Notes de cours

# CHAÎNES DE MARKOV, ETC.

January 14, 2013

# Contents

# Chapter 1

# The basic theory

## 1.1 Conditional expectation

We start with the *conditional expectation given an event*. Let $Z$ be a discrete random variable with values in $E$, and let $f : E \to \mathbb{R}$ be a non-negative function. Let $A$ be some event of positive probability. The conditional expectation of $f(Z)$ given $A$, denoted by $E[f(Z) \mid A]$, is by definition the expectation when the distribution of $Z$ is replaced by its conditional distribution given $A$:

$$P(Z = z \mid A).$$

Therefore

$$E[f(Z) \mid A] = \sum_z f(z) P(Z = z \mid A).$$

Let $\{A_i\}_{i \in \mathbb{N}}$ be a partition of the sample space. The following formula is then a direct consequence of Bayes's formula of total causes:

$$E[f(Z)] = \sum_{i \in \mathbb{N}} E[f(Z) \mid A_i] P(A_i)$$

The following elementary result will be often used, and therefore, we shall promote it to the rank of theorem:

**Theorem 1.1.1** *Let $Z$ be a discrete random variable with values in $E$, and let $f : E \mapsto \mathbb{R}$ be a non-negative function. Let $A$ be some event of positive probability. Then*

$$E[f(Z) 1_A] = E[f(Z) \mid A] P(A).$$

**Proof.**

$$E\left[f(Z) \mid A\right] P(A) = \sum_{z \in E} f(z) P(Z = z \mid A) P(A) = \sum_{z \in E} f(z) P(Z = z\,,\, A)\,.$$

Now, the random variable $f(Z)1_A$ takes a non-null value if and only if this value is of the form $f(z) > 0$, and this with probability $P(Z = z\,,\, A)$. Therefore

$$E\left[f(Z)1_A\right] = \sum_{z\,;f(z)>0} f(z) P(Z = z\,,\, A) = \sum_{z \in E} f(z) P(Z = z\,,\, A)\,.$$

□

Next, we now introduce the notion of conditional expectation of some discrete random variable $Z$ given some other discrete random variable $Y$.

**Definition 1.1.1** *Let $X$ and $Y$ be two discrete random variables taking their values in the denumerable sets $F$ and $G$, respectively, and let the function $g : F \times G \to \mathbb{R}$ be either non-negative, or such that $E[|g(X,Y)|] < \infty$,. Define for each $y \in G$*

$$\psi(y) = \sum_{x \in F} g(x,y) P(X = x \mid Y = y) \tag{1.1}$$

*if $P(Y = y) > 0$, $= 0$ otherwise. This quantity is called the* conditional expectation *of $g(X,Y)$ given $Y = y$, and is denoted by $E^{Y=y}[g(X,Y)]$, or $E[g(X,Y) \mid Y = y]$. The random variable $\psi(Y)$ is called the conditional expectation of $g(X,Y)$ given $Y$, and is denoted by $E^{Y}[g(X,Y)]$ or $E[g(X,Y) \mid Y]$.*

The sum in (1.1) is well-defined (possibly infinite however) when $g$ is non-negative. Note that in the non-negative case, we have that

$$\begin{aligned}\sum_{y \in G} \psi(y) P(Y = y) &= \sum_{y \in G} \sum_{x \in F} g(x,y) P(X = x \mid Y = y) P(Y = y) \\ &= \sum_{x} \sum_{y} g(x,y) P(X = x, Y = y) \\ &= E[g(X,Y)].\end{aligned}$$

In particular, if $E[g(X,Y)] < \infty$, then

$$\sum_{y \in G} \psi(y) P(Y = y) < \infty,$$

which implies that (Theorem **??**) $P(\psi(Y) < \infty) = 1$.

Let now $g : F \times G \to \mathbb{R}$ be a function of arbitrary sign such that $E[|g(X,Y)|] < \infty$, and in particular $E[g^{\pm}(X,Y)] < \infty$. Denote by $\psi^{\pm}$ the functions associated to $g^{\pm}$ as in (1.1). As we just saw, for all $y \in G$, $\psi^{\pm}(y) < \infty$, and therefore $\psi(y) = \psi^{+}(y) - \psi^{-}(y)$ is well-defined (not an indeterminate $\infty - \infty$ form). Thus, the conditional expectation is well-defined also in the integrable case. From the observation made a few lines above, in this case,

$$|E^{Y}[g(X,Y)]| < \infty,\ P - a.s.$$

---

EXAMPLE 1.1.1: Let $X_1$ and $X_2$ be independent binomial random variables of same size $N$ and same parameter $p$. We are going to show that

$$E^{X_1+X_2}[X_1] = \psi(X_1 + X_2) = \frac{X_1 + X_2}{2}\,.$$

We have

$$\begin{aligned} P(X_1 = k | X_1 + X_2 = n) &= \frac{P(X_1 = k, X_1 + X_2 = n)}{P(X_1 + X_2 = n)} \\ &\frac{P(X_1 = k, X_2 = n - k)}{P(X_1 + X_2 = n)} \\ &\frac{P(X_1 = k) P(X_2 = n - k)}{P(X_1 + X_2 = n)} \end{aligned}$$

Expliciting the probabilities thereof, and using the fact that the sum of two independent binomial random variables with size $N$ and parameter $p$ is a binomial random variable with size $2N$ and parameter $p$, we find after a straightforward computation

$$P(X_1 = k | X_1 + X_2 = n) = \frac{\binom{N}{k}\binom{N}{n-k}}{\binom{2N}{n}}\,.$$

This is the *hypergeometric distribution*. The right-hand side of the last display is the probability of obtaining $k$ black balls when a sample of $n$ balls is randomly selected from an urn containing $N$ black balls and $N$ red balls. The mean of such a distribution is (by symmetry) $\frac{n}{2}$, therefore

$$E^{X_1+X_2=n}[X_1] = \frac{n}{2} = \psi(n)$$

and this gives the announced result.

---

EXAMPLE 1.1.2: Let $X_1$ and $X_2$ be two independent Poisson random variables with respective means $\theta_1 > 0$ and $\theta_2 > 0$. We seek to compute $E^{X_1+X_2}[X_1]$, that is $E^Y[X]$, where $X = X_1$, $Y = X_1 + X_2$. For $y \geq x$, the same computations as in Example 1.1.1 give

$$P(X = x \mid Y = y) = \frac{P(X_1 = x)P(X_2 = y - x)}{P(X_1 + X_2 = y)}.$$

Expliciting the probabilities thereof, and using the fact that the sum of two independent Poisson random variables with parameter $\theta_1$ and $\theta_2$ is a Poisson random variable with parameter $\theta_1 + \theta_2$, we find after a straightforward computation

$$P(X = x \mid Y = y) = \binom{y}{x} \left(\frac{\theta_1}{\theta_1 + \theta_2}\right)^x \left(\frac{\theta_2}{\theta_1 + \theta_2}\right)^{y-x}.$$

Therefore, with $\alpha = \frac{\theta_1}{\theta_1+\theta_2}$,

$$\psi(y) = E^{Y=y}[X] = \sum_{x=0}^{y} x \binom{y}{x} \alpha^x (1-\alpha)^{y-x} = \alpha y.$$

Finally, $E^Y[X] = \psi(Y) = \alpha Y$, that is,

$$E^{X_1+X_2}[X_1] = \frac{\theta_1}{\theta_1 + \theta_2}(X_1 + X_2).$$

### Properties of conditional expectation

The first property of conditional expectation, *linearity*, is obvious from the definitions: For all $\lambda_1, \lambda_2 \in \mathbb{R}$,

$$E^Y[\lambda_1 g_1(X, Y) + \lambda_2 g_2(X, Y)] = \lambda_1 E^Y[g_1(X, Y)] + \lambda_2 E^Y[g_2(X, Y)]$$

whenever the conditional expectations thereof are well-defined and do not produce $\infty - \infty$ forms. *Monotonicity* is equally obvious: if $g_1(x, y) \leq g_2(x, y)$, then

$$E^Y[g_1(X, Y)] \leq E^Y[g_2(X, Y)].$$

**Theorem 1.1.2** *If $g$ is non-negative or such that $E[|g(X, Y)|] < \infty$, we have*

$$E[E^Y[g(X, Y)]] = E[g(X, Y)].$$

**Proof.**

$$\begin{aligned} E[E^Y[g(X,Y)]] = E[\psi(Y)]] &= \sum_{y\in G} \psi(y)P(Y=y) \\ &= \sum_{y\in G}\sum_{x\in F} g(x,y)P(X=x \mid Y=y)P(Y=y) \\ &= \sum_{x}\sum_{y} g(x,y)P(X=x, Y=y) = E[g(X,Y)]. \end{aligned}$$

□

**Theorem 1.1.3** *If $w$ is non-negative or such that $E[|w(Y)|] < \infty$,*

$$E^Y[w(Y)] = w(Y),$$

*and more generally,*

$$E^Y[w(Y)h(X,Y)] = w(Y)E^Y[h(X,Y)].$$

*assuming that the left-hand side is well-defined.*

**Proof.** We prove the more general identity. We do the case where $w$ and $h$ are non-negative, since the general case follows easily from this special case. We have,

$$\begin{aligned} E^{Y=y}[w(Y)h(X,Y)] &= \sum_{x\in F} w(y)h(x,y)P(X=x \mid Y=y) \\ &= w(y)\sum_{x\in F} h(x,y)P(X=x \mid Y=y) \\ &= w(y)E^{Y=y}[h(X,Y)]. \end{aligned}$$

□

**Theorem 1.1.4** *If $X$ and $Y$ are independent and if $v$ is non-negative or such that $E[|v(X)|] < \infty$, then*

$$E^Y[v(X)] = E[v(X)].$$

**Proof.** We have

$$\begin{aligned} E^{Y=y}[v(X)] &= \sum_{x\in F} v(x)P(X=x \mid Y=y) \\ &= \sum_{x\in F} v(x)P(X=x) = E[v(X)]. \end{aligned}$$

□

We now give the *successive conditioning* rule. Suppose that $Y = (Y_1, Y_2)$, where $Y_1$ and $Y_2$ are discrete random variables. In this situation, we use the more developed notation

$$E^{Y}[g(X,Y)] = E^{Y_1,Y_2}[g(X,Y_1,Y_2].$$

**Theorem 1.1.5** *Let $Y = (Y_1, Y_2)$ be as above, and let $g$ be either non-negative or such that $E[|g(X,Y)|] < \infty$. Then*

$$E^{Y_2}[E^{Y_1,Y_2}[g(X,Y_1,Y_2)]] = E^{Y_2}[g(X,Y_1,Y_2)].$$

**Proof.** Let

$$\psi(Y_1,Y_2) = E^{Y_1,Y_2}[g(X,Y_1,Y_2)].$$

We must show that

$$E^{Y_2}[\psi(Y_1,Y_2)] = E^{Y_2}[g(X,Y_1,Y_2)].$$

But

$$\psi(y_1,y_2) = \sum_x g(x,y_1,y_2)P(X=x \mid Y_1=y_1, Y_2=y_2)$$

and

$$E^{Y_2=y_2}[\psi(Y_1,Y_2)] = \sum_{y_1} \psi(y_1,y_2)P(Y_1=y_1 \mid Y_2=y_2),$$

that is,

$$E^{Y_2=y_2}[\psi(Y_1,Y_2)] = \sum_{y_1}\sum_x g(x,y_1,y_2)P(X=x \mid Y_1=y_1, Y_2=y_2)P(Y_1=y_1 \mid Y_2=y_2).$$

But

$$\begin{aligned} &P(X=x \mid Y_1=y_1, Y_2=y_2)P(Y_1=y_1 \mid Y_2=y_2) \\ &\qquad = \frac{P(X=x,\ Y_1=y_1, Y_2=y_2)}{P(Y_1=y_1, Y_2=y_2)} \frac{P(Y_1=y_1, Y_2=y_2)}{P(Y_2=y_2)} \\ &\qquad = P(X=x, Y_1=y_1 \mid Y_2=y_2)\,. \end{aligned}$$

Therefore

$$\begin{aligned} E^{Y_2=y_2}[\psi(Y_1,Y_2)] &= \sum_{y_1}\sum_x g(x,y_1,y_2)P(X=x, Y_1=y_1 \mid Y_2=y_2) \\ &= E^{Y_2=y_2}[g(X,Y_1,Y_2)]. \end{aligned}$$

□

## 1.2 The transition matrix

There is a particle moving on a denumerable set $E$. If at time $n$, the particle is in position $i$, it will be at time $n+1$ in a position $j$ choosen independently of the past trajectory $X_{n-1}$, $X_{n-2}$ with probability $p_{ij}$. This can be represented by a labeled oriented graph, whose set of vertices is $E$, and for which there is an oriented 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", correspondings to positions $i$ such that $p_{ii} > 0$. This graphical interpretation of a homogeneous Markov chain (to be defined soon) in terms of a "random walk" on a set $E$ will be emphasized later on when we shall study symmetric 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.

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

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

*theis 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 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*[1].

[1]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 extends easily (exercise) to

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

where

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

This is in turn equivalent to

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

In other words, $A$ and $B$ are conditionaly independent given $X_n = i$. In words, the future at time $n$ and the past at time $n$ are conditionally independent given the present state $X_n = i$. In particular that 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 $\mu$ 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 $\mu$.

The distribution at time $n$ of the chain is the vector $\nu_n$, 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

$$\nu_n^T = \nu_0^T \mathbf{P}^n. \tag{1.2}$$

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

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

Indeed, the Bayes sequential rule and the Markov property give for the right-hand side of the latter equality

$$\sum_{i_1,\ldots,i_{m-1}\in E} p_{ii_1}p_{i_1i_2}\cdots p_{i_{m-1}j},$$

which is the general term of the $m$-th power of $\mathbf{P}$.

The probability distribution $\nu_0$ of the *initial state* 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,

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_k = i_k) = \nu_0(i_0)p_{i_0i_1}\cdots p_{i_{k-1}i_k}\,.$$

Therefore, t distribution of a discrete-time HMC is uniquely determined by its initial distribution and its transition matrix.

### Sample path realization of the transition matrix

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

**Theorem 1.2.1** *Let $\{Z_n\}_{n\geq 1}$ be an* IID *sequence of random variables with values in an arbitrary measurable space $F$. Let $E$ be a countable space, and $f : E\times F \to E$ be some measurable function. Let $X_0$ be a random variable with values in $E$, independent of $\{Z_n\}_{n\geq 1}$. The recurrence equation*

$$X_{n+1} = f(X_n, Z_{n+1}) \tag{1.3}$$

*then defines a* HMC.

**Proof.** Iteration of recurrence (1.3) shows that for all $n \geq 1$, there is a measurable function $g_n$ such that $X_n = g_n(X_0, Z_1, \ldots, Z_n)$, and therefore $P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(f(i, Z_{n+1}) = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(f(i, Z_{n+1}) = j)$, since the event $\{X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i\}$ is expressible in terms of $X_0, Z_1, \ldots, Z_n$ 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:

$$p_{ij} = P(f(i, Z_1) = j). \tag{1.4}$$

□

Not all homogeneous Markov chains receive a "natural" description of the type featured in Theorem 1.2.1. However, it is always possible to find a "theoretical" description of the kind. More exactly,

**Theorem 1.2.2** *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 1.2.1.*

**Proof.** Let

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

where $\{Z_n\}_{n\geq 1}$ is IID, uniform on $[0, 1]$. By application of Theorem 1.2.1 and of formula (1.4), we check that this HMC has the announced transition matrix. □

EXAMPLE 1.2.1: 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 1.2.1, an HMC, called a *random walk* on $\mathbb{Z}$. It is called a "symmetric" random walk if $p = \frac{1}{2}$.

