *sampled distribution*.

There are infinitely many transition matrices with a given target distribution, and among them there are infinitely many that correspond to a reversible chain, that is, such that

$$\pi(i)p_{ij} = \pi(j)p_{ji}.$$

We seek solutions of the form

$$p_{ij} = q_{ij}\alpha_{ij} \tag{3.22}$$

for $j \neq i$, where $Q = \{q_{ij}\}_{i,j \in E}$ is an arbitrary irreducible transition matrix on E, called the *candidate-generator* matrix. When the present state is i, the next *tentative* state j is chosen with probability q_{ij}. When $j \neq i$, this new state is accepted with probability α_{ij}. Otherwise, the next state is the same state i. Hence, the resulting probability of moving from i to j when $i \neq j$ is given by (3.22). It remains to select the *acceptance* probabilities α_{ij}.

EXAMPLE 3.3.1: METROPOLIS, TAKE 1. In the *Metropolis algorithm*

$$\alpha_{ij} = \min\left(1, \frac{\pi(j)q_{ji}}{\pi(i)q_{ij}}\right).$$

In Physics, it often arises, and we shall understand why later, that the distribution π is of the form 3.23.

$$\pi(i) = \frac{e^{-U(i)}}{Z}, \tag{3.23}$$

where $U : E \to \mathbb{R}$ is the "energy function" and Z is the "partition function", the normalizing constant ensuring that π is indeed a probability vector. The acceptance probability of the transition from i to j is then, assuming the candidate-generating matrix to be *symmetric*,

$$\alpha_{ij} = \min\left(1, e^{-(U(j)-U(i))}\right).$$

EXAMPLE 3.3.2: BARKER'S ALGORITHM. The *Barker algorithm*, corresponds to the choice

$$\alpha_{ij} = \frac{\pi(j)q_{ji}}{\pi(j)q_{ji} + \pi(i)q_{ij}}. \tag{3.24}$$

When the distribution π is of the form 3.23, the acceptance probability of the transition from i to j is, assuming the candidate-generating matrix to be *symmetric*,

$$\alpha_{ij} = \frac{e^{-U(i)}}{e^{-U(i)} + e^{-U(j)}}.$$

This corresponds to the basic principle of statistical thermodynamics: when two states 1 and 2 with energies E_1 and E_2, choose 1 with probability $\frac{e^{-E_1}}{e^{-E_1}+e^{-E_2}}$.

EXAMPLE 3.3.3: THE GIBBS ALGORITHM. Consider a multivariate probability distribution

$$\pi(x(1), \ldots, x(N))$$

on a set $E = \Lambda^N$, where Λ is countable. The basic step of the *Gibbs sampler* for the multivariate distribution π consists in selecting a coordinate number $1 \leq i \leq$, at random, and choosing the new value $y(i)$ of the corresponding coordinate, given the present values $x(1), \ldots, x(i-1), x(i+1), \ldots, x(N)$ of the other coordinates, with probability

$$\pi(y(i) \mid x(1), \ldots, x(i-1), x(i+1), \ldots, x(N)).$$

One checks as above that π is the stationary distribution of the corresponding chain.

Exact sampling

We attempt to construct an *exact* sample of a given π on a finite state space E, that is a random variable Z such that $P(Z = i) = \pi(i)$ for all $i \in E$. The following algorithm (*Propp–Wilson algorithm*) is based on a coupling idea. One starts as usual from an *ergodic* transition matrix $\mathbf{P}$ with stationary distribution π, just as in the classical MCMC method.

The algorithm is based on a representation of $\mathbf{P}$ in terms of a recurrence equation, that is, for given a function f and an IID sequence $\{Z_n\}_{n\geq 1}$ independent of the initial state, the chain satisfies the recurrence

$$X_{n+1} = f(X_n, Z_{n+1})\,. \tag{3.25}$$

The Propp-Wilson algorithm constructs a family of HMC with this transition matrix with the help of a unique IID sequence of random vectors $\{Y_n\}_{n\in -}$, called the *updating sequence*, where $Y_n = (Z_{n+1}(1), \cdots, Z_{n+1}(r))$ is a r-dimensional random vector, and where the coordinates $Z_{n+1}(i)$ have a common distribution, that of Z_1. For each $N \in$ and each $k \in E$, a process $\{X_n^N(k)\}_{n\geq N}$ is defined recursively by:

$$X_N^N(k) = k,$$

and, for $n \geq N$,

$$X_{n+1}^N(k) = f(X_n^N(k), Z_{n+1}(X_n^N(k))).$$

(Thus, if the chain is in state i at time n, it will be at time $n+1$ in state $j = f(i, Z_{n+1}(i)$.) Each of these processes is therefore a HMC with the transition matrix $\mathbf{P}$. Note that for all $k, \ell \in E$, and all $M, N \in$, the HMC's $\{X_n^N(k)\}_{n\geq N}$ and $\{X_n^M(\ell)\}_{n\geq M}$ use at any time $n \geq \max(M, N)$ the same updating random vector Y_{n+1}.

If, in addition to the independence of $\{Y_n\}_{n\in -}$, the components $Z_{n+1}(1)$, $Z_{n+1}(2)$, ..., $Z_{n+1}(r)$ are, for each $n \in$, independent, we say that the updating is *componentwise independent*.

Definition 3.3.1 *The random time*

$$\tau^+ = \inf\{n \geq 0; X_n^0(1) = X_n^0(2) = \cdots = X_n^0(r)\}$$

is called the forward coupling time *(Fig. 3.1). The random time*

$$\tau^- = \inf\{n \geq 1; X_0^{-n}(1) = X_0^{-n}(2) = \cdots = X_0^{-n}(r)\}$$

is called the backward coupling time *(Fig. 3.1).*

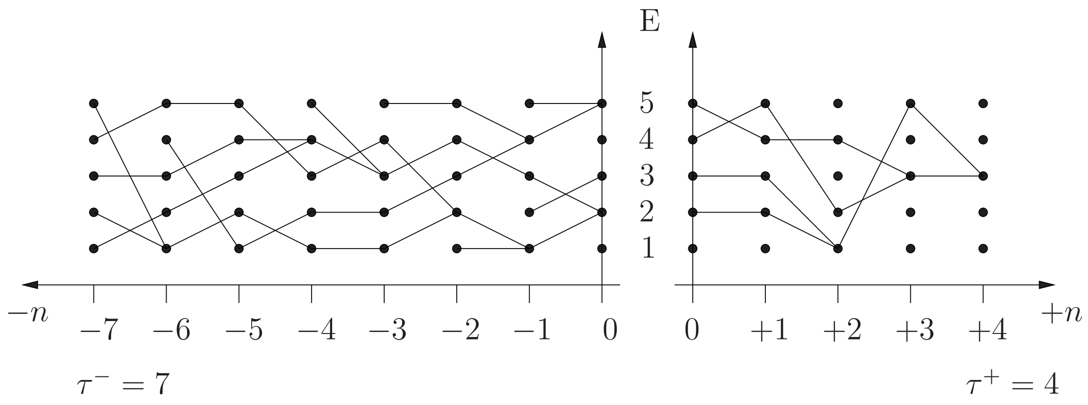


Figure 1. Backward and forward coupling

Thus, τ^+ is the first time at which the chains $\{X_n^0(i)\}_{n\geq 0}$, $1 \leq i \leq r$, *coalesce*.

Lemma 3.3.1 *When the updating is componentwise independent, the forward coupling time τ^+ is almost surely finite.*

Proof. Consider the (immediate) extension of Lemma 3.2.2 to the case of r independent HMC's with the same transition matrix. It cannot be applied directly to our situation, because the chains are not independent. However, the probability of coalescence in our situation is bounded below by the probability of coalescence in the completely independent case. To see this, first construct the independent chains model, using r independent IID componentwise independent updating sequences. The difference with our model is that we use too many updatings. In order to construct from this a set of r chains as in our model, it suffices to use for two chains the same updatings as soon as they meet. Clearly, the forward coupling time of the so modified model is smaller than or equal to that of the initial completely independent model. □

For easier notation, we set $\tau^- = \tau$. Let

$$Z = X_0^{-\tau}(i).$$

(This random variable is independent of i. In Figure 1, $Z = 2$.) Then,

Theorem 3.3.1 *With a componentwise independent updating sequence, the backwardward coupling time τ is almost surely finite. Also, the random variable Z has the distribution π.*

Proof. We shall show at the end of the current proof that for all $k \in \mathbb{Z}$, $P(\tau \leq k) = P(\tau^+ \leq k)$, and therefore the finiteness of τ follows from that of τ^+ proven in the last lemma. Now, since for $n \geq \tau$, $X_0^{-n}(i) = Z$,

$$\begin{aligned} P(Z = j) &= P(Z = j, \tau > n) + P(Z = j, \tau \leq n) \\ &= P(Z = j, \tau > n) + P(X_0^{-n}(i) = j, \tau \leq n) \\ &= P(Z = j, \tau > n) - P(X_0^{-n}(i) = j, \tau > n) + P(X_0^{-n}(i) = j) \\ &= P(Z = j, \tau > n) - P(X_0^{-n}(i) = j, \tau > n) + p_{ij}(n) \\ &= A_n - B_n + p_{ij}(n) \end{aligned}$$

But A_n and B_n are bounded above by $P(\tau > n)$, a quantity that tends to 0 as $n \uparrow \infty$ since τ is almost-surely finite. Therefore

$$P(Z = j) = \lim_{n \uparrow \infty} p_{ij}(n) = \pi(j).$$

It remains to prove the equality of the distributions of the forwards and backwards coupling time. For this, select an arbitrary integer $k \in \mathbb{Z}$. Consider an updating sequence constructed from a *bona fide* updating sequence $\{Y_n\}_{n \in \mathbb{Z}}$, by replacing $Y_{-k+1}, Y_{-k+2}, \ldots, Y_0$ by $Y_1, Y_2, \ldots, Y_k$. Call τ' the backwards coupling time in the modified model. Clearly τ an τ' have the same distribution.