---

EXAMPLE 1.2.2: REPAIR SHOP, TAKE 2. 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 $n$,

$$X_{n+1} = (X_n - 1)^+ + Z_{n+1}, \tag{1.5}$$

where $a^+ = \max(a, 0)$. In particular, if $\{Z_n\}_{n\geq 1}$ is an IID sequence independent of the initial state $X_0$, then $\{X_n\}_{n\geq 0}$ is a homogeneous Markov chain. In terms of the probability distribution

$$P(Z_1 = k) = a_k,\ k \geq 0,$$

its transition probabilities are, by formula (1.4),

$$p_{ij} = P((i-1)^+ + Z_1 = j) = P(Z_1 = j - (i-1)^+) = a_{j-(i-1)^+}.$$

---

As we already mentioned, not all homogeneous Markov chains are *naturally* described by the model of Theorem 1.2.1. A slight modification of this result considerably enlarges its scope.

**Theorem 1.2.3** *Let things be as in Theorem 1.2.1 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 1.2.1 (exercise). □

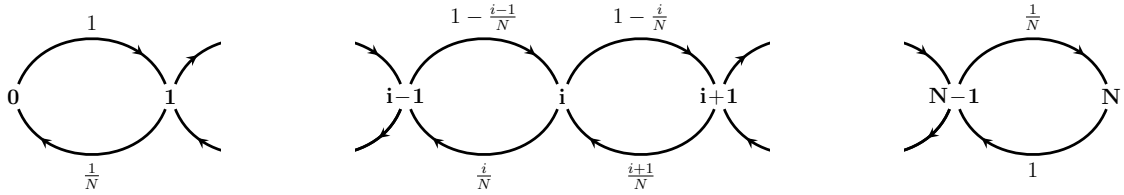
EXAMPLE 1.2.3: THE EHRENFEST URN, TAKE 1. This simplified model of diffusion through a porous membrane was proposed in 1907 by the Austrian physicists Tatiana and Paul Ehrenfest to describe in terms of statistical mechanics the exchange of heat between two systems at different temperatures. (This model considerably helped understanding the phenomenon of thermodynamic irreversibility.)

There are $N$ 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 1.2.3. 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

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

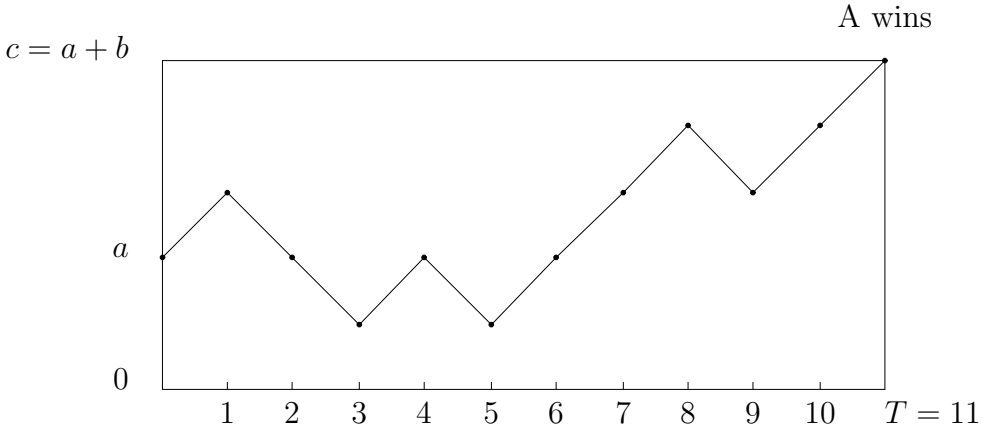



**First-step analysis**

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 1.2.4: 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 1.2.1, except that it is restricted to $E = \{0, \ldots, a, a+1, \ldots, a+b = c\}$. 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 $u(a) = P(X_T = c \mid X_0 = a)$.




The gambler's ruin

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 $i$, $1 \leq i \leq c-1$ (see, however, a rigorous proof at the close of the example),

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

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

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 $p \neq q$, and a double root, $r_1 = 1$, if $p = q = \frac{1}{2}$. 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 $\lambda$ and $\mu$. 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 $p \neq q$.

---

### Communication

All the properties defined in the present subsection are *topological* in the sense that they concern only the *naked* transition graph (without the labels).

**Definition 1.2.2** *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 an oriented 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 1.2.3** *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 1.2.4** *If there exists only one communication class, then the chain, its transition matrix, and its transition graph are said to be* irreducible.

Consider the random walk on $\mathbb{Z}$ (Example 1.2.1). Since $p \in (0, 1)$, 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 $C'_0$, $C'_1, \ldots, C'_{d'-1}$ with $d' > d$). The proof follows directly from Theorem 1.2.5 below.

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

### Period

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 1.2.5** *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 1.2.4** *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 $X_0 = i, X_M = j, X_{M+k} = j, \ldots, X_{M+nk} = j, X_{M+nk+N} = i$ 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.

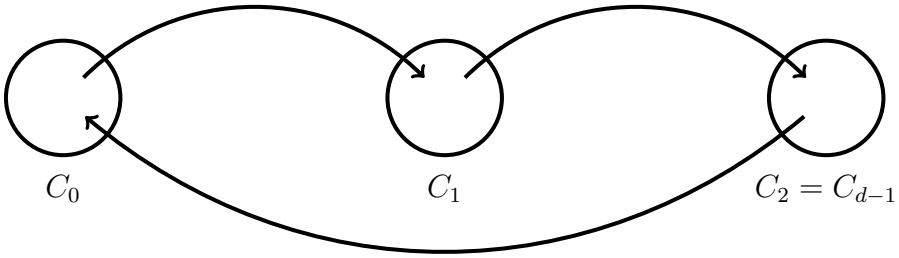
The important result concerning periodicity is the following.

**Theorem 1.2.5** *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*

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

**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. 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

### Stationarity

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

**Definition 1.2.6** *A probability distribution $\pi$ satisfying*

$$\pi^T = \pi^T \mathbf{P} \tag{1.6}$$

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

The *global balance equation* (1.6) says that for all states $i$,

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

Iteration of (1.6) gives $\pi^T = \pi^T \mathbf{P}^n$ for all $n \geq 0$, and therefore, in view of (1.2), 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,

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

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 1.2.6** *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 $\pi$ be a probability vector, i.e., $\pi^T \mathbf{1} = 1$ (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 1.2.5: TWO-STATE MARKOV CHAIN. Take $E = \{1, 2\}$ and define the transition matrix

$$\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cc} 1 & \quad 2 \end{array} \\ \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \end{array},$$

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

$$\begin{aligned} \pi(1) &= \pi(1)(1-\alpha) + \pi(2)\beta, \\ \pi(2) &= \pi(1)\alpha + \pi(2)(1-\beta). \end{aligned}$$

This is a dependent system which reduces to the single equation $\pi(1)\alpha = \pi(2)\beta$, to which must be added $\pi(1) + \pi(2) = 1$ expressing that $\pi$ is a probability vector. We obtain

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

---

EXAMPLE 1.2.6: 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 $i = 0, 1, \ldots, N$ successively, to obtain

$$\pi(i) = \pi(0)\binom{N}{i}.$$

The value of $\pi(0)$ is then determined by writing that $\pi$ 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 $\pi$ 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.

Recurrence equations can be used to obtain the stationary distribution when the latter exists and is unique. Generating functions sometimes usefully exploit the dynamics.

EXAMPLE 1.2.7: REPAIR SHOP, TAKE 3. For any complex number $z$ with modulus not larger than 1, it follows from the recurrence equation (1.5) that

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

and therefore

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

From the independence of $X_n$ and $Z_{n+1}$, $E[z^{X_n} z^{Z_{n+1}}] = E[z^{X_n}] g_Z(z)$, where $g_Z(z)$ is the generating function of $Z_{n+1}$, and $E[1_{\{X_n=0\}} z^{Z_{n+1}}] = \pi(0) g_Z(z)$, where $\pi(0) = P(X_n = 0)$. Therefore,

$$z E[z^{X_{n+1}}] - g_Z(z) E[z^{X_n}] = (z-1)\pi(0) g_Z(z).$$

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

$$g_X(z)\,(z - g_Z(z)) = \pi(0)(z-1) g_Z(z)\,. \qquad (\star)$$

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 $(\star)$:

$$g_X'(z)\,(z - g_Z(z)) + g_X(z)\,(1 - g_Z'(z)) = \pi(0)\,(g_Z(z) + (z-1) g_Z'(z))\,,$$

and let $z = 1$, to obtain, taking into account the equalities $g_X(1) = g_Z(1) = 1$ and $g_Z'(1) = E[Z]$,

$$\pi(0) = 1 - E[Z]. \qquad (1.7)$$

Since $\pi(0)$ must be non-negative, this immediately gives the necessary condition $E[Z] \le 1$. Actually, one must have, if the trivial case $Z_{n+1} \equiv 1$ is excluded,

$$E[Z] < 1. \qquad (1.8)$$

Indeed, if $E[Z] = 1$, implying $\pi(0) = 0$, it follows from (**??**) that

$$g_X(x)(x - g_Z(x)) = 0$$

for all $x \in [0,1]$. But excluding the case $Z_{n+1} \equiv 1$ (that is, $g_Z(x) \equiv x$), the equation $x - g_Z(x) = 0$ has only $x = 1$ for a solution when $g_Z'(1) = E[Z] \le 1$.

Therefore, $g_X(x) \equiv 0$ for $x \in [0,1)$, and consequently $g_X(z) \equiv 0$ on $\{|z| < 1\}$ (where $g_Z$ is analytic). This leads to a contradiction, since the generating function of an integer-valued random variable cannot be identically null.

We shall prove later that $E[Z] < 1$ is also a sufficient condition for the existence of a steady state. For the time being, we learn from (**??**) and (1.7) that, if the stationary distribution exists, then its generating function is given by the formula

$$\sum_{i=0}^{\infty} \pi(i) z^i = (1 - E[Z]) \frac{(z-1) g_Z(z)}{z - g_Z(z)}. \tag{1.9}$$

---

**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 $\pi$ such that

$$\pi > 0$$

(that is, $\pi(i) > 0$ for all states $i$). Define the matrix $\mathbf{Q}$, indexed by $E$, by

$$\pi(i) q_{ij} = \pi(j) p_{ji}. \tag{1.10}$$

This matrix is stochastic, 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 $\pi$, 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 (1.10),

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

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 1.2.7** *Let* $\mathbf{P}$ *be a stochastic matrix indexed by a countable set* $E$*, and let* $\pi$ *be a probability distribution on* $E$*. Define the matrix* $\mathbf{Q}$ *indexed by* $E$ *by (1.10). If* $\mathbf{Q}$ *is a stochastic matrix, then* $\pi$ *is a stationary distribution of* $\mathbf{P}$*.*

**Proof.** For fixed $i \in E$, sum equalities (1.10) with respect to $j \in E$ to obtain

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

But the left-hand side is equal to $\pi(i)\sum_{j\in E} q_{ij} = \pi(i)$, and therefore, for all $i \in E$,

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

□

A stationary HMC with time index $\mathbb{N}$ can be extended to $\mathbb{Z}$ while preserving stationarity.

**Theorem 1.2.8** *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* $\pi$*. Define the matrix* $\mathbf{Q} = \{q_{ij}\}_{i,j\in E}$ *by (1.10). Construct* $\{X_{-n}\}_{n\geq 1}$*, independent of* $\{X_n\}_{n\geq 1}$ *given* $X_0$*, as follows:*

$$\begin{aligned} &P(X_{-1} = i_1, X_{-2} = i_2, \ldots, X_{-k} = i_k \mid X_0 = i, X_1 = j_1, \ldots, X_n = j_n) \\ &\quad = P(X_{-1} = i_1, X_{-2} = i_2, \ldots, X_{-k} = i_k \mid X_0 = i) = q_{ii_1}q_{i_1i_2}\cdots q_{i_{k-1}i_k} \end{aligned}$$

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

**Proof.** Exercise. □

**Definition 1.2.7** *One calls* reversible *a stationary Markov chain with initial distribution* $\pi$ *(a stationary distribution) if for all* $i, j \in E$*, we have the so-called* detailed balance equations

$$\pi(i)p_{ij} = \pi(j)p_{ji}. \tag{1.11}$$

*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 1.2.7.

**Theorem 1.2.9** *Let* $\mathbf{P}$ *be a transition matrix on the countable state space* $E$*, and let* $\pi$ *be some probability distribution on* $E$*. If for all* $i, j \in E$*, the detailed balance equations (1.11) are satisfied, then* $\pi$ *is a stationary distribution of* $\mathbf{P}$*.*

EXAMPLE 1.2.8: THE EHRENFEST URN, TAKE 3. Recall that we obtained the expression

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

for the stationary distribution. Checking the detailed balance equations

$$\pi(i)p_{i,i+1} = \pi(i+1)p_{i+1,i}$$

is immediate.

EXAMPLE 1.2.9: RANDOM WALK ON A GROUP. Let $G$ be a finite associative group with respect to the operation $*$ (here called the "product")and let the inverse of $a \in G$ be denoted by $a^{-1}$ and the identity by $e$. Let $\mu$ be a probability distribution on $G$. Let $X_0$ be an arbitrary random element of $G$, and let $\{Z_n\}_{n\geq 1}$ be a sequence of IID random elements of $G$, independent of $X_0$, with common distribution $\mu$. The recurrence equation

$$X_{n+1} = Z_{n+1} * X_n \tag{1.12}$$

defines according to Theorem 1.2.1 a HMC whose transition probabilities are

$$P_{g,h*g} = \mu(h)$$

for all $g, h \in G$.

For $H \subset G$, denote by $\langle H\rangle$ the smallest subgroup of $G$ containing $H$. Recall that $\langle H\rangle$ consists of all elements of the type $b_r * b_{r-1} * \cdots * b_1$ where the $b_i$'s are elements of $H$ or inverses of elements of $H$. Let $S = \{g \in G;\ \mu(g) > 0\}$. The random walk is irreducible if and only if $S$ generates $G$, that is, $\langle S\rangle = G$.

**Proof.** Assume irreducibility. Let $a \in G$. There exists $r > 0$ such that $p_{e,a}(r) > 0$, that is, there exists a sequence $s_1, \ldots, s_r$ of $S$ such that $a = s_r * \cdots * s_1$. Therefore $a \in \langle S\rangle$. Conversely, suppose that $S$ generates $G$. Let $a, b \in G$. The element

$b * a^{-1}$ is therefore of the type $u_r * u_{r-1} * \cdots * u_1$ where the $u_i$'s are elements of $S$ or inverses of elements of $S$. Now, every element of $G$ is of finite order, that is, can be written as a power of some element of $G$. Therefore $b * a^{-1}$ can be written as $b * a^{-1} = s_r * \cdots * s_1$ where the $s_i$'s are in $S$. In particular, $p_{a,b}(r) > 0$. □

The uniform distribution $U$ on $G$ is a stationary distribution of the chain.

**Proof.** In fact

$$\sum_{g\in G} U(g)p_{g,f} = \frac{1}{|G|}\sum_{h\in G} p_{h^{-1}f,f}$$
$$= \frac{1}{|G|}\sum_{h\in G}\mu(h) = \frac{1}{|G|}\,.$$

□

The probability distribution $\mu$ on $G$ is called *symmetric* iff $\mu(g) = \mu(g^{-1})$ for all $g \in G$. If this is the case, then the chain is reversible. We just have to check the detailed balance equations

$$U(g)p_{g,h} = U(h)p_{h,g}$$

that is

$$\frac{1}{|G|}\mu(hg^{-1}) = \frac{1}{|G|}\mu(gh^{-1})\,,$$

which is true because of the assumed symmetry of $\mu$.

## 1.3 Recurrence

### Strong Markov property and recurrence

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

$$X_\infty = \Delta.$$

By definition, the "process after $\tau$" is the stochastic process

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

The "process before $\tau$," or the "process stopped at $\tau$," is the process

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

which freezes at time $\tau$ at the value $X_\tau$.

We introduce the notion of stopping time. 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

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

In other terms, this event is expressible in terms of $X_0(\omega), \ldots, X_n(\omega)$. Let now $\tau$ be a random variable with values in $\bar{\mathbb{N}}$. It is called a $X_0^n$-stopping time if for all $m \in \mathbb{N}$, $\{\tau = m\} \in X_0^m$. 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 $\tau$ is a $X_0^n$-stopping time, then so is $\tau + n$ for all $n \geq 1$.

---

EXAMPLE 1.3.1: 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 $\mathbb{N}$. This is a $X_0^n$-stopping time since for all $m \in \mathbb{N}$,

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

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 $X_0 = i$.

---

EXAMPLE 1.3.2: SUCCESSIVE RETURN TIMES. This continues the previous example. Let us fix a state, conventionally named 0, and let $T_0$ be the return time to 0. We define the successive return times to 0, $\tau_k$, $k \geq 1$ 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 $\tau_k$ is a $X_0^n$-stopping time.

---

**Theorem 1.3.1** *Let $\{X_n\}_{n\geq 0}$ be an* HMC *with state space $E$ and transition matrix $\mathbf{P}$. Let $\tau$ be a $X_0^n$-stopping time. Then for any state $i \in E$,*

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

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

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

$$\begin{aligned} P(X_{\tau+1} = j_1, \ldots, X_{\tau+k} = j_k \mid X_\tau = i, X_{\tau\wedge 0} = i_0, \ldots, X_{\tau\wedge n} = i_n) \\ = P(X_{\tau+1} = j_1, \ldots, X_{\tau+k} = j_k \mid X_\tau = i). \end{aligned}$$

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

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

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

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

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

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

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 $(\alpha)$ is proven.

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

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

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

### Regenerative cycles

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 $\tau_1 = T_0, \tau_2, \ldots$ be the successive return times to 0, and set $\tau_0 \equiv 0$.

By the strong Markov property, for any $k \geq 1$, the process after $\tau_k$ is independent of the process before $\tau_k$ (observe that condition $X_{\tau_k} = 0$ is always satisfied), and the process after $\tau_k$ 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

$$\{X_{\tau_k}, X_{\tau_k+1}, \ldots, X_{\tau_{k+1}-1}\}, k \geq 0,$$

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

---

EXAMPLE 1.3.3: 1-D RANDOM WALK, TAKE 2. Let $\{X_n\}_{n\geq 0}$ be a symmetric random walk on $\mathbb{Z}$. We show that for all positive integers $j$, $k$ and $n$,

$$P_k(T_0 < n,\, X_n = j) = P_k(X_n = -j)\,,$$

and therefore, summing over $j > 0$,

$$P_k(T_0 < n,\, X_n > 0) = P_k(X_n < 0)\,.$$

**Proof.** By the strong Markov property, for $m < n$,

$$P_k(T_0 = m,\, X_n = j) = P_k(T_0 = m)P_0(X_{n-m} = j)\,.$$

Since the distribution of $X_n$ is symmetric when the initial position is 0, the right-hand side is

$$P_k(T_0 = m)P_0(X_{n-m} = -j) = P_k(T_0 = m,\, X_n = -j)\,,$$

and therefore

$$P_k(T_0 = m,\, X_n = +j) = P_k(T_0 = m,\, X_n = -j)\,.$$

Summing over $m < n$, this gives

$$P_k(T_0 < n,\, X_n = j) = P_k(T_0 < n,\, X_n = -j) = P_k(X_n = -j),$$

where we have observed for the last equality that starting from a positive position and reaching a negative position at time $n$ implies that position 0 has been reached for the first time strictly before time $n$. □

---

EXAMPLE 1.3.4: THE GAMBLER'S RUIN, TAKE 2. A gambler with initial fortune 1 plays a heads and tails fair coin game with a one dollar stake at each toss. What is the distribution of the duration of the game until he is broke? In other terms, what is the distribution of the return time to 0 of a symmetric random walk starting from position 1? Note that in this case $T_0$ is necessarily odd. We have by the strong Markov property and the reflection principle (Example 1.3.3)

$$\begin{aligned}
P_1(T_0 = 2m+1) &= P_1(T_0 > 2m, X_{2m} = 1, X_{2m+1} = 0) \\
&= P_1(T_0 > 2m, X_{2m} = 1)P_1(X_{2m+1} = 0 \,|\, X_{2m} = 1) \\
&= P_1(T_0 > 2m, X_{2m} = 1)\frac{1}{2} \\
&= \frac{1}{2}\left\{P_1(X_{2m} = 1) - P_1(T_0 \le 2m, X_{2m} = 1)\right\} \\
&= \frac{1}{2}\left\{P_1(X_{2m} = 1) - P_1(X_{2m} = -1)\right\} \\
&= \frac{1}{2}\left\{\binom{2m}{m}2^{-2m} - \binom{2m}{m-1}2^{-2m}\right\} = \frac{\binom{2m}{m}}{m+1}2^{-2m-1}\,.
\end{aligned}$$

---

**Visits to a given state**

Let the number of visits to state $i$ strictly after time 0 be denoted by

$$N_i = \sum_{n \geq 1} 1_{\{X_n = i\}}.$$

The distribution of $N_i$ given $X_0 = j$ is

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

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

**Proof.** An informal proof goes like this: 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 independent cycle property, all these "jumps" are independent, so that the successive probabilities multiplicate. Here is a formal proof if someone needs it.

For $r = 0$, this is just the definition of $f_{ji}$. Now let $r \geq 1$, and suppose that $P_j(N_i = k) = f_{ji} f_{ii}^{k-1}(1 - f_{ii})$ is true for all $k$, $1 \leq k \leq r$. In particular,

$$P_j(N_i > r) = f_{ji} f_{ii}^{r}.$$

Denoting by $\tau_r$ the $r$th return time to state $i$,

$$\begin{aligned} P_j(N_i = r+1) &= P_j(N_i = r+1, X_{\tau_{r+1}} = i) \\ &= P_j(\tau_{r+2} - \tau_{r+1} = \infty, X_{\tau_{r+1}} = i) \\ &= P_j(\tau_{r+2} - \tau_{r+1} = \infty \mid X_{\tau_{r+1}} = i) P_j(X_{\tau_{r+1}} = i). \end{aligned}$$

But

$$P_j(\tau_{r+2} - \tau_{r+1} = \infty \mid X_{\tau_{r+1}} = i) = 1 - f_{ii}$$

by the strong Markov property ($\tau_{r+2} - \tau_{r+1}$ is the return time to $i$ of the process after $\tau_{r+1}$). Also,

$$P_j(X_{\tau_{r+1}} = i) = P_j(N_i > r)$$

Therefore,

$$P_j(N_i = r+1) = P_i(T_i = \infty) P_j(N_i > r) = (1 - f_{ii}) f_{ji} f_{ii}^{r}.$$

The result then follows by induction. □

The distribution of $N_i$ given $X_0 = j$ and given $N_i \geq 1$ is geometric. This has two main consequences. Firstly,

$$P_i(T_i < \infty) = 1 \Longleftrightarrow P_i(N_i = \infty) = 1.$$

In words: if starting from $i$ you almost surely return to $i$, then you will visit $i$ infinitely often. Secondly, we have

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

In particular,

$$P_i(T_i < \infty) < 1 \Longleftrightarrow E_i[N_i] < \infty.$$

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

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

and

$$P_i(T_i < \infty) < 1 \Longleftrightarrow P_i(N_i = \infty) = 0 \Longleftrightarrow E_i[N_i] < \infty. \tag{1.13}$$

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 $i$.

### Recurrence and the potential matrix

For the time being, we introduce the relevant definitions. First recall that $T_i$ denotes the *return* time to state $i$.

**Definition 1.3.1** *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 1.3.2** *State $i \in E$ is recurrent if and only if*

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

**Proof.** This merely rephrases Eqn. (1.13). □

EXAMPLE 1.3.5: 1-D RANDOM WALK, TAKE 3. The corresponding Markov chain was described in Example 1.2.1. The non-null terms of its transition matrix are

$$p_{i,i+1} = p \ , \ p_{i,i-1} = 1 - p,$$

where $p \in (0, 1)$. We shall study 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

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

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 1.3.6: 1-D RANDOM WALK, TAKE 4. We show that for the *symmetric* ($p = \frac{1}{2}$) 1-D random walk, the states are in fact *null* recurrent, as we proceed to prove. Let $\tau_1 = T_0, \tau_2, \ldots$ be the successive return times to state 0. Observe that for $n \geq 1$,

$$P_0(X_{2n} = 0) = \sum_{k \geq 1} P_0(\tau_k = 2n),$$

and therefore, for all $z \in \mathbb{C}$ such that $|z| < 1$,

$$\sum_{n \geq 1} P_0(X_{2n} = 0) z^{2n} = \sum_{k \geq 1} \sum_{n \geq 1} P_0(\tau_k = 2n) z^{2n} = \sum_{k \geq 1} E_0[z^{\tau_k}].$$

But $\tau_k = \tau_1 + (\tau_2 - \tau_1) + \cdots + (\tau_k - \tau_{k-1})$ and therefore, in view of Theorem 7.4, and since $\tau_1 = T_0$,

$$E_0[z^{\tau_k}] = (E_0[z^{T_0}])^k.$$

In particular,

$$\sum_{n\geq 0} P_0(X_{2n} = 0)z^{2n} = \frac{1}{1 - E_0[z^{T_0}]}$$

(note that the latter sum includes the term for $n = 0$, that is, 1). Direct evaluation of the left-hand side yields

$$\sum_{n\geq 0} \frac{1}{2^{2n}} \frac{(2n)!}{n!n!} z^{2n} = \frac{1}{\sqrt{1 - z^2}}.$$

Therefore, the generating function of the return time to 0 given $X_0 = 0$ is

$$E_0[z^{T_0}] = 1 - \sqrt{1 - z^2}.$$

Its first derivative

$$\frac{z}{\sqrt{1 - z^2}}$$

tends to $\infty$ as $z \to 1$ from below via real values. Therefore, by Abel's theorem,

$$E_0[T_0] = \infty.$$

We see that although the return time to 0 is almost surely finite, it has an infinite expectation.

---

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

**Theorem 1.3.3** *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 conclusion follows from the potential matrix 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 1.3.2** *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*

$$x^T = x^T \mathbf{P} \tag{1.14}$$

**Theorem 1.3.4** *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 $i \in E$*

$$x_i = E_0 \left[ \sum_{n=1}^{T_0} 1_{\{X_n = i\}} \right] \tag{1.15}$$

*(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 (1.15) 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

$$\sum_{i\in E} x_i = E_0[T_0]. \tag{1.16}$$

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 $x$,

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

We first prove (1.14). Observe that

$$ {}_0p_{0i}(1) = p_{0i} $$

and, using 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, (1.14), since $x_0 = 1$.

Next we show that $x_i > 0$ for all $i \in E$. Indeed, iterating (1.14), 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 1.3.5** *The invariant measure of an irreducible recurrent* HMC *is unique up to a multiplicative factor.*

**Proof.** In the proof of Theorem 1.3.4, 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 $y_0\ {}_0p_{0i}(1) = y_0p_{0i}$ 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 (1.16) and the definition of positive recurrence give the following.

**Theorem 1.3.6** *An irreducible recurrent* HMC *is positive recurrent if and only if its invariant measures $x$ satisfy*

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

### The stationary distribution criterion

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

**Proof.** The direct part follows from Theorems 1.3.4 and 1.3.6. For the converse part, assume the existence of a stationary distribution $\pi$. 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 $i, j$,

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

(In fact, $P_j(X_n = i) \leq P_j(\tau \geq n)$ where $\tau$ is the time of the last visit to $i$. But this time is *finite*, and therefore $\lim_{n\uparrow\infty} P_j(\tau \geq n) = 0$).) Since $p_{ji}(n)$ is bounded uniformly in $j$ and $n$ by 1 , by dominated convergence (Theorem **??**):

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

This contradicts the assumption that $\pi$ is a stationary distribution ($\sum_{i \in E} \pi(i) = 1$). The chain must therefore be recurrent, and by Theorem 1.3.6, it is positive recurrent.

The stationary distribution $\pi$ of an irreducible positive recurrent chain is unique (use Theorem 1.3.5 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 1.3.4). □

**Theorem 1.3.8** *Let $\pi$ be the unique stationary distribution of an irreducible positive recurrent* HMC*, and let $T_i$ be the return time to state $i$ . Then*

$$\pi(i) E_i[T_i] = 1. \tag{1.17}$$

**Proof.** This equality is a direct consequence of expression (1.15) for the invariant measure. Indeed, $\pi$ 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 (1.16),

$$\pi(0) = \frac{x_0}{\sum_{j \in E} x_j} = \frac{1}{E_0[T_0]}.$$

Since state 0 does not play a special role in the analysis, (1.17) is true for all $i \in E$. □

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

**Theorem 1.3.9** *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 1.3.7, 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 1.3.4 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 1.3.6. □

---

EXAMPLE 1.3.7: BIRTH-AND-DEATH, TAKE 1. 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 1.3.9), and it has a unique stationary distribution. Motivated by the Ehrenfest HMC which is reversible in the stationary state, we make the educated 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 $\pi$ 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

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

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

---

EXAMPLE 1.3.8: BIRTH-AND-DEATH, TAKE 2. 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) but this time with $r_i = 0$ for all $i \geq 0$. The same conditions that guarantee irreducibility are otherwise assumed. The invariant measure equation $x^T = x^T\mathbf{P}$ takes in this case the form

$$\begin{aligned} x_0 &= x_1 q_1, \\ x_i &= x_{i-1}p_{i-1} + x_{i+1}q_{i+1},\ i \geq 1\,. \end{aligned}$$

The general solution is, for $i \geq 1$, $x_i = x_0 w_i$, where $w_i$ is as in Example 1.3.7. The positive recurrence condition $\sum_{i \in E} x_i < \infty$ is

$$\sum_{i \geq 0} w_i < \infty,$$

and if it is satisfied, the stationary distribution $\pi$ is obtained by normalization of the general solution. This gives for all $i \geq 1$,

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

In the special case where $p_i = p$, $q_i = q = 1 - p$, the positive recurrence condition becomes $\sum_{j \geq 0} \left(\frac{p}{q}\right)^j < \infty$, that is to say $p < q$, or equivalently, $p < \frac{1}{2}$.

## 1.4 Foster's theorem

The stationary distribution criterion of positive recurrence of an irreducible chain requires solving the balance equation, and this is not always feasible. The need arises for more efficient conditions guaranteeing positive recurrence since. The following result (*Foster's theorem*) gives a more tractable, and quite powerful sufficient condition.