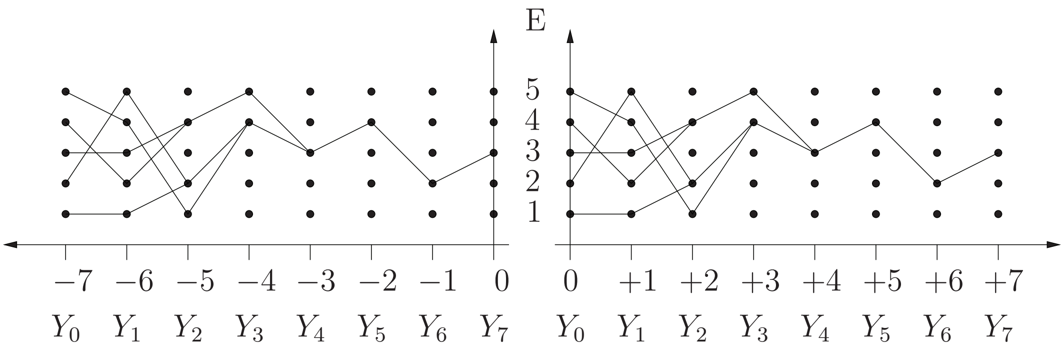


Figure 2. $\tau^+ \leq k$ implies $\tau' \leq k$

Suppose that $\tau^+ \leq k$. Consider in the modified model the chains starting at time $-k$ from states $1, \dots, r$. They coalesce at time $-k + \tau^+ \leq 0$ (see Fig. 2), and consequently $\tau' \leq k$. Therefore $\tau^+ \leq k$ implies $\tau' \leq k$, so that

$$P(\tau^+ \leq k) \leq P(\tau' \leq k) = P(\tau \leq k).$$

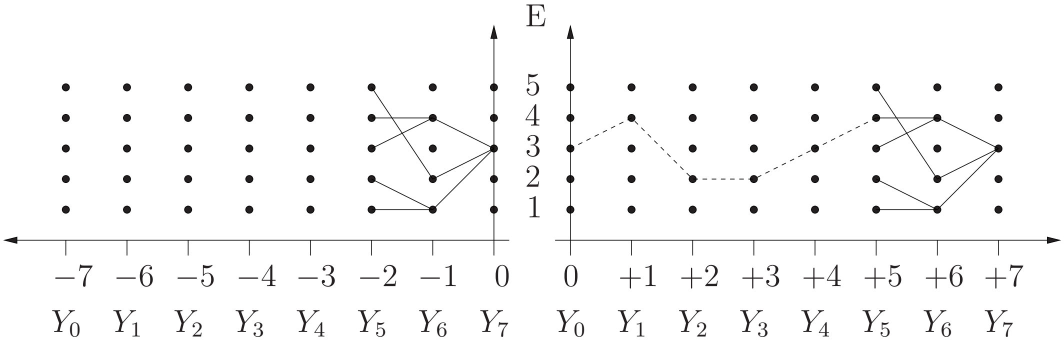


Figure 3. $\tau' \leq k$ implies $\tau^+ \leq k$

Now, suppose that $\tau' \leq k$. Then, in the modified model, the chains starting at time $k - \tau'$ from states $1, \dots, r$ must at time $-k + \tau^+ \leq 0$ coalesce at time k. Therefore (see Fig. 3), $\tau^+ \leq k$. Therefore $\tau' \leq k$ implies $\tau^+ \leq k$, so that

$$P(\tau \leq k) = P(\tau' \leq k) \leq P(\tau^+ \leq k).$$

□

Note that the coalesced value at the forward coupling time is not a sample of π (see Exercise 3.5.12).

Sandwiching

The above exact sampling algorithm is often prohibitively time-consuming when the state space is large. However, if the algorithm required the coalescence of *two*, instead of r processes, then it would take less time. The Propp and Wilson algorithm does this in a special, yet not rare, case.

It is now assumed that there exists a partial order relation on E, denoted by $\preceq$, with a minimal and a maximal element (say, respectively, 1 and r), and that we can perform the updating in such a way that for all $i, j \in E$, all $N \in$, all $n \geq N$,

$$i \preceq j \Rightarrow X_n^N(i) \preceq X_n^N(j).$$

However we do not require componentwise independent updating (but the updating vectors sequence remains IID). The corresponding sampling procedure is called the *monotone Propp–Wilson algorithm*.

Define the backwards *monotone* coupling time

$$\tau_m = \inf\{n \geq 1; X_0^{-n}(1) = X_0^{-n}(r)\} .$$

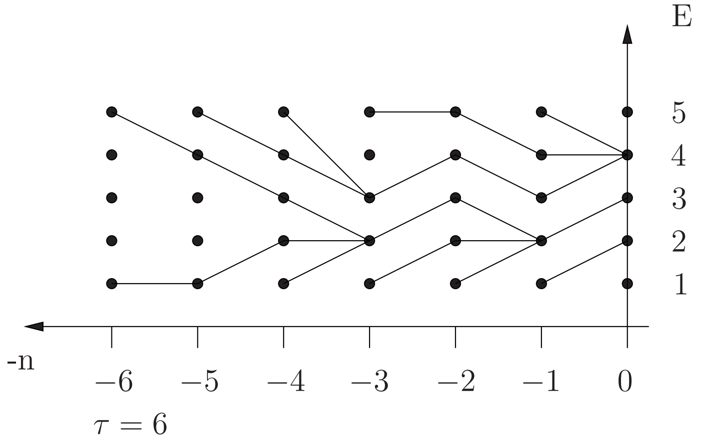


Figure 4. Monotone Propp–Wilson algorithm

Theorem 3.3.2 *The monotone backwards coupling time τ_m is almost surely finite. Also, the random variables $X_0^{-\tau_m}(1) = X_0^{-\tau_m}(r)$ has the distribution π.*