**Theorem 1.4.1** *Let the transition matrix* $\mathbf{P}$ *on the countable state space* $E$ *be irreducible and suppose that there exists a function* $h : E \to \mathbb{R}$ *such that* $\inf_i h(i) > -\infty$ *and*

$$\sum_{k \in E} p_{ik} h(k) < \infty \text{ for all } i \in F, \tag{1.18}$$

$$\sum_{k \in E} p_{ik} h(k) \leq h(i) - \epsilon \text{ for all } i \notin F, \tag{1.19}$$

*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 $\tau$ the return time to $F$, and define $Y_n = h(X_n) 1_{\{n < \tau\}}$. Equality (1.19) 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 (1.18). 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 1.4.1** *Let $\{X_n\}_{n\geq 0}$ 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 $\{Y_n\}_{n\geq 0}$ 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 $\{X_n\}$ and the stopping time $\tau_k$, and the fact that the event $\{k < \tilde{T}_i\}$ belongs to the past of $\{X_n\}$ at time $\tau_k$,

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

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

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

□

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 called *Pakes's lemma*.

**Corollary 1.4.1** *Let $\{X_n\}_{n\ge 0}$ be an irreducible* HMC *on $E = \mathbb{N}$ such that for all $n \ge 0$ and all $i \in E$,*

$$E[X_{n+1} \mid X_n = i] < \infty \tag{1.20}$$

*and*

$$\limsup_{i\uparrow\infty} E[X_{n+1} - X_n \mid X_n = i] < 0. \tag{1.21}$$

*Such an* HMC *is positive recurrent.*

**Proof.** Let $-2\epsilon$ be the left-hand side of (1.21). In particular, $\epsilon > 0$. By (1.21), 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 $F = \{i; i \le i_0\}$. □

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

**Theorem 1.4.2** *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* $h : E \to \mathbb{R}_+$ *such that*

$$\text{there exists } j \notin F \text{ such that } h(j) > \max_{i \in F} h(i) \tag{1.22}$$

$$\sup_{i \in E} \sum_{k \in E} p_{ik} |h(k) - h(i)| < \infty, \tag{1.23}$$

$$\sum_{k \in E} p_{ik} h(k) \leq h(i) \text{ for all } i \notin F. \tag{1.24}$$

*Then the chain cannot be positive recurrent.*

**Proof.** Let $\tau$ be the return time to $F$. Observe that

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

Now

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

for some finite positive constant $K$ by (1.22. Therefore, if the chain is positive recurrent, the latter bouund is $K E_j[\tau] < \infty$. We can therefore apply Fubini's theorem to obtain

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

by (1.24). Choosing $j$ that satisfies (1.22), 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 □

---

EXAMPLE 1.4.1: INSTABILITY OF THE ALOHA PROTOCOL. A typical situation in a multiple-access satellite communications system is the following. Users—each

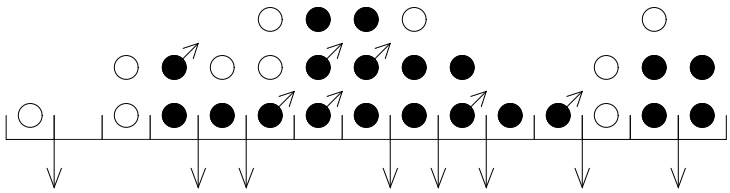
one identified with a message—contend for access to a single-channel satellite communications link for the purpose of transmitting messages. Two or more messages in the air at the same time jam each other, and are not successfully transmitted. The users are somehow able to detect a collision of this sort and will try to retransmit later the message involved in a collision. The difficulty in such communications systems resides mainly in the absence of cooperation among users, who are all unaware of the intention to transmit of competing users.

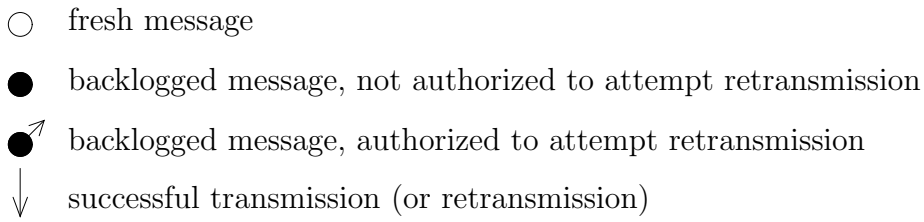
The *slotted* ALOHA protocol imposes on the users the following rules (see the Figure below):

(i) Transmissions and retransmissions of messages can start only at equally spaced moments; the interval between two consecutive (re-)transmission times is called a *slot*; the duration of a slot is always larger than that of any message.

(ii) All *backlogged* messages, i.e., those messages having already tried unsuccessfully (maybe more than once) to get through the link, require retransmission independently of one another with probability $\nu \in (0, 1)$ at each slot. This is the so-called *Bernoulli retransmission policy*.

(iii) The *fresh messages*—those presenting themselves for the first time—immediately attempt to get through.






The ALOHA protocol

Let $X_n$ be the number of backlogged messages at the beginning of slot $n$. The backlogged messages behave independently, and each one has probability $\nu$ of attempting retransmission in slot $n$. In particular, if there are $X_n = k$ backlogged

messages, the probability that $i$ among them attempt to retransmit in slot $n$ is

$$b_i(k) = \binom{k}{i} \nu^i (1-\nu)^{k-i}.$$

Let $A_n$ be the number of fresh requests for transmission in slot $n$. The sequence $\{A_n\}_{n\geq 0}$ is assumed i.i.d with the distribution

$$P(A_n = j) = a_j.$$

The quantity

$$\lambda = E[A_n] = \sum_{i=1}^{\infty} i a_i$$

is the *traffic intensity*. We suppose that $a_0 + a_1 \in (0,1)$, so that $\{X_n\}_{n\geq 0}$ is an irreducible HMC. Its transition matrix is

$$p_{ij} = \begin{cases} b_1(i)a_0 & \text{if } j = i-1, \\ [1-b_1(i)]a_0 + b_0(i)a_1 & \text{if } j = i, \\ [1-b_0(i)]a_1 & \text{if } j = i+1, \\ a_{j-i} & \text{if } j \geq i+2. \end{cases}$$

The proof is by accounting. For instance, the first line corresponds to one among the $i$ backlogged messages having succeeded to retransmit, and for this there should be no fresh arrival (probability $a_0$) and only one of the $i$ backlogged messages allowed to retransmit (probability $b_1(i)$). The second line corresponds to one of the two events "no fresh arrival and zero or strictly more than two retransmission requests from the backlog" and "zero retransmission request from the backlog and one fresh arrival."

We show that the system using the Bernoulli retransmission policy is *not stable*, in the sense that the chain $\{X_n\}_{n\geq 0}$ is *not positive recurrent*. Later on, in the next subsection, a (theoretical) remedy to this situation will be proposed. To prove unstability, we must, in view of the stationary distribution criterion, contradict the existence of a stationary distribution $\pi$. An elementary computation yields, for the ALOHA model,

$$E[X_{n+1} - X_n \mid X_n = i] = \lambda - b_1(i)a_0 - b_0(i)a_1. \tag{1.25}$$

Note that $b_1(i)a_0 + b_0(i)a_1$ is the probability of one successful (re-)transmission in a slot given that the backlog at the beginning of the slot is $i$. Equivalently, since there is at most one successful (re-)transmission in any slot, this is the average

number of successful (re-)transmissions in a slot given the backlog $i$ at the start of the slot. An elementary computation shows that $\lim_{i\uparrow\infty}(b_1(i)a_0 + b_0(i)a_1) = 0$. Therefore, outside a finite set $F$, the conditions of Theorem 1.4.2 are satisfied when we take $h$ to be the identity, and remember the hypothesis that $e\,[A_1] < \infty$.

EXAMPLE 1.4.2: STABILIZATION OF ALOHA. Since the original ALOHA protocol with a fixed retransmission probability $\nu$ is unstable, it seems natural to try a retransmission probability $\nu = \nu(k)$ depending on the number $k$ of backlogged messages. We show that there is a choice of the function $\nu(k)$ that achieves stability of the protocol. The probability that $i$ among the $k$ backlogged messages at the beginning of slot $n$ retransmit in slot $n$ is now $\nu(k)$. The same is true for the transition probabilities.

According to Pakes's lemma and using (1.25), it suffices to find a function $\nu(k)$ guaranteeing that

$$\lambda \le \lim_{i\uparrow\infty}(b_1(i)a_0 + b_0(i)a_1) - \epsilon, \tag{1.26}$$

for some $\epsilon > 0$. We shall therefore study the function

$$g_k(\nu) = (1-\nu)^k a_1 + k\nu(1-\nu)^{k-1}a_0,$$

since conditition (1.26) is just $\lambda \le g_i(\nu(i)) - \epsilon$. The derivative of $g_k(\nu)$ is, for $k \ge 2$,

$$g_k'(\nu) = k(1-\nu)^{k-2}[(a_0 - a_1) - \nu(ka_0 - a_1)].$$

We first assume that $a_0 > a_1$. In this case, for $k \ge 2$, the derivative is zero for

$$\nu = \nu(k) = \frac{a_0 - a_1}{ka_0 - a_1},$$

and the corresponding value of $g_k(\nu)$ is a maximum equal to

$$g_k(\nu(k)) = a_0\left(\frac{k-1}{k - a_1/a_0}\right)^{k-1}.$$

Therefore, $\lim_{k\uparrow\infty} g_k(\nu(k)) = a_0 \exp\left\{\frac{a_1}{a_0} - 1\right\}$, and we see that

$$\lambda < a_0 \exp\left\{\frac{a_1}{a_0} - 1\right\} \tag{1.27}$$

is a sufficient condition for stability of the protocol.

For instance, with a Poisson distribution of arrivals

$$a_i = e^{-\lambda}\frac{\lambda^i}{i!},$$

condition (1.27) reads

$$\lambda < e^{-1}$$

(in particular, the condition $a_0 > a_1$ is satisfied a posteriori).

If $a_0 \leq a_1$, the protocol can be shown to be unstable, whatever retransmission policy $\nu(k)$ is adopted (the reader is invited to check this).

---

## 1.5 The Markov chain ergodic theorem

This section gives conditions which guarantee that empirical averages of the type

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

converge to probabilistic averages. As a matter of fact, if the chain is irreducible positive recurrent with the stationary distribtion $\pi$, the above empirical average converges $P_\mu$-almost-surely to $E_\pi[g(X_0, \ldots, X_L)]$ for any initial distribution $\mu$ (Corollary 1.5.1), at least if $E_\pi[|g(X_0, \ldots, X_L)|] < \infty$.

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

**Proposition 1.5.1** *Let $\{X_n\}_{n\geq 0}$ be an irreducible recurrent* HMC*, and let $x$ denote the canonical invariant measure associated with state $0 \in E$,*

$$x_i = E_0\left[\sum_{n\geq 1} 1_{\{X_n=i\}} 1_{\{n\leq T_0\}}\right], \tag{1.28}$$

*where $T_0$ is the return time to $0$. Define for $n \geq 1$*

$$\nu(n) = \sum_{k=1}^{n} 1_{\{X_k=0\}}. \tag{1.29}$$

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

$$\sum_{i\in E} |f(i)|x_i < \infty. \tag{1.30}$$

*Then, for any initial distribution $\mu$, $P_\mu$-a.s.,*

$$\lim_{N\uparrow\infty} \frac{1}{\nu(N)} \sum_{k=1}^{N} f(X_k) = \sum_{i\in E} f(i)x_i. \tag{1.31}$$

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

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

In view of the regenerative cycle theorem (Theorem **??**), $\{U_p\}_{p\geq 1}$ 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,

$$\lim_{n\uparrow\infty}\frac{1}{n}\sum_{k=T_0+1}^{\tau_{n+1}} f(X_k) = \sum_{i\in E} f(i)x_i. \tag{1.32}$$

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 (1.32), the extreme terms of the above chain of inequality tend to $\sum_{i\in E} f(i)x_i$ as $n$ goes to $\infty$, and this implies (1.31). The case of a function $f$ of arbitrary sign is obtained by considering (1.31) 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 (1.30). □

**Theorem 1.5.1** *Let $\{X_n\}_{n\geq 0}$ be an irreducible positive recurrent Markov chain with the stationary distribution $\pi$, and let $f : E \to \mathbb{R}$ be such that*

$$\sum_{i\in E} |f(i)|\pi(i) < \infty. \tag{1.33}$$

*Then for any initial distribution $\mu$, $P_\mu$-a.s.,*

$$\lim_{n\uparrow\infty} \frac{1}{N} \sum_{k=1}^{N} f(X_k) = \sum_{i\in E} f(i)\pi(i). \tag{1.34}$$

**Proof.** Apply Proposition 1.5.1 to $f \equiv 1$. Condition (1.30) is satisfied, since in the positive recurrent case, $\sum_{i\in E} x_i = E_0[T_0] < \infty$. Therefore, $P_\mu$-a.s.,

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

Now, $f$ satisfying (1.33) also satisfies (1.30), since $x$ and $\pi$ are proportional, and therefore, $P_\mu$-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, $P_\mu$-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 (1.34) follows, since $\pi$ is obtained by normalization of $x$. □

**Corollary 1.5.1** *Let $\{X_n\}_{n\geq 1}$ be an irreducible positive recurrent Markov chain with the stationary distribution $\pi$, and let $g : E^{L+1} \to \mathbb{R}$ be such that*

$$\sum_{i_0,\ldots,i_L} |g(i_0,\ldots,i_L)|\pi(i_0)p_{i_0i_1}\cdots p_{i_{L-1}i_L} < \infty\,.$$

*Then for all initial distributions $\mu$, $P_\mu$-a.s.*

$$\lim \frac{1}{N} \sum_{k=1}^{N} g(X_k, X_{k+1},\ldots,X_{k+L}) = \sum_{i_0,i_1,\ldots,i_L} g(i_0,i_1,\ldots,i_L)\pi(i_0)p_{i_0i_1}\cdots p_{i_{L-1}i_L}.$$

**Proof.** Apply Theorem 1.5.1 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

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

---

EXAMPLE 1.5.1: FIXED-AGE RETIREMENT POLICY. Consider the forward recurrence time HMC defined as follows. 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 machine, the $n$th 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. We assume positive recurrence, that is $E[U] < \infty$, where $U = U_1$. In which case he stationary distribution is given by the formula

$$\pi(i) = \frac{P(U > i)}{E[U]}. \tag{1.35}$$

A visit of the chain to state 0 corresponds to a breakdown of a machine, and therefore, in view of the ergodic theorem,

$$\pi(0) = \lim_{N\uparrow\infty} \frac{1}{N} \sum_{k=1}^{N} 1_{\{X_k=0\}}$$

is the empirical frequency of breakdowns. Recall that

$$\pi(0) = E_0[T_0]^{-1},$$

where $T_0$ is the return time to 0. Here,

$$E_0[T_0] = E[U],$$

and therefore

$$\lim_{N\uparrow\infty} \frac{1}{N} \sum_{k=1}^{N} 1_{\{X_k=0\}} = \frac{1}{E[U]}. \tag{1.36}$$

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. We are interested in computing the empirical frequency of breakdowns (not replacements).

The forward recurrence chain corresponding to this situation is the same as before, except that the times $U_n$ are replaced by $V_n = U_n \wedge T$. Also, a replacement (not breakdown) occurs at time $n$ if and only if $X_n = 0$ and $X_{n-1} = T - 1$. But $X_{n-1} = T - 1$ implies $X_n = 0$, and therefore a replacement occurs at time $n$ if and only if

$$X_{n-1} = T - 1.$$

The empirical frequency of replacements is, therefore, in view of the ergodic theorem,

$$\lim_{N\uparrow\infty} \frac{1}{N} \sum_{k=1}^{N} 1_{\{X_k=T-1\}} = \pi(T-1).$$

Formula (1.35) applied to the new situation gives

$$\pi(T-1) = \frac{P(V \geq T)}{E[V]},$$

and therefore, since $V = U \wedge T$,

$$\pi(T-1) = \frac{P(U \geq T)}{E[U \wedge T]}.$$

The empirical frequency of visits to state 0 is, by (1.36),

$$\frac{1}{E[U \wedge T]}.$$

The empirical frequency of breakdowns is therefore

$$\frac{1}{E[U \wedge T]} - \frac{P(U \geq T)}{E[U \wedge T]} = \frac{P(U < T)}{E[U \wedge T]}.$$

## 1.6 Coupling of Markov chains

### Variation distance

In rather general terms, the coupling of two given probability distributions of random elements (random variables or stochastic processes) consists in the construction on the same probability space of a pair of random elements with the given distributions as marginal distributions, and correlated in an *ad hoc* manner so as to obtain bounds on their variation distance, a notion that is now introduced

**Definition 1.6.1** *Let $E$ be a countable space. The* distance in variation *between two probability distributions $\alpha$ and $\beta$ on $E$ is the quantity*

$$d_V(\alpha,\beta) := \frac{1}{2}\sum_{i\in E}|\alpha(i)-\beta(i)|. \tag{1.37}$$

That $d_V$ is indeed a distance is clear.

**Lemma 1.6.1** *Let $\alpha$ and $\beta$ 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 $g(i) = \alpha(i)-\beta(i)$,

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

since $\sum_{i\in E} g(i) = 0$. □

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 $\alpha$ 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 of two discrete probability distributions, $\pi'$ on $E'$ and $\pi''$ on $E''$, is the construction of a probability distribution $\pi$ on $E := E' \times E''$ such that the marginal distributions of $\pi$ on $E'$ and $E''$ respectively are $\pi'$ and $\pi''$, 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 $\alpha$ and $\beta$ on the countable set $E$, let $\mathcal{D}(\alpha,\beta)$ be the collection of random vectors $(X,Y)$ taking their values in $E\times E$, and with marginal distributions $\alpha$ and $\beta$, that is,

$$P(X=i) = \alpha(i),\ P(Y=i) = \beta(i)\,. \tag{1.38}$$

**Theorem 1.6.1** *For any pair $(X,Y) \in \mathcal{D}(\alpha,\beta)$, we have the fundamental* coupling inequality

$$d_V(\alpha,\beta) \le P(X\neq Y), \tag{1.39}$$

*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 is given by

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

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, $P(X = Y) = P(U = 1) = 1 - d_V(\alpha, \beta)$. □

**The coupling inequality**

**Definition 1.6.2** *(A) A sequence $\{\alpha_n\}_{n\geq 0}$ of probability distributions on $E$ is said to converge in variation to the probability distribution $\beta$ 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 $\pi$ on $E$,*

$$\lim_{n\uparrow\infty} d_V(X_n, \pi) = 0, \tag{1.40}$$

*is said to converge in variation to $\pi$.*

Observe that Definition 1.6.2 concerns only the marginal distributions of the random sequence (or *stochastic process* in the following context), not the stochastic process itself. Therefore, if there exists another stochastic process $\{X'_n\}_{n\geq 0}$ with

$X_n \overset{\mathcal{D}}{\sim} X'_n$ 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 (1.40) follows from

$$\lim_{n\uparrow\infty} d_V(X'_n, X''_n) = 0. \tag{1.41}$$

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 $\tau$ such that $X'_n = X''_n$ for all $n \geq \tau$.

**Definition 1.6.3** *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 $\tau$ such that*

$$n \geq \tau \Rightarrow X'_n = X''_n. \tag{1.42}$$

*The random variable $\tau$ is called a* coupling time *of the two processes.*

**Theorem 1.6.2** *For any coupling time $\tau$ of $\{X'_n\}_{n\geq 0}$ and $\{X''_n\}_{n\geq 0}$, we have the* coupling inequality

$$d_V(X'_n, X''_n) \leq P(\tau > n)\,. \tag{1.43}$$

**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 (1.43) then follows from Lemma 1.6.1. □

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

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

This situation will be exploited in the context of homogeneous Markov chains.

### Convergence in variation of ergodic HMC's

Consider an HMC that is irreducible and positive recurrent. In particular, 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 $\mu$ is *arbitrary*? For instance, will it *converge to equilibrium*? The classical form of the result is that for arbitrary states $i$ and $j$,

$$\lim_{n\uparrow\infty} p_{ij}(n) = \pi(j), \tag{1.44}$$

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

**Definition 1.6.4** *An irreducible positive recurrent and* aperiodic HMC *is called* ergodic.

We shall prove a much more powerful result.

**Theorem 1.6.3** *Let* $\mathbf{P}$ *be an ergodic transition matrix on the countable state space* $E$ *with stationary distribution* $\pi$*, and let* $\mu$ *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,*

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

**Proof.**

We prove that, for all probability distributions $\mu$ and $\nu$ 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 $\nu$ is the stationary distribution $\pi$, and particularizing further, $\mu = \delta_j$. From the discussion preceding Definition 1.6.3, it suffices to construct two coupling chains with initial distributions $\mu$ and $\nu$, respectively. This is done in the next theorem. □

**Theorem 1.6.4** *Let* $\{X_n^{(1)}\}_{n\geq 0}$ *and* $\{X_n^{(2)}\}_{n\geq 0}$ *be two independent* ergodic HMCs *with the same transition matrix* $\mathbf{P}$ *and initial distributions* $\mu$ *and* $\nu$*, respectively. Let* $\tau = \inf\{n \geq 0;\ X_n^{(1)} = X_n^{(2)}\}$*, with* $\tau = \infty$ *if the chains never intersect. Then* $\tau$ *is, in fact, almost surely finite. Moreover, the process* $\{X'_n\}_{n\geq 0}$ *defined by*

$$X'_n = \begin{cases} X_n^{(1)} & \text{if } n \leq \tau, \\ X_n^{(2)} & \text{if } n \geq \tau \end{cases} \tag{1.45}$$

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

**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 $(j,\ell)$ 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 $(j,\ell)$ in $n$ steps is $p_{ij}(n)p_{k\ell}(n)$. Since $\mathbf{P}$ is irreducible and *aperiodic*, by Theorem 1.2.5, there exists $m$ such that for all pairs $(i,j)$ and $(k,\ell)$, $n \geq m$ 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 $\mathbb{Z}$, which is irreducible but of period 2. The product of two indepencent such HMC's is the symmetric random walk on $\mathbb{Z}^2$ 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 $\pi$ 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 (1.45) is an HMC with transition matrix $\mathbf{P}$. For this we use the following lemma.

**Lemma 1.6.2** *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* $\tau$ *be a non-negative integer-valued random variable such that for all* $m \in \mathbb{N}$*, the event* $\{\tau = m\}$ *is expressible in terms of* $X_0^1, X_0^2, Z_n^1, Z_n^2\,(n \leq m)$*, that is, more formally* $\{\tau = m\} \in \sigma(X_0^1, X_0^2, Z_n^1, Z_n^2\,(n \leq m))$*. 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.**

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

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

$$
\begin{aligned}
&= \sum_{m=0}^{k} P(X_0^1 \in C_1, X_0^2 \in C_2, Z_1^1 \in A_1, \ldots, Z_m^1 \in A_m, \tau = m) P(Z_{m+1}^2 \in A_{m+1}, \ldots, Z_k^2 \in A_k) \\
&\quad + P(X_0^1 \in C_1, X_0^2 \in C_2, Z_1^1 \in A_1) \\
&= \sum_{m=0}^{k} P(X_0^1 \in C_1, X_0^2 \in C_2, Z_1^1 \in A_1, \ldots, Z_m^1 \in A_m, \tau = m) P(Z_{m+1}^1 \in A_{m+1}, \ldots, Z_k^1 \in A_k) \\
&\quad + P(X_0^1 \in C_1, X_0^2 \in C_2, Z_1^1 \in A_1) \\
&= \sum_{m=0}^{k} P(X_0^1 \in C_1, X_0^2 \in C_2, Z_1^1 \in A_1, \ldots, Z_m^1 \in A_m, \tau = m), Z_{m+1}^1 \in A_{m+1}, \ldots, Z_k^1 \in A_k) \\
&\quad + P(X_0^1 \in C_1, X_0^2 \in C_2, Z_1^1 \in A_1, \ldots, Z_k^1 \in A_k, \tau > k) \\
&= P(X_0^1 \in C_1, X_0^2 \in C_2, Z_1^1 \in A_1, \ldots, Z_k^1 \in A_k)\,.
\end{aligned}
$$

□

**STEP 3.** We now finish 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 1.6.2. We verify that $\tau$ satisfies the condition of Lemma 1.6.2. Defining $\{Z_n\}_{n\geq 1}$ in the same manner as in this lemma, we have

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

which proves the announced result. □

EXAMPLE 1.6.1: Consider the random walk $\{X_n\}_{n\geq 0}$ on a group $G$ (defined by (1.12) and the lines following this equation) with increment measure $\mu$ and transition matrix $\mathbf{P}$. Let $\{\hat{X}_n\}_{n\geq 0}$ be another random walk on $G$, this time corresponding to the increment measure $\hat{\mu}$ that is the symmetric of $\mu$, that is, for all $g \in G$, $\hat{\mu}(g) = \mu(g^{-1})$. Let $\hat{\mathbf{P}}$ be the corresponding transition matrix. Then, for all $n \geq 1$, denoting by $\pi$ the common stationary distribution of the above two chains (equal the uniform distribution on $G$), we have that

$$d_V(\delta_e^T \mathbf{P}^n, \pi) = d_V(\delta_e^T \hat{\mathbf{P}}^n, \pi)$$

**Proof.** The sequences of increments $g_1, g_2, \ldots, g_k$ for the first chain and $g_k^{-1}, g_{k-1}^{-1}, \ldots, g_1^{-1}$ for the second chain are equally likely, and therefore for all $a = a_k * a_{k-1} * \cdots * a_1$, $\delta_e^T \mathbf{P}^n(a) = \delta_e^T \hat{\mathbf{P}}^n(a^{-1})$, so that

$$\sum_{a\in G} |\delta_e^T \mathbf{P}^n(a) - \frac{1}{|G|}| = \sum_{a\in G} |\delta_e^T \hat{\mathbf{P}}^n(a^{-1}) - \frac{1}{|G|}| = \sum_{b\in G} |\delta_e^T \hat{\mathbf{P}}^n(b) - \frac{1}{|G|}| \,.$$

□

## 1.7 Monte Carlo

The problem that we adress now is to generate a discrete random variable with prescribed probability. For this, one is allowed to use a random generator that produces at will independent random variables, uniformly distributed on $[0, 1]$. In practice, the numbers that such random generators produce are not quite random, but they look as if they were (they are called *pseudo-random generators*). The topic of how to devise a good pseudo-random generator is out of our scope, and we shall admit that we can trust our favorite computer for providing us with an IID sequence of random variables uniformly distributed on $[0, 1]$ (from now on we call them *random numbers*).

In order to generate a discrete random variable $Z$ with distribution $P(Z = a_i) = p_i \quad (0 \leq i \leq K)$ we can apply the sampling method of the inverse. A crude algorithm based on this method would perform successively the tests $U \leq p_0$?, $U \leq p_0 + p_1$?, ..., until the answer is positive. If it is positive at the $i$-th stage $(i \geq 0)$ set $Z = a_i$. The average number of iterations required would therefore be $\sum_{i\geq 0}(i + 1)p_i = 1 + E[Z]$. This number may be too large, but there are ways of improving it, as the Example below will show for the Poisson random variable.

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) In random field simulation, another, maybe more important, reason is the necessity to enumerate the configurations, which implies coding and decoding of a mapping from the integers to the usually very large configuration space.

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

### Monte Carlo

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 $\{X_n\}_{n\geq 0}$ with state space $E$ and stationary distribution $\pi$. Since the state space is finite, the chain is ergodic, and therefore, by Theorem 1.6.3, for any initial distribution $\mu$ and all $i \in E$,

$$\lim_{n\to\infty} P_\mu(X_n = i) = \pi(i) \tag{1.46}$$

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

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 $\pi$.

In the Monte Carlo context, the transition mechanism of the chain is called a *sampling algorithm*, and the asymptotic distribution $\pi$ 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

$$\pi(i)p_{ij} = \pi(j)p_{ji}. \tag{1.47}$$

We seek solutions of the form

$$p_{ij} = q_{ij}\alpha_{ij} \tag{1.48}$$

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 $\alpha_{ij}$. 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 (1.48). It remains to select the *acceptance* probabilities $\alpha_{ij}$.

EXAMPLE 1.7.1: THE METROPOLIS ALGORITHM. In the *Metropolis algorithm*

$$\alpha_{ij} = \min\left(1, \frac{\pi(j)q_{ji}}{\pi(i)q_{ij}}\right). \tag{1.49}$$

In Physics, it often arises, and we shall understand why later, that the distribution $\pi$ is of the form 1.50.

$$\pi(i) = \frac{e^{-U(i)}}{Z}, \tag{1.50}$$

where $U : E \to \mathbb{R}$ is the "energy function" and $Z$ is the "partition function", the normalizing constant ensuring that $\pi$ 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 1.7.2: THE MODIFIED RANDOM WALK. Consider the usual random walk on a graph. Its stationary distribution is in general non-uniform. We are going to modify it so as to obtain a HMC with uniform stationary distribution. This can be done by accepting a transition from vertex $i$ to vertex $j$ of the original random walk with probability $\alpha_{ij} = \frac{d(i)}{d(j)} \wedge 1$. In this case the resulting probability transition from $i$ to $j$, $i \neq j$, is

$$p_{ij} = \frac{1}{d(i)}\left(\alpha_{ij} = \frac{d(i)}{d(j)} \wedge 1\right)$$

equal to $\frac{1}{d(i)}$ if $d(i) > d(j)$, equal to $\frac{1}{d(j)}$ if $d(i) < d(j)$. In particular, say with $d(i) > d(j)$, $p_{ij} = p_{ji} = \frac{1}{d(i)}$ and therefore, in all cases, $p_{ij} = p_{ji}$, and therefore the detailed balance equations are satisfied by the uniform distribution.

EXAMPLE 1.7.3: THE BARKER ALGORITHM. The *Barker algorithm*, corresponds to the choice

$$\alpha_{ij} = \frac{\pi(j)q_{ji}}{\pi(j)q_{ji} + \pi(i)q_{ij}}. \tag{1.51}$$

When the distribution $\pi$ is of the form 1.50, 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 $\frac{e^{-E_1}}{e^{-E_1}+e^{-E_2}}$.

What makes the Metropolis algorithm and the Barker algorithms, so interesting is the fact that their implementation requires the knowledge of the target distribution $\pi$ only up to a normalizing constant, since it depends only on the ratios $\pi(j)/\pi(i)$ (This in particular avoids the need to compute the normalizing constant $Z$ in 1.50, which is too often inaccessible to exact computation). The latter statement is true only as long as the candidate-generating matrix $Q$ is known. This is *not* the case in the following trivial example.

EXAMPLE 1.7.4: THE GIBBS ALGORITHM. Consider a multivariate probability distribution

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

on a set $E = \Lambda^N$, where $\Lambda$ is countable. The basic step of the *Gibbs sampler* for the multivariate distribution $\pi$ consists in selecting a coordinate number $i \in [1, N]$, 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 $\pi$ is the stationary distribution of the corresponding chain.