Proof. We can use most of the proof of Theorem 3.3.1. We need only to prove independently that τ^+ is finite. It is so because τ^+ is dominated by the first time $n \geq 0$ such that $X_n^0(r) = 1$, and the latter is finite in view of the recurrence assumption. □

Monotone coupling will occur with representations of the form (3.25) such that for all z,

$$i \preceq j \Rightarrow f(i, z) \preceq f(j, z),$$

and if for all $n \in$, all $i \in \{1, \ldots, r\}$,

$$Z_{n+1}(i) = Z_{n+1}.$$

EXAMPLE 3.3.4: A DAM MODEL. We consider the following model of a dam reservoir. The corresponding HMC, with values in $E = \{0, 2, \ldots, r\}$ satisfies the recurrence equation

$$X_{n+1} = \min\{r, \max(0, X_n + Z_{n+1})\},$$

where, as usual, $\{Z_n\}_{n\geq 1}$ is IID. In this specific model, X_n is the content at time n of a dam reservoir with maximum capacity r, and $Z_{n+1} = A_{n+1} - c$, where A_{n+1} is the input into the reservoir during the time period from n to $n+1$, and c is the maximum release during the same period. The updating rule is then monotone.

3.4 Absorption

Before absorption

We now consider the absorption problem for HMC's based only on the transition matrix $\mathbf{P}$, not necessarily assumed irreducible. The state space E is then decomposable as $E = T + \sum_j R_j$, where $R_1, R_2, \dots$ are the disjoint recurrent classes and T is the collection of transient states. (Note that the number of recurrent classes as well as the number of transient states may be infinite.) The transition matrix can therefore be block-partitioned as

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{P}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ B(1) & B(2) & \cdots & \mathbf{Q} \end{pmatrix}$$

or in condensed notation,

$$\mathbf{P} = \begin{pmatrix} D & 0 \\ B & \mathbf{Q} \end{pmatrix}. \tag{3.26}$$

This structure of the transition matrix accounts for the fact that one cannot go from a state in a given recurrent class to any state not belonging to this recurrent class. In other words, a recurrent class is closed.

What is the probability of being absorbed by a given recurrent class when starting from a given transient state? This kind of problem was already addressed when the first-step analysis method was introduced. It led to systems of linear equations with boundary conditions, for which the solution was unique, due to the finiteness of the state space. With an infinite state space, the uniqueness issue cannot be overlooked, and the absorption problem will be reconsidered with this in mind, and also with the intention of finding general matrix-algebraic expressions for the solutions. Another phenomenon not manifesting itself in the finite case is the possibility, when the set of transient states is infinite, of never being absorbed by the recurrent set. We shall consider this problem first, and then proceed to derive the distribution of the time to absorption by the recurrent set, and the probability of being absorbed by a given recurrent class.

Let A be a subset of the state space E (typically the set of transient states, but not necessarily). We aim at computing for any initial state $i \in A$ the probability of remaining forever in A,

$$v(i) = P_i(X_r \in A;\ r \geq 0).$$

Defining $v_n(i) := P_i(X_1 \in A, \dots, X_n \in A)$, we have, by monotone sequential continuity,

$$\lim_{n\uparrow\infty} \downarrow v_n(i) = v(i).$$

But for $j \in A$, $P_i(X_1 \in A, \dots, X_{n-1} \in A, X_n = j) = \sum_{i_1 \in A} \cdots \sum_{i_{n-1} \in A} p_{ii_1} \cdots p_{i_{n-1}j}$ is the general term $q_{ij}(n)$ of the n-th iterate of the restriction $\mathbf{Q}$ of $\mathbf{P}$ to the set A. Therefore $v_n(i) = \sum_{j \in A} q_{ij}(n)$, that is, in vector notation,

$$v_n = \mathbf{Q}^n \mathbf{1}_A,$$

where $\mathbf{1}_A$ is the column vector indexed by A with all entries equal to 1. From this equality we obtain

$$v_{n+1} = \mathbf{Q} v_n,$$

and by dominated convergence $v = \mathbf{Q}v$. Moreover, $\mathbf{0}_A \leq v \leq \mathbf{1}_A$, where $\mathbf{0}_A$ is the column vector indexed by A with all entries equal to 0. The above result can be refined as follows:

Theorem 3.4.1 *The vector v is the* maximal *solution of*

$$v = \mathbf{Q}v, \mathbf{0}_A \leq v \leq \mathbf{1}_A.$$

Moreover, either $v = \mathbf{0}_A$ or $sup_{i \in A} v(i) = 1$. In the case of a finite transient set T, the probability of infinite sojourn in T is null.

Proof. Only maximality and the last statement remain to be proven. To prove maximality consider a vector u indexed by A such that $u = \mathbf{Q}u$ and $\mathbf{0}_A \leq u \leq \mathbf{1}_A$. Iteration of $u = \mathbf{Q}u$ yields $u = \mathbf{Q}^n u$, and $u \leq \mathbf{1}_A$ implies that $\mathbf{Q}^n u \leq \mathbf{Q}^n \mathbf{1}_A = v_n$. Therefore, $u \leq v_n$, which gives $u \leq v$ by passage to the limit.

To prove the last statement of the theorem, let $c = \sup_{i \in A} v(i)$. From $v \leq c\mathbf{1}_A$, we obtain $v \leq c v_n$ as above, and therefore, at the limit, $v \leq cv$. This implies either $v = \mathbf{0}_A$ or $c = 1$.

When the set T is *finite*, the probability of infinite sojourn in T is null, because otherwise at least one transient state would be visited infinitely often. □

Equation $v = \mathbf{Q}v$ reads

$$v(i) = \sum_{j \in A} p_{ij} v(j) \ (i \in A) .$$

First-step analysis gives this equality as a *necessary* condition. However, it does not help to determine which solution to choose, in case there are several.

EXAMPLE 3.4.1: THE REPAIR SHOP ONCE MORE. We shall prove in a different way a result already obtained in Subsection 2.4, that is: the chain is recurrent if and only if $\rho \le 1$,. Observe that the restriction of $\mathbf{P}$ to $A_i := \{i+1, i+2, \dots\}$, namely

$$\mathbf{Q} = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & \cdots \\ & a_0 & a_1 & \cdots \\ & & & \cdots \end{pmatrix} ,$$

does not depend on $i \ge 0$. In particular, the maximal solution of $v = \mathbf{Q}v$, $\mathbf{0}_A \le v \le \mathbf{1}_A$ when $A \equiv A_i$ has, in view of Theorem 3.4.1, the following two interpretations. Firstly, for $i \ge 1$, $1 - v(i)$ is the probability of visiting 0 when starting from $i \ge 1$. Secondly, $(1 - v(1))$ is the probability of visiting $\{0, 1, \dots, i\}$ when starting from $i+1$. But when starting from $i+1$, the chain visits $\{0, 1, \dots, i\}$ if and only if it visits i, and therefore $(1 - v(1))$ is also the probability of visiting i when starting from $i+1$. The probability of visiting 0 when starting from $i+1$ is

$$1 - v(i+1) = (1 - v(1))(1 - v(i)),$$

because in order to go from $i+1$ to 0 one must first reach i, and then go to 0. Therefore, for all $i \ge 1$,

$$v(i) = 1 - \beta^i ,$$

where $\beta = 1 - v(1)$. To determine β, write the first equality of $v = \mathbf{Q}v$:

$$v(1) = a_1 v(1) + a_2 v(2) + \cdots ,$$

that is,

$$(1 - \beta) = a_1(1 - \beta) + a_2(1 - \beta^2) + \cdots .$$

Since $\sum_{i \ge 0} a_i = 1$, this reduces to

$$\beta = g(\beta) , \qquad (\star)$$

where g is the generating function of the probability distribution $(a_k, k \geq 0)$. Also, all other equations of $v = \mathbf{Q}v$ reduce to $(\star)$.

Under the irreduciblity assumptions $a_0 > 0$, $a_0 + a_1 < 1$, $(\star)$ has only one solution in $[0,1]$, namely $\beta = 1$ if $\rho \leq 1$, whereas if $\rho > 1$, it has two solutions in $[0,1]$, this probability is $\beta = 1$ and $\beta = \beta_0 \in (0,1)$. We must take the smallest solution. Therefore, if $\rho > 1$, the probability of visiting state 0 when starting from state $i \geq 1$ is $1 - v(i) = \beta_0^i < 1$, and therefore the chain is transient. If $\rho \leq 1$, the latter probability is $1 - v(i) = 1$, and therefore the chain is recurrent.

EXAMPLE 3.4.2: 1-D RANDOM WALK, TAKE 5. The transition matrix of the random walk on $\mathbb{N}$ with a reflecting barrier at 0,

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & & & & \\ q & 0 & p & & & \\ & q & 0 & p & & \\ & & q & 0 & p & \\ & & & \ddots & \ddots & \ddots \end{pmatrix},$$

where $p \in (0,1)$, is clearly irreducible. Intuitively, if $p > q$, there is a drift to the right, and one expects the chain to be transient. This will be proven formally by showing that the probability $v(i)$ of never visiting state 0 when starting from state $i \geq 1$ is strictly positive. In order to apply Theorem 3.4.1 with $A = \mathbb{N} - \{0\}$, we must find the general solution of $u = \mathbf{Q}u$. This equation reads

$$\begin{aligned} u(1) &= pu(2), \\ u(2) &= qu(1) + pu(3), \\ u(3) &= qu(2) + pu(4), \\ \cdots \end{aligned}$$