### The Propp–Wilson algorithm

We now raise our ambitions and construct an *exact* sample of a given $\pi$ 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 folowing algorithm is based on a coupling idea. One starts as usual from an *ergodic* transition matrix $\mathbf{P}$ with stationary distribution $\pi$, just as in the classical MCMC method.

We shall use 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

$$X_{n+1} = f(X_n, Z_{n+1})\,. \tag{1.52}$$

The Propp-Wilson algorithm constructs a family of HMC with this transition matrix with the help of a unique IID sequence of random vectors $\{Y_n\}_{n\in\mathbb{Z}}$, 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 $N \in \mathbb{Z}$ 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 $M, N \in \mathbb{Z}$, 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 $\{Y_n\}_{n\in\mathbb{Z}}$, the components $Z_{n+1}(1)$, $Z_{n+1}(2)$, ..., $Z_{n+1}(r)$ are, for each $n \in \mathbb{Z}$, independent, we say that the updating is *componentwise independent*.

**Definition 1.7.1** *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).*

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

**Lemma 1.7.1** *When the updating is componentwise independent, the forward coupling time $\tau^+$ is almost surely finite.*

**Proof.** Consider the (immediate) extension of Theorem 1.6.4 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

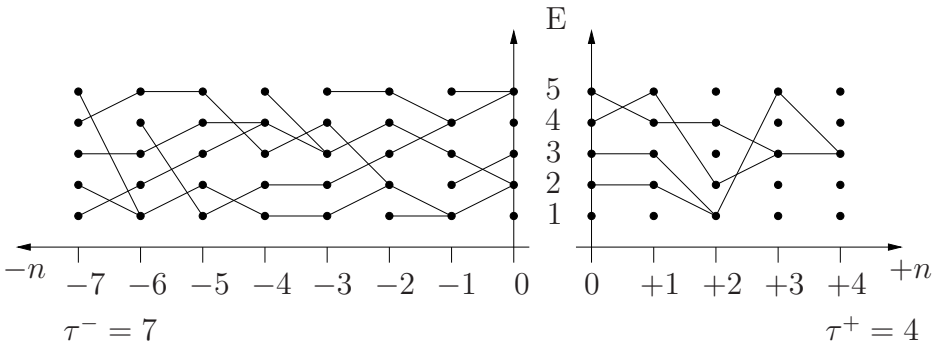



Figure 1. Backward and forward coupling

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 1.7.1** *With a componentwise independent updating sequence, the backwardward coupling time $\tau$ is almost surely finite. Also, the random variable $Z$ has the distribution $\pi$.*

**Proof.** We shall show at the end of the current proof that for all $k \in \mathbb{N}$, $P(\tau \leq k) = P(\tau^+ \leq k)$, and therefore the finiteness of $\tau$ follows from that of $\tau^+$ 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 $\tau$ 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 $k \in \mathbb{N}$. 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 $\tau'$ the backwards coupling time in the modified model. Clearly $\tau$ an $\tau'$ have the same distribution.

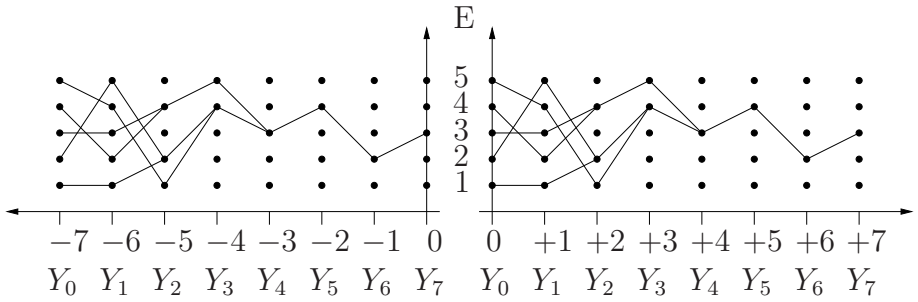



Figure 2. $\tau^+ \le k$ implies $\tau' \le k$

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

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

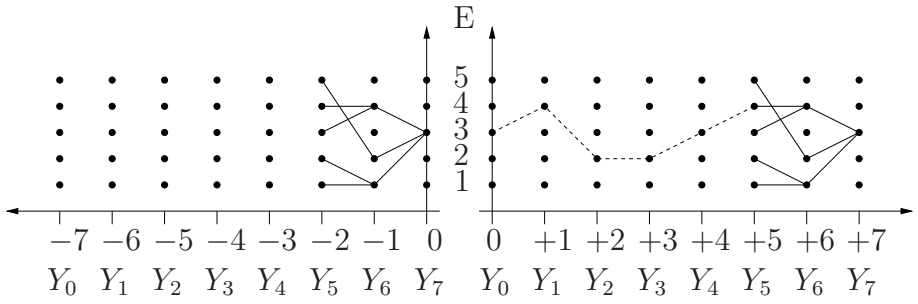



Figure 3. $\tau' \le k$ implies $\tau^+ \le k$

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

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

□

Note that the coalesced value at the forward coupling time is not in general a sample of $\pi$. (Exercise: try a counterexample with a two-state HMC.)

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 $N \in \mathbb{Z}$, 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

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

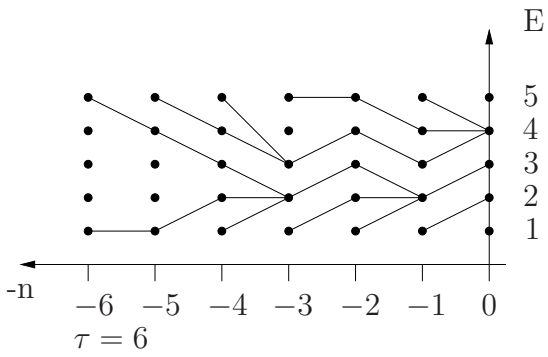



Figure 4. Monotone Propp–Wilson algorithm

**Theorem 1.7.2** *The monotone backwards coupling time $\tau_m$ is almost surely finite. Also, the random variables $X_0^{-\tau_m}(1) = X_0^{-\tau_m}(r)$ has the distribution $\pi$.*

**Proof.** We can use most of the proof of Theorem 1.7.1. We need only to prove independently that $\tau^+$ is finite. It is so because $\tau^+$ 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 (1.52) such that for all $z$,

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

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

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

---

EXAMPLE 1.7.5: 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.

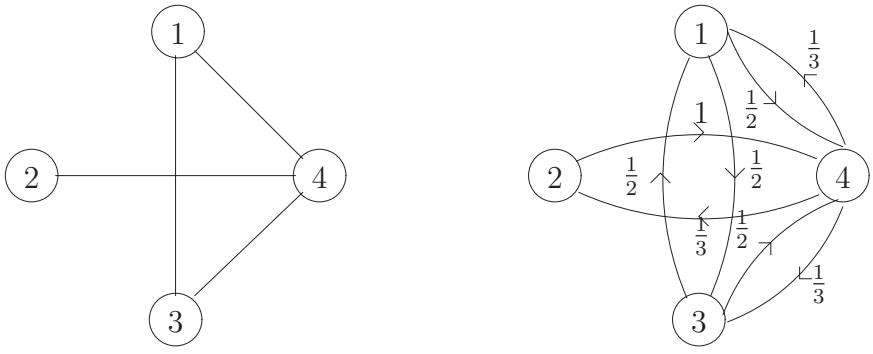
---

# Chapter 2

# Related topics

## 2.1 Random walks and spanning trees

Consider a finite non-oriented graph and call $E$ the set of vertices, or nodes, of this graph. Let $d_i$ be the *index* of vertex $i$ (the number of edges "adjacent" to node $i$). Transform this graph into an oriented graph by splitting each edge into two oriented edges of opposite directions, and make it a transition graph by associating to the oriented edge from $i$ to $j$ the transition probability $\frac{1}{d_i}$ (see the figure below). It will be assumed, as is the case in the figure, that $d_i > 0$ for all states $i$ (That is, the graph is *connected*).




A random walk on a graph

We attempt to find a stationary distribution via Theorem 1.2.9. Let $i$ and $j$ be connected in the graph, 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

$$\pi(i) = K\frac{1}{d_i},$$

where $K$ is obtained by normalization: $K = \left(\sum_{j\in E} d_j\right)^{-1}$.

---

EXAMPLE 2.1.1: 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 coordivate. For instance, in two 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.

---

EXAMPLE 2.1.2: COVER TIME OF THE RANDOM WALK. The *cover time* of a HMC is the number of steps it takes to visit all the states. We derive a bound on the maximum (with respect to the initial state) average cover time of the random walk on a graph. For this we shall first observe that the average return time to a given state $i \in E$ is $E_i[T_i] = \frac{1}{\pi(i)} = \frac{d_i}{|E|}$, where $d_i$ is the index of $i$. By first-step analysis, denoting by $N_i$ the set of states (vertices) adjacent to $i$,

$$\frac{|E|}{d_i} = E_i[T_i] = \frac{1}{d_i}\sum_{j\in N_i}(1 + E_j[T_i])$$

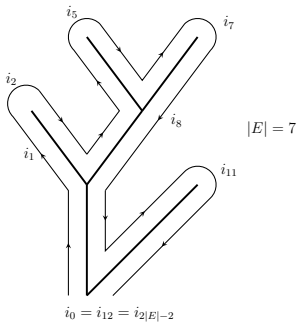
and therefore

$$|E| = \sum_{j\in N_i}(1 + E_j[T_i]),$$

from which we obtain the rough bound

$$E_j[T_i] \leq |E|$$

for any pair $(i,j)$ of states. Let now $i_0$ be an arbitrary state and consider the spanning circuit obtained by a depth-first census of the vertices of the graph (see the figure below), say $i_0, i_1, i_{2|E|-2} = i_0$.




Clearly, the average cover time from $i_0$ is lesser than or equal to

$$\sum_{\ell=0}^{2|E|-3} E_{i_\ell}\left[T_{i_{\ell+1}}\right] < (2|E|-2) \times |E| < 2|E|^2 .$$

---

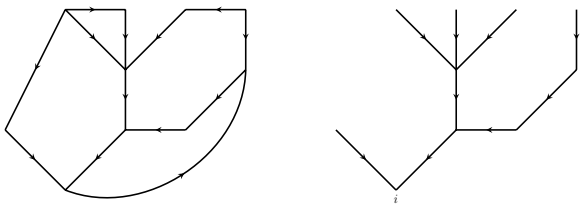
Let $\{X_n\}_{n\in\mathbb{Z}}$ be an irreducible stationary HMC with the finite state space $E$, transition matrix $\mathbf{P}$ and stationary distribution $\pi$. Let $G = (E, A)$ be the associated directed graph, where $A$ is the set of directed edges (arrows), that is of ordered pairs of states $(i, j)$ such that $p_{ij} > 0$. The weight of an edge $(i, j)$ is $p_{ij}$. A rooted spanning tree of $G$ is a directed subgraph of $G$ with the following properties:

(i) As an undirected graph it is a connected graph with $E$ as set of vertices.

(ii) As an undirected graph it is without cycles.

(iii) As a directed graph, each of its vertex has out degree 1, except one vertex, the root, that has out degree 0.

Denote by $\mathcal{S}$ the set of spanning trees of $G$, and by $\mathcal{S}_i$ the subset of $\mathcal{S}$ consisting of rooted spanning trees with vertex $i \in E$. The weight $w(S)$ of a rooted spanning tree of $S \in \mathcal{S}$ is the product of the weights of all the directed edges in $S$.

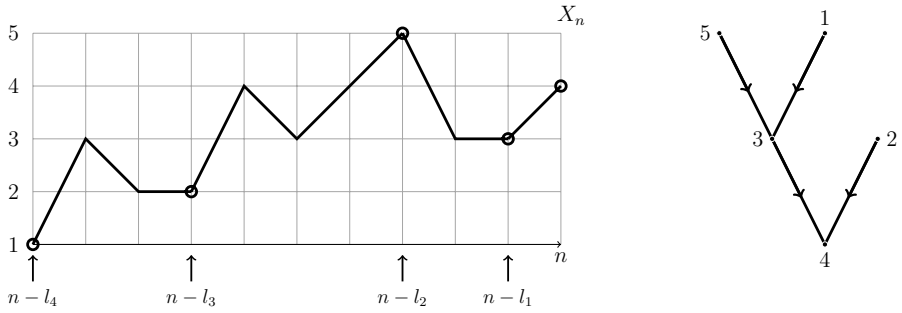



A directed graph and one of his directed spanning tree

**Theorem 2.1.1** *The stationary distribution $\pi$ of $\mathbf{P}$ is given by*

$$\pi(i) = \frac{\sum_{S\in\mathcal{S}_i} w(S)}{\sum_{S\in\mathcal{S}} w(S)}\,.$$

**Proof.** We define a stochastic process $\{Y_n\}_{n\in\mathbb{Z}}$ taking its values in $\mathcal{S}$ as follows. The root of $Y_n$ is $X_n$, say $X_n = i$. Now, by screening the past values $X_{n-1}, X_{n-2}, \ldots$ in this order, let $X_{n-\ell_1}$ be the first value different from $X_n$, let $X_{n-\ell_2}$, $\ell_2 > \ell_1$, be the first value different from $X_n$ and $X_{n-\ell_1}$, let $X_{n-\ell_3}$, $\ell_3 > \ell_2$, be the first value different from $X_n$ $X_{n-\ell_1}$ and $X_{n-\ell_2}$. Continue this procedure until you have exhausted the (finite) state space $E$. The spanning tree $Y_n$ is the one with directed edges $(X_{n-\ell_1}, X_{n-\ell_1+1} = X_n)$, $(X_{n-\ell_2}, X_{n-\ell_2+1})$, $(X_{n-\ell_3}, X_{n-\ell_3+1})$ ...




The stochastic process $\{Y_n\}_{n\in\mathbb{Z}}$ is a HMC.

The forward procedure, that is the procedure allowing to pass from $S \in \mathcal{S}$ with root $i$ to $T \in \mathcal{S}$ with root $j$ in one step is the following:

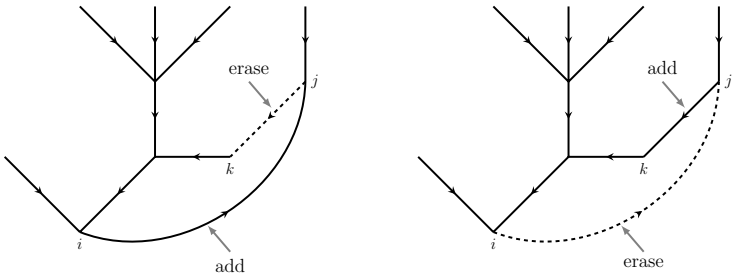
(a) Add to $S$ the directed $(i, j)$, thus creating a directed spanning graph with a unique directed loop that contains $i$ and $j$ (this may be a self-loop at $i$)

(b) Delete the unique directed edge of $S$ out of $j$, say $(j, k)$, thus breaking the loop and producing a rooted spanning tree $T \in \mathcal{S}$ with root $j$.

A rooted spanning tree $T$ with root $j$ can be obtained from the spanning tree $S$ if and only if $S$ can be constructed from $T$ by the following reverse procedure based on a suitable vertex $k$:

($\alpha$) Add to $T$ the directed edge $(j, k)$, thus creating a directed spanning graph with unique directed loop containing $j$ and $k$ (possibly a self-loop at $j$).