and its general solution is $u(i) = u(1)\sum_{j=0}^{i-1}\left(\frac{q}{p}\right)^j$. The largest value of $u(1)$ respecting the constraint $u(i) \in [0,1]$ is $u(1) = 1 - \left(\frac{q}{p}\right)$. The solution $v(i)$ is therefore

$$v(i) = 1 - \left(\frac{q}{p}\right)^i.$$

Time to absorption

We now turn to the determination of the distribution of τ, the time of exit from the transient set T. Theorem 3.4.1 tells that $v = \{v(i)\}_{i\in T}$, where $v(i) = P_i(\tau = \infty)$, is the largest solution of $v = \mathbf{Q}v$ subject to the constraints $\mathbf{0}_T \le v \le \mathbf{1}_T$, where $\mathbf{Q}$ is the restriction of $\mathbf{P}$ to the transient set T. The probability distribution of τ when the initial state is $i \in T$ is readily computed starting from the identity

$$P_i(\tau = n) = P_i(\tau \ge n) - P_i(\tau \ge n+1)$$

and the observation that for $n \ge 1$ $\{\tau \ge n\} = \{X_{n-1} \in T\}$, from which we obtain, for $n \ge 1$,

$$P_i(\tau = n) = P_i(X_{n-1} \in T) - P(X_n \in T) = \sum_{j\in T}(p_{ij}(n-1) - p_{ij}(n)).$$

Now, $p_{ij}(n)$ is for $i, j \in T$ the general term of $\mathbf{Q}^n$, and therefore:

Theorem 3.4.2

$$P_i(\tau = n) = \{(\mathbf{Q}^{n-1} - \mathbf{Q}^n)\mathbf{1}_T\}_i\,. \tag{3.27}$$

In particular, if $P_i(\tau = \infty) = 0$,

$$P_i(\tau > n) = \{\mathbf{Q}^n \mathbf{1}_T\}_i\,.$$

Proof. Only the last statement remains to be proved. From (3.27),

$$\begin{aligned} P_i(n < \tau \le n+m) &= \sum_{j=0}^{m-1}\{(\mathbf{Q}^{n+j} - \mathbf{Q}^{n+j-1})\mathbf{1}_T\}_i \\ &= \left\{\left(\mathbf{Q}^n - \mathbf{Q}^{n+m}\right)\mathbf{1}_T\right\}_i\,, \end{aligned}$$

and therefore, if $P_i(\tau = \infty) = 0$, we obtain (3.27) by letting $m \uparrow \infty$. $\square$

Absorption destination

We seek to compute the probability of absorption by a given recurrent class when starting from a given transient state. As we shall see later, it suffices for the theory to treat the case where the recurrent classes are singletons. We therefore suppose that the transition matrix has the form

$$\mathbf{P} = \begin{pmatrix} I & 0 \\ B & \mathbf{Q} \end{pmatrix}. \tag{3.28}$$

Let f_{ij} be the probability of absorption by recurrent class $R_j = \{j\}$ when starting from the transient state i. We have

$$\mathbf{P}^n = \begin{pmatrix} I & 0 \\ L_n & \mathbf{Q}^n \end{pmatrix},$$

where $L_n = (I + \mathbf{Q} + \cdots + \mathbf{Q}^n)B$. Therefore, $\lim_{n\uparrow\infty} L_n = SB$. For $i \in T$, the (i,j) term of L_n is

$$L_n(i,j) = P(X_n = j | X_0 = i).$$

Now, if T_{R_j} is the first time of visit to R_j after time 0, then

$$L_n(i,j) = P_i(T_{R_j} \le n),$$

since R_j is a closed state. Letting n go to ∞ gives the following:

Theorem 3.4.3 *For a* HMC *with transition matrix* $\mathbf{P}$ *of the form (3.28), the probability of absorption by recurrent class* $R_j = \{j\}$ *starting from transient state* i *is*

$$P_i(T_{R_j} < \infty) = (SB)_{i,R_j}.$$

The general case, where the recurrence classes are not necessarily singletons, can be reduced to the singleton case as follows. Let $\mathbf{P}^*$ be the matrix obtained from the transition matrix $\mathbf{P}$, by grouping for each j the states of recurrent class R_j into a single state $\hat{\jmath}$:

$$\mathbf{P}^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ b_{\hat{1}} & b_{\hat{2}} & \cdots & \mathbf{Q} \end{pmatrix} \tag{3.29}$$

where $b_{\hat{\jmath}} = B(j)\mathbf{1}_T$ is obtained by summation of the columns of $B(j)$, the matrix consisting of the columns $i \in R_j$ of B. The probability f_{iR_j} of absorption by class R_j when starting from $i \in T$ equals $\hat{f}_{i\hat{\jmath}}$, the probability of ever visiting $\hat{\jmath}$ when starting from i, computed for the chain with transition matrix $\mathbf{P}^*$.

EXAMPLE 3.4.3: SIBMATING. In the reproduction model called *sibmating* (sister–brother mating), two individuals are mated and two individuals from their offspring are chosen at random to be mated, and this incestuous process goes on through the subsequent generations.

We shall denote by X_n the genetic type of the mating pair at the nth generation. Clearly, $\{X_n\}_{n\ge 0}$ is a HMC with six states representing the different pairs of genotypes $AA \times AA$, $aa \times aa$, $AA \times Aa$, $Aa \times Aa$, $Aa \times aa$, $AA \times aa$, denoted respectively 1, 2, 3, 4, 5, 6. The following table gives the probabilities of occurrence of the 3 possible genotypes in the descent of a mating pair:

	AA	*Aa*	*aa*
AA AA	1	0	0
aa aa	0	0	1
AA Aa	1/2	1/2	0
Aa Aa	1/4	1/2	1/4
Aa aa	0	1/2	1/2
AA aa	0	1	0

(rows: parents' genotype; columns: descendant's genotype)

The transition matrix of $\{X_n\}_{n\geq 0}$ is then easily deduced:

$$\mathbf{P} = \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ 1/4 & & 1/2 & 1/4 & & \\ 1/16 & 1/16 & 1/4 & 1/4 & 1/4 & 1/8 \\ & 1/4 & & 1/4 & 1/2 & \\ & & & 1 & & \end{pmatrix}.$$

The set $R = \{1, 2\}$ is absorbing, and the restriction of the transition matrix to the transient set $T = \{3, 4, 5, 6\}$ is

$$\mathbf{Q} = \begin{pmatrix} 1/2 & 1/4 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/8 \\ 0 & 1/4 & 1/2 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

We find

$$S = (1 - \mathbf{Q})^{-1} = \frac{1}{6}\begin{pmatrix} 16 & 8 & 4 & 1 \\ 8 & 16 & 8 & 2 \\ 4 & 8 & 16 & 1 \\ 8 & 16 & 8 & 8 \end{pmatrix},$$

and the absorption probability matrix is

$$SB = S\begin{pmatrix} 1/4 & 0 \\ 1/16 & 1/16 \\ 0 & 1/4 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \\ 1/4 & 3/4 \\ 1/2 & 1/2 \end{pmatrix}.$$

For instance, the $(3, 2)$entry, $\frac{3}{4}$, is the probability that when starting from a couple of ancestors of type $Aa \times aa$, the race will end up in genotype $aa \times aa$.

3.5 Exercises

Exercise 3.5.1. ABBABAA!
A sequence of A's and B's is formed as follows. The first item is chosen at random, $P(A) = P(B) = \frac{1}{2}$, as is the second item, independently of the first one. When the first $n \geq 2$ items have been selected, the $(n+1)$st is chosen, independently of the letters in positions $k \leq n-2$ conditionally on the pair at position $n-1$ and n, as follows:

$$P(A \mid AA) = \frac{1}{2}, P(A \mid AB) = \frac{1}{2}, P(A \mid BA) = \frac{1}{4}, P(A \mid BB) = \frac{1}{4}.$$

What is the proportion of A's and B's in a long chain?

Exercise 3.5.2. Fixed-age retirement policy.
Let $\{U_n\}_{n\geq 1}$ be a sequence of IID random variables taking their values in $\mathbb{N}_+ = \{1, 2, \dots, \}$. The random variable U_n is interpreted as the lifetime of some equipment, or "machine", the nth one, which is replaced by the $(n+1)$st one upon failure. Thus at time 0, machine 1 is put in service until it breaks down at time U_1, whereupon it is immediately replaced by machine 2, which breaks down at time $U_1 + U_2$, and so on. The time to next failure of the current machine at time n is denoted by X_n. More precisely, the process $\{X_n\}_{n\geq 0}$ takes its values in $E = \mathbb{N}$, equals 0 at time $R_k = \sum_{i=1}^{k} U_i$, equals $U_{k+1} - 1$ at time $R_k + 1$, and then decreases of one unit per unit of ime until it reaches the value 0 at time R_{k+1}. It is assumed that for all $k \in \mathbb{N}_+$, $P(U_1 > k) > 0$, so that the state space E is $\mathbb{N}$. Then $\{X_n\}_{n\geq 0}$ is an irreducible HMC called thr forward recurrence chain. We assume positive recurrence, that is $E[U] < \infty$, where $U = U_1$.

A. Show that the chain is irreducible. Give the necessary and sufficient condition for positive recurrence. Assuming positive recurrence, what is the stationary distribution? A visit of the chain to state 0 corresponds to a breakdown of a machine. What is the empirical frequency of breakdowns?

B. Suppose that the cost of a breakdown is so important that it is better to replace a working machine during its lifetime (breakdown implies costly repairs, whereas replacement only implies moderate maintenance costs). The *fixed-age retirement policy* fixes an integer $T \geq 1$ and requires that a machine having reached age T be immediately replaced. What is the empirical frequency of breakdowns (not replacements)?