($\beta$) Delete the unique directed edge $(i, j)$ that lies in the loop, thus breaking the loop and producing a rooted spanning tree $T \in \mathcal{S}$ with root $i$.




Let $Q_{ST}$ be the one-step transition probability of the rooted spanning tree process $\{Y_n\}_{n\in\mathbb{Z}}$ from $S$ with root $i$ to $T$ with root $j$, and let $k$ be the unique vertex used in the reveres procedure. We have that $Q_{ST} = p_{ij}$. Remarking that to pass from $T$ to $S$ we first added the edge $(i, j)$ and then deleted the unique directed edge $(j, k)$, and that to pass from $S$ to $T$, we added the directed edge $(j, k)$ and then deleted the edge $(j, i)$. Therefore

$$w(S)Q_{ST} = w(T)R_{TS}$$

where $R_{TS} := p_{jk}$. It follows that

$$\sum_S w(S)Q_{ST} = \sum_S w(T)R_{TS} = w(T)$$

Therefore, the stationary distribution of the chain, $\{\rho(S)\}_{S\in\mathcal{S}}$ is

$$\rho(S) = \frac{w(S)}{\sum_{S'} w(S')}\,,$$

and therefore,

$$\pi(i) = \sum_{T\in\mathcal{S}_i} \rho(T) = \frac{\sum_{T\in\mathcal{S}_i} w(T)}{\sum_{T\in\mathcal{S}} w(T)}\,.$$

□

**Corollary 2.1.1** *Let $\{X_n\}_{n\in\mathbb{Z}}$ be the stationary random walk on the complete graph built on the finite state space $E$. (In particular $p_{ij} = \frac{1}{|E|}$ for all $j \neq i$ and the stationary distribution is the uniform distribution on $E$.) Let $\tau_i = \inf\{n \geq 0\,;\, X_n = i\}$. The directed graph with directed edges*

$$(X_{\tau_i}, X_{\tau_i - 1}), \qquad i \neq X_0$$

*is uniformly distributed over $\mathcal{S}$.*

**Proof.** Use the proof of Theorem 2.1.1 and the time-reversibility of the random walk. □

## 2.2 Martingales and harmonic functions

### Basic examples

Let $\{X_n\}_{n\geq 0}$ be a sequence of discrete random variables[1].

**Definition 2.2.1** *A real-valued stochastic process $\{Y_n\}_{n\geq 0}$ such that for each $n \geq 0$*

*(i) $Y_n$ is a function of $X_0^n := (X_0, \ldots, X_n)$, and*

*(ii) $E[|Y_n|] < \infty$ or $Y_n \geq 0$,*

*is called a* martingale *(resp.,* submartingale, supermartingale*) with respect to $\{X_n\}_{n\geq 0}$ if, moreover,*

$$E[Y_{n+1} \mid X_0^n] = Y_n \,(\textit{resp.},\ \geq Y_n, \leq Y_n). \tag{2.1}$$

For short, we sometimes say "$X_0^n$-martingale" for "martingale with respect to $\{X_n\}_{n\geq 0}$", with similar abbreviations for supermartingales and submartingales.

Observe that a martingale is a submartingale *and* a supermartingale.

EXAMPLE 2.2.1: SUMS OF IID. Let $X = \{X_n\}_{n\geq 0}$ be a sequence of *centered and integrable* IID random variables. The stochastic process

$$Y_n = X_0 + X_1 + \cdots + X_n, \quad n \geq 1$$

is a $X_0^n$–martingale. Indeed, for all $n \geq 1$,

$$E[Y_{n+1} \,|\, X_0^n] = E[Y_n \,|\, X_0^n] + E[X_{n+1} \,|\, X_0^n] = Y_n + E[X_{n+1}] = Y_n\,,$$

where the second equality is due to the facts that $Y_n$ is a function of $X_0^n$ (Theorem 1.1.3) and that $X_0^n$ and $X_{n+1}$ are independent (Theorem 1.1.4).

EXAMPLE 2.2.2: PRODUCTS OF IID. Let $X = \{X_n\}_{n\geq 0}$ be a sequence of integrable IID random variables with mean 1. The stochastic process

$$Y_n = \prod_{k=0}^{n} X_k, \quad n \geq 0$$

[1]We are dealing in these notes with discrete random variables, and therefore the mention "discrete" will be generally omitted.

is a $X_0^n$–martingale. Indeed

$$E[Y_{n+1} \mid X_0^n] = E\left[X_{n+1} \prod_{k=0}^{n} X_k \mid X_0^n\right] = E[X_{n+1} \mid X_0^n] \prod_{k=0}^{n} X_k$$
$$= E[X_{n+1}] \prod_{k=1}^{n} X_k = 1 \times Y_n = Y_n \,,$$

where the second equality is due to the fact that $\prod_{k=0}^{n} X_k$ is a function of $X_0^n$ (Theorem 1.1.3) and the third to the fact that $X_0^n$ and $X_{n+1}$ are independent (Theorem 1.1.4).

---

EXAMPLE 2.2.3: GAMBLING. Consider the stochastic process $\{Y_n\}_{n\geq 0}$ with values in $\mathbb{R}_+$ defined by $Y_0 = a \in \mathbb{R}_+$ and, for $n \geq 0$,

$$Y_{n+1} = Y_n + X_{n+1}\, b_{n+1}(X_0^n),$$

where $\{X_n\}_{n\geq 1}$ is an IID sequence of random variables taking the values $+1$ or $-1$ equiprobably, and where the family of functions $b_n : \{0,1\}^n \to \mathbb{N}$, $n \geq 1$, is a given *betting strategy*, that is, $b_{n+1}(X_0^n)$ is the stake at time $n+1$ of a gambler given the observed history $X_0^n := (X_0, \ldots, X_n)$ of the chance outcomes up to time $n$. The initial conditions are $X_0 = Y_0 = a$. Admissible bets must guarantee that the fortune $Y_n$ remains non-negative at all times $n$, that is $b_{n+1}(X_0^n) \leq Y_n$. The process so defined is a $X_0^n$–martingale. Indeed,

$$E\left[Y_{n+1} \mid X_0^n\right] = E\left[Y_n \mid X_0^n\right] + E\left[X_{n+1} b_{n+1}(X_0^n) \mid X_0^n\right]$$
$$= Y_n + E\left[X_{n+1} \mid X_0^n\right] b_{n+1}(X_0^n) = Y_n \,,$$

where the second equality uses Theorem 1.1.3 and the assumption that $X_{n+1}$ is independent of $X_0^n$ and centered (Theorem 1.1.4).

---

### Harmonic functions of Markov chains

Let $\{X_n\}_{n\geq 0}$ be a HMC on the countable space $E$ with transition matrix $\mathbf{P}$. A function $h : E \to \mathbb{R}$ is called *harmonic* (resp., *subharmonic, superharmonic*) iff

$$\mathbf{P}h = h \,(\text{resp.},\ \geq h, \leq h). \tag{2.2}$$

In developed form, for all $i \in E$,

$$\sum_{j\in E} p_{ij} h(j) = h(i) \,(\text{resp.},\ \geq h(i), \leq h(i)).$$

Superharmonic functions are also called *excessive* functions.

Equation (2.2) is equivalent to

$$E[h(X_{n+1}) \mid X_n = i] = h(i) \,(\text{resp.}, \ \geq h(i), \leq h(i)), \tag{2.3}$$

for all $i \in E$. In view of the Markov property, the left-hand side of the above equality is also equal to

$$E[h(X_{n+1}) \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0],$$

and therefore (2.3) is equivalent to

$$E[h(X_{n+1} \mid X_0^n] = h(X_n) \,(\text{resp.}, \ \leq h(X_n), \geq h(X_n)). \tag{2.4}$$

Therefore, if either $E[|h(X_n)|] < \infty$ for all $n \geq 0$, or $h \geq 0$, the process $\{h(X_n)\}_{n\geq 0}$ is, with respect to $\{X_n\}_{n\geq 0}$, a martingale (resp., submartingale, supermartingale).

### Convex functions of martingales

**Theorem 2.2.1** *Let $I$ be an interval of $\mathbb{R}$ of arbitrary nature (closed, open, semi-closed, infinite, etc.) and let $\phi : I \to \mathbb{R}$ be a convex function.*

*A. Let $Y = \{Y_n\}_{n\geq 0}$ be a $X_0^n$-martingale, such that $P(Y_n \in I) = 1$ for all $n \geq 0$. Assume that $E\left[|\phi(Y_n)|\right] < \infty$ for all $n \geq 0$. Then, the process $\{\phi(Y_n)\}_{n\geq 0}$ is a $X_0^n$–submartingale.*

*B. Assume moreover that $\phi$ is non-decreasing and suppose this time that $Y$ is a $X_0^n$—submartingale. Then, the process $\{\phi(Y_n)\}_{n\geq 0}$ is a $X_0^n$–submartingale.*

**Proof.** By Jensen's inequality for conditional expectations,

$$E\left[\phi(Y_{n+1})|X_0^n\right] \geq \phi(E\left[Y_{n+1}|X_0^n\right]).$$

Therefore (case A)

$$\begin{aligned} E\left[\phi(Y_{n+1})|X_0^n\right] &\geq \phi(E\left[Y_{n+1}|X_0^n\right]) \\ &= \phi(Y_n), \end{aligned}$$

and (case B)

$$\begin{aligned} E\left[\phi(Y_{n+1})|X_0^n\right] &\geq \phi(E\left[Y_{n+1}|X_0^n\right]) \\ &\geq \phi(Y_n). \end{aligned}$$

(For the last inequality, use the submartingale property $E\left[Y_{n+1}|X_0^n\right] \geq Y_n$ and the hypothesis that $\phi$ is non-decreasing.) □

EXAMPLE 2.2.4: POWERS OF MARTINGALES. Let $Y = \{Y_n\}_{n\geq 0}$ be a $X_0^n$-martingale and let $p \geq 1$. Applying Theorem 2.2.1 with the convex function $x \to |x|^p$, we have that if $E\left[|Y_n|^p\right] < \infty$, $\{|Y_n|^p\}_{n\geq 0}$ is a $X_0^n$–submartingale.

EXAMPLE 2.2.5: MODULUS OF A MARTINGALE. Let $Y = \{Y_n\}_{n\geq 0}$ be a $X_0^n$-martingale and let $p \geq 1$. Applying Theorem 2.2.1 with the convex function $x \to x^+$, we have that $\{Y_n^+\}_{n\geq 0}$ is a $X_0^n$–submartingale.

### Optional sampling

Martingale theory rests on two pillars. The first pillar of martingale theory is the *optional sampling theorem*. It has many versions, and that given next is a rather simple one, but already very powerful.

We first recall the definition of a stopping time.

**Definition 2.2.2** *Let $\{X_n\}_{n\geq 0}$ be some sequence of random variables taking their values in $\mathcal{X}$. A randome variable $T$ taking integer values and possibly the value $\infty$ is called a $X_0^n$-stopping time if for all integers $m$ the event $\{T = m\}$ is expressible in terms of $X_0^n$, that is, more precisely, there exists a function $g_m : \mathcal{X}_0^{m+1} \to \{0, 1\}$ such that*

$$1_{\{T=m\}} = g_m(X_0^{m+1}) .$$

**Theorem 2.2.2** *Let $\{M_n\}_{n\geq 0}$ be a $X_0^n$–martingale, and let $T$ be a $X_0^n$-stopping time. Suppose that at least one of the following conditions holds:*

*($\alpha$) $P(T \leq n_0) = 1$ for some $n_0 \geq 0$.*

*($\beta$) $P(T < \infty) = 1$ and $|M_n| \leq K < \infty$ when $n \leq T$.*

*Then*

$$E[M_T] = E[M_0]. \tag{2.5}$$

**Proof.** ($\alpha$) Write

$$M_T - M_0 = \sum_{k=0}^{n_0-1} (M_{k+1} - M_k) 1_{\{k<T\}}.$$

Since $T$ is a stopping time of $\{X_n\}_{n\geq 0}$,

$$1_{\{k<T\}} = \varphi(X_0^k)$$

for some function $\varphi$, and therefore, using the basic rules of conditioning (Theorems 1.1.2 and 1.1.5)

$$\begin{aligned} E[(M_{k+1} - M_k)1_{\{k<T\}}] &= E[(M_{k+1} - M_k)\varphi(X_0^k)] \\ &= E\left[E[(M_{k+1} - M_k)\varphi(X_0^k)|X_0^k)]\right] \\ &= E\left[E[(M_{k+1} - M_k)|X_0^k)]\varphi(X_0^k)\right] = 0. \end{aligned}$$

Therefore,

$$E[M_T - M_0] = \sum_{k=0}^{n_0-1} E[(M_{k+1} - M_k)1_{\{k<T\}}] = 0.$$

($\beta$) Apply the result of ($\alpha$) to the stopping time $T \wedge n_0$ to obtain

$$E[M_{T\wedge n_0}] = E[M_0].$$

Therefore,

$$\begin{aligned} |E[M_T] - E[M_0]| &= |E[M_T] - E[M_{T\wedge n_0}]| \\ &\leq E[|M_T - M_{T\wedge n_0}]| \\ &= E[\sum_{k=n_0+1}^{\infty} |M_k - M_{k\wedge n_0}|\, 1_{\{k=T\}}] \\ &\leq E[\sum_{k=n_0+1}^{\infty} 2K 1_{\{k=T\}}] \\ &= 2KP(T > n_0). \end{aligned}$$

Since $T$ is finite, $\lim_{n_0\uparrow\infty} P(T > n_0) = 0$, and therefore $E[M_T] = E[M_0]$. □

---

EXAMPLE 2.2.6: THE RUIN PROBLEM VIA MARTINGALES. Consider the symmetric random walk $\{X_n\}_{n\geq 0}$ on $\mathbb{Z}$ with $X_0 = 0$. It is a martingale (with respect to itself). Let $T$ be the first time $n$ for which $X_n = -a$ or $+b$, where $a, b > 0$. This is a stopping time, and moreover $T < \infty$. We can apply Theorem 2.2.2 (optional sampling), part ($\beta$), with $K = \sup(a, b)$, to obtain

$$0 = E[X_0] = E[X_T].$$

Writing

$$v = P(-a \text{ is hit before } b),$$

we have

$$E[X_T] = -av + b(1 - v),$$

and therefore

$$v = \frac{b}{a+b}.$$

---

EXAMPLE 2.2.7: A COUNTEREXAMPLE. Consider the symmetric random walk of the previous example, but now define $T$ to be the hitting time of $b > 0$. We know that $T < \infty$, since the symmetric walk on $\mathbb{Z}$ is recurrent. If the optional sampling theorem applied, we would have

$$0 = E[X_0] = E[X_T] = b,$$

an obvious contradiction. The optional sampling theorem (Theorem 2.2.2) does not apply because neither condition $(\alpha)$ nor $(\beta)$ thereof is satisfied.