Exercise 3.5.3. Convergence speed via coupling.
Suppose that the coupling time τ in Theorem 3.2.3 satisfies

$$E[\psi(\tau)] < \infty$$

for some non-decreasing function $\psi : \mathbb{N} \to \mathbb{R}_+$ such that $\lim_{n\uparrow\infty} \psi(n) = \infty$. Show that for any initial distributions μ and ν

$$|\mu^T \mathbf{P}^n - \nu^T \mathbf{P}^n| = o\left(\frac{1}{\psi(n)}\right).$$

Exercise 3.5.4.
Let $\{Z_n\}_{n\geq 1}$ be an IID sequence of IID $\{0,1\}$-valued random variables, $P(Z_n = 1) = p \in (0,1)$. Show that for all $k \geq 1$,

$$\lim_{n\uparrow\infty} P(Z_1 + Z_2 + \cdots Z_n \text{ is divisible by } k) = 1\,.$$

Hint: modulo k.

Exercise 3.5.5.
Let $\mathbf{P}$ be an ergodic transition matrix on the *finite* state space E. Prove that for any initial distributions μ and ν, one can construct two HMC's $\{X_n\}_{n\geq 0}$ and $\{Y_n\}_{n\geq 0}$ on E with the same transition matrix $\mathbf{P}$, and the respective initial distributions μ and ν, in such a way that they couple at a finite time τ such that $E[e^{\alpha\tau}] < \infty$ for some $\alpha > 0$.

Exercise 3.5.6. THE LAZY RANDOM WALK ON THE CIRCLE.
Consider N points on the circle forming the state space $E := \{0, 1, \ldots, N-1\}$. Two points i, j are said to be neighbours if $j = i \pm 1$ *modulo* n. Consider the Markov chain $\{(X_n, Y_n)\}_{n\geq 0}$ with state space $E \times E$ and representing two particles moving on E as follows. At each time n choose X_n or Y_n with probability $\frac{1}{2}$ and move the corresponding particle to the left or to the right, equiprobably while the other particle remains still. The initial positions of the particles are a and b respectively. Compute the average time it takes until the two particles collide (the average coupling time of two lazy random walks).

Exercise 3.5.7. COUPLING TIME FOR THE 2-STATE HMC.
Find the distribution of the first meeting time of two independent HMC with state space $E = \{1, 2\}$ and transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix},$$

where $\alpha, \beta \in (0,1)$, when their initial states are different.

Exercise 3.5.8. THE SNAKE CHAIN.

Let $\{X_n\}_{n\geq 0}$ be an HMC with state space E and transition matrix $\mathbf{P}$. Define for $L \geq 1, Y_n = (X_n, X_{n+1}, \ldots, X_{n+L})$.

(a) The process $\{Y_n\}_{n\geq 0}$ takes its values in $F = E^{L+1}$. Prove that $\{Y_n\}_{n\geq 0}$ is an HMC and give the general entry of its transition matrix. (The chain $\{Y_n\}_{n\geq 0}$ is called the *snake chain* of length $L+1$ associated with $\{X_n\}_{n\geq 0}$.)

(b) Show that if $\{X_n\}_{n\geq 0}$ is irreducible, then so is $\{Y_n\}_{n\geq 0}$ if we restrict the state space of the latter to be $F = \{(i_0, \ldots, i_L) \in E^{L+1};\ p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{L-1} i_L} > 0\}$. Show that if the original chain is irreducible aperiodic, so is the snake chain.

(c) Show that if $\{X_n\}_{n\geq 0}$ has a stationary distribution π, then $\{Y_n\}_{n\geq 0}$ also has a stationary distribution. Which one?

Exercise 3.5.9. Target time.
Let π be the stationary distribution of an ergodic Markov chain with finite state space, and denote by T_i the return time to state i. Let S_Z be the time necessary to visit for the first time the random state Z chosen according to the distribution π, independently of the chain. Show that $E_i[S_Z]$ is independent of i, and give its expression in terms of the fundamental matrix.

Exercise 3.5.10. Mean time between successive visits of a set.
Let $\{X_n\}_{n\geq 0}$ be an irreducible positive recurrent HMC with stationary distribution π. Let A be a subset of the state space E and let $\{\tau(k)\}_{k\geq 1}$ be the sequence of return times to A. Show that

$$\lim_{k\uparrow\infty} \frac{\tau(k)}{k} = \frac{1}{\sum_{i\in A} \pi(i)}.$$

(This extends Formula (2.5)).

Exercise 3.5.11. Irreducibility of the Barker sampling chain.
Show that for both the Metropolis and Barker samplers, if Q is irreducible and U is not a constant, then $\mathbf{P}(T)$ is irreducible and aperiodic for all $T > 0$.

Exercise 3.5.12. Forward coupling does not yield exact sampling.
Refer to the Propp–Wilson algorithm. Show that the coalesced value at the forwards coupling time is not a sample of π. For a counterexample use the two-state HMC with $E = \{1, 2\}$, $p_{1,2} = 1$, $p_{2,2} = p_{2,1} = 1/2$.

Exercise 3.5.13. The modified random walk.
Consider the usual random walk on a graph. Its stationary distribution is in general non-uniform. We wish to modify it so as to obtain a HMC with uniform

stationary distribution. Now accept a transition from vertex i to vertex j of the original random walk with probability α_{ij}. Find one such acceptance probability depending only $d(i)$ and $d(j)$ that guarantees that the corresponding Monte Carlo Markov chain admits the uniform distribution as stationary distribution.