---

EXAMPLE 2.2.8: THE BALLOT PROBLEM, TAKE 3. In the ballot problem, let $X_k$ be the number of votes in advance (can be negative) for candidate I and define for $0 \leq k \leq n-1$, where $n := a+b$,

$$M_k := \frac{X_{n-k}}{n-k}\,.$$

In Exercise **??**, you are invited to prove that the sequence $M_0, M_1, \ldots, M_{n-1}$ forms a martingale. Admitting this, let $\tau$ be the first $M_0^k$-stopping time $\tau$ at which $X_k = 0$ if such $k$ exists, or $n-1$ otherwise. By the optional sampling theorem (Theorem 2.2.2),

$$E\left[M_\tau\right] = E\left[M_0\right] = \frac{E\left[X_n\right]}{n} = \frac{a-b}{a+b}\,. \qquad (\star)$$

Let $A$ be the event that candidate I leads all the way to victory. If $A$ occurs, then $\tau = n-1$ and $M_\tau = M_{n-1} = X_1$ and $X_1 = 1$ (in this case, candidate I has the first vote). If $A$ does not occut, there is an intermediate time when the candidates have an equal count, and therefore, $\tau$ being the first such time, $\tau < n-1$ and $M\tau = 0$. Therefore

$$E\left[M_\tau\right] = P(A)\,.$$

Comparison with $(\star)$ gives

$$P(A) = \frac{a-b}{a+b}\,.$$

---

### The martingale convergence theorem

We now introduce the second pillar of martingale theory, the *martingale convergence theorem*. This result is the probabilistic counterpart of the convergence of a non-negative non-increasing, or bounded nondecreasing, sequence of real numbers to a finite limit.

**Theorem 2.2.3** *Let $S$ be a $X_0^n$-submartingale, $L_1$-bounded, that is such that*

$$\sup_{n\geq 0} E[|S_n|] < \infty. \tag{2.6}$$

*Then $S$ converges $P$-a.s. to an integrable random variable $S_\infty$.*

The proof is omitted.

**Corollary 2.2.1** *(a) Any non-positive submartingale $S$ converges to an integrable random variable.*
*(b) Any non-negative supermartingale converges to an integrable random variable.*

**Proof.** (b) follows from (a) by changing signs. For (a), we have

$$E[|S_n|] = -E[S_n] \leq -E[S_0] = E[|S_0|] < \infty.$$

Therefore (2.6) is satisfied and the conclusion follows from Theorem 2.2.3. □

We shall apply this result to absorption problems concerning Markov chains.

---

EXAMPLE 2.2.9: BRANCHING PROCESS VIA MARTINGALES. We are going to illustrate the power of the concept of martingale by revisiting the branching process of Section **??** (to which we refer for the notation) and obtaining the results thereof via martingale theory. It is assumed that $P(Z = 0) < 1$ and $P(Z \geq 2) > 0$ (to get rid of trivialities). The stochastic process

$$Y_n = \frac{X_n}{m^n},$$

where $m$ is the average number of sons of a given individual, is a martingale with respect to $\{X_n\}_{n\geq 0}$.

Indeed, each of the $X_n$ members of the $n$th generation gives birth on the average to $m$ sons, and does this independently of the rest. Therefore, $E[X_{n+1}|X_n] = mX_n$, and

$$E\left[\frac{X_{n+1}}{m^{n+1}}|X_0^n\right] = E\left[\frac{X_{n+1}}{m^{n+1}}|X_n\right] = \frac{X_n}{m^n}.$$

By the martingale convergence theorem, almost surely

$$\lim_{n\uparrow\infty} \frac{X_n}{m^n} = Y < \infty.$$

In particular, if $m < 1$, then $\lim_{n\uparrow\infty} X_n = 0$ almost surely. Since $X_n$ takes integer values, this implies that the branching process eventually becomes extinct.

If $m = 1$, then $\lim_{n\uparrow\infty} X_n = X_\infty < \infty$, and it is easily argued that this limit must be 0. Therefore, in this case as well the process eventually becomes extinct.

For the case $m > 1$, we consider the unique solution in $(0, 1)$ of $x = g(x)$ (see Theorem **??**). Suppose we can show that $Z_n = x^{X_n}$ is a martingale. Then, by the martingale convergence theorem, it converges to a finite limit, and therefore $X_n$ has a limit $X_\infty$, which, however, can be infinite. One can easily argue that this limit cannot be other than 0 (extinction) or $\infty$ (nonextinction). Since $\{Z_n\}_{n\geq 0}$ is a martingale, $x = E[Z_0] = E[Z_n]$, and therefore, by dominated convergence, $x = E[Z_\infty] = E[x^{X_\infty}] = P(X_\infty = 0)$. Therefore, $x$ is the probability of extinction.

It remains to show that $\{Z_n\}_{n\geq 0}$ is a martingale. We have

$$E[x^{X_{n+1}}|X_n = i] = x^i.$$

This is obvious if $i = 0$, and if $i > 0$, $X_{n+1}$ is the sum of $i$ independent random variables with the same generating function $g$. Therefore, $E[x^{X_{n+1}}|X_n = i] = g(x)^i = x^i$. From this last result and the Markov property,

$$E[x^{X_{n+1}}|X_0^n] = E[x^{X_{n+1}}|X_n] = x^{X_n}.$$

---

EXAMPLE 2.2.10: A CELLULAR AUTOMATON. Consider a chessboard of size $N \times N$, on which are placed stones, exactly one on each square. Each stone has one among $k$ possible colors. The state $X_n$ of the process at time $n$ is the $N \times N$ matrix with elements in $\{1, \ldots, k\}$ describing the chessboard and the color of the stone in each square. The evolution of $\{X_n\}_{n\geq 0}$ is that of a homogeneous Markov chain, where the transition from $X_n$ to $X_{n+1}$ is as follows. Select one case of the chessboard at random, and change the color of the stone there, the new color being the color of a stone chosen at random among the 4 neighboring stones. To avoid boundary effects, we shall consider that the chessboard is a bi-torus in the sense of the Figure below.

We shall see that the only absorbing states are the monochromatic states and prove, using a martingale argument, that probability of being absorbed in a specific state is equal to the initial proportion of states of this color.

Neighbors in the cellular automaton model

The chain has $2^k - 1$ absorbing classes. Each such class corresponds to a given nonempty subset of $\ell$ different colors. For instance, with $\ell = 3$, we consider all the configurations of the chessboard with a combination of three given colors, say blue, white and red. It is straightforward to verify that one can pass (in several steps) from a configuration with at least one stone of each color, blue, white or red, to any other such configuration. Any monochromatic state is of course closed.

Denote by $Y_n$ the proportion of red stones at stage $n$. The process $\{Y_n\}_{n\geq 0}$ is a martingale with respect to $\{X_n\}_{n\geq 0}$. Indeed, $Y_n$ is a function of $X_n$ and is integrable, since it is bounded by 1. Also, $E[Y_{n+1}|X_0^n] = Y_n$, as the following exchange argument shows.

Let $\alpha_{n+1}$ be the box selected at time $n+1$ and let $\beta_{n+1}$ be the selected neighbor of $\alpha_{n+1}$. Then, for any pair $(\alpha, \beta)$ of boxes, $P(\alpha_{n+1} = \alpha, \beta_{n+1} = \beta|X_0^n) = P(\alpha_{n+1} = \beta, \beta_{n+1} = \alpha|X_0^n) = \frac{1}{8N^2}$. Clearly, if the result $\alpha_{n+1} = \alpha, \beta_{n+1} = \beta$ changes $Y_n$ to $Y_{n+1} = Y_n + \Delta Y_{n+1}$, the result $\alpha_{n+1} = \beta$, $\beta_{n+1} = \alpha$ changes $Y_n$ to $Y_{n+1} = Y_n - \Delta Y_{n+1}$. Since these two situations are equiprobable, the martingale property easily follows.

By the martingale convergence theorem, $\lim_{n\uparrow\infty} Y_n = Y$ exists, and by dominated convergence $E[Y] = \lim_{n\uparrow\infty} E[Y_n]$. Therefore, since $E[Y_n] = E[Y_0]$, we have $E[Y] = E[Y_0] = y_0$, where $y_0$ is the initial proportion of red stones. Because $|\Delta Y_n| = 0$ or $\frac{1}{N^2}$ for all $n$, $\{Y_n\}_{n\geq 0}$ can converge only if it remains constant after some (random) time, and this constant is either 0 or 1. Since the limit 1 corresponds to absorption by the "all-red" state, we see that the probability of being absorbed by the "all-red" state is equal to the initial proportion of red stones.

---

EXAMPLE 2.2.11: FAIR GAME NOT SO FAIR. Consider the situation in Example 2.2.3, assuming that the initial fortune $a$ is a positive integer, and that the bets are also positive integers (that is the functions $b_{n+1}(X_0^n) \in \mathbb{N}_+$ except if $Y_n = 0$, in which case the gambler is not allowed to bet anymore, or equivalently

$b_n(X_0^{n-1}, 0) = 0$). In particular, for all $n \geq 0$, $Y_n \geq 0$. Therefore the process $Y$ is a non-negative $\mathcal{F}_n^X$-martingale, and by the martingale convergence theorem, it almost surely has a finite limit. Since the bets are assumed positive integers when the fortune of the player is positive, this limit cannot be other than 0. Since $Y_n$ is a non-negative integer for all $n \geq 0$, this can happen only if the fortune of the gambler becomes null in finite time.

---

**Theorem 2.2.4** *An irreducible recurrent* HMC *has no non-negative superharmonic or bounded subharmonic functions besides the constant functions.*

**Proof.** If $h$ is non-negative superharmonic (resp., bounded subharmonic), then the stochastic sequence $\{h(X_n)\}_{n\geq 0}$ is a non-negative supermartingale (resp., bounded submartingale), and therefore, by the martingale convergence theorem it converges to a finite limit $Y$. Since $\{X_n\}_{n\geq 0}$ visits any state $i \in E$ infinitely often, one must have $Y = h(i)$ almost surely for all $i \in E$. In paticular, $h$ is a constant. □

**Corollary 2.2.2** *A necessary and sufficient condition for an irreducible* HMC *to be transient is the existence of some state conventionally called 0 and of a bounded function $h : E \to \mathbb{R}$, not identically null and satisfying*

$$h(j) = \sum_{k\neq 0} p_{jk} h(k), \text{ for all } j \neq 0. \tag{2.7}$$

**Proof.** Let $T_0$ be the return time to state 0. First-step analysis shows that the bounded function $h$ defined by

$$h(j) = P_j(T_0 = \infty)$$

satisfies (2.7). If the chain is transient, $h$ is nontrivial (not identically null). This proves necessity.

Conversely, suppose that (2.7) holds for a not identically null bounded function. Define $\tilde{h}$ by $\tilde{h}(0) = 0$ and

$$\tilde{h}(j) = h(j) \text{ if } j \neq 0,$$

and let $\alpha = \sum_{k\in E} p_{0k}\tilde{h}(k)$. Changing the sign of $\tilde{h}$ if necessary, $\alpha$ can be assumed $\geq 0$. Then $\tilde{h}$ is subharmonic and bounded. If the chain were recurrent, then by Theorem 2.2.4, $\tilde{h}$ would be a constant. This constant would be equal to $\tilde{h}(0) = 0$, and this contradicts the assumed nontriviality of $h$. □

---

EXAMPLE 2.2.12: REPAIR SHOP, TAKE 6. The state equation is

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

where $\{Z_n\}_{n\geq 1}$ is an IID sequence independent of the initial state $X_0$. In terms of the probability distribution $P(Z_1 = k) = a_k,\ k \geq 0$, its transition matrix is

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

We assume that the chain is irreducible, that is $P(Z_1 = 0) < 1$ and $P(Z_1 \geq 2) > 0$. We shall show that, if $E[Z_1] > 1$, the system of equations (2.7) admits a bounded nontrivial solution, and therefore, by Corollary 2.2.2, the chain is transient.

Indeed, trying $h(j) = 1 - \zeta^j$ for a solution, we can check that equations (2.7) reduce to a single equation in $\zeta$,

$$\sum_{k\geq 0} P(Z_1 = k)\zeta^k = \zeta, \tag{2.8}$$

for which there is, under condition $E[Z_1] > 1$ and the irreducibility condition, a unique solution $\zeta \in (0, 1)$ (see Theorem **??**).

Therefore, $h(i) = 1 - \zeta^i$ is a solution of (2.7) that is nontrivial and bounded.

---

**Theorem 2.2.5** *Let the* HMC *with transition matrix* $\mathbf{P}$ *be irreducible, and suppose that there exists a function* $h : E \to \mathbb{R}$ *such that* $\{i\ ;\ h(i) < K\}$ *is finite for all finite* $K$*, and such that*

$$\sum_{k\in E} p_{ik}h(k) \leq h(i),\ \textit{for all } i \notin F, \tag{2.9}$$

*for some finite subset* $F \subset E$*. Then the chain is recurrent.*

The conditions of the above result are also necessary (we shall not prove this here), and this is why it is called a criterion. Note that it might then as well be called a transience criterion.

**Proof.** Since $\{i\ ;\ h(i) < 0\}$ is finite, $\inf h(i) > -\infty$, and therefore, adding a constant if necessary, one may assume without loss of generality that $h \geq 0$. Let $\tau = \tau(F)$ be the return time to $F$, and define $Y_n = h(X_n)1_{\{n<\tau\}}$. Equality (2.9) is just $E[h(X_{n+1}) \mid X_n = i] \leq h(i)$ for all $i \notin F$. For $i \notin F$, we have using the basic

rules for conditional expectation (Theorems 1.1.4, 1.1.3 and 1.1.2)

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

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., for $i \notin F$, $P_i$-a.s,

$$E_i[Y_{n+1}|X_0^n] \leq Y_n,$$

that is, $\{Y_n\}_{n\geq 0}$ is, under $P_i$, a non-negative supermartingale with respect to $\{X_n\}_{n\geq 0}$. By the martingale convergence theorem, $\lim_{n\uparrow\infty} Y_n = Y_\infty$ exists and is finite, $P_i$-a.s.

Suppose, in view of contradiction, that the chain is transient. It must then visit any finite subset of the state space only a finite number of times. In particular, for arbitrary $K$, we can have $h(X_n) < K$ only for a finite (random) number of indices $n$. This implies that $\lim_{n\to\infty} h(X_n) = +\infty$, $P_j$-a.s. (for any $j \in E$). For this to be compatible with the fact that $\{1_{\{n<\tau\}}h(X_n)\}$ has $P_i$-a.s. a finite limit for $i \notin F$, we must have $P_i(\tau < \infty) = 1$.

In summary, $P_i(\tau < \infty) = 1$ for all $i \notin F$. Since $F$ is finite, some state in $F$ must be recurrent, hence the announced contradiction. □

---

EXAMPLE 2.2.13: REPAIR SHOP, TAKE 7. We know that this HMC is positive recurrent only if $E[Z_1] < 1$, and that it is transient if $E[Z_1] > 1$. We now 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. Indeed, one easily verifies that Theorem 2.2.5 applies with $h(i) = i$ and $F = \{0\}$. Therefore, the chain is recurrent. Since it is not positive recurrent, it is null-recurrent.

---

Here is another application of the martingale convergence theorem in the vein of the previous results.

**Theorem 2.2.6** *Let the* HMC $\{X_n\}_{n\geq 0}$ *with transition matrix* $\mathbf{P}$ *be irreducible and let* $h : E \to \mathbb{R}$ *be a bounded function such that*

$$\sum_{k\in E} p_{ik}h(k) \leq h(i), \text{ for all } i \notin F, \tag{2.10}$$

*for some set* $F$*, not assumed finite. Suppose, moreover, that there exists* $i \notin F$ *such that*

$$h(i) < h(j), \text{ for all } j \in F. \tag{2.11}$$

*Then the chain is transient.*

**Proof.** Let $\tau$ be the return time in $F$ and let $i \notin F$ satisfy (2.11). Defining $Y_n = h(X_{n\wedge\tau})$, we have that, under $P_i$, $\{Y_n\}_{n\geq 0}$ is a (bounded) supermartingale (exercise) with respect to $\{X_n\}_{n\geq 0}$. By the martingale convergence theorem, the limit $Y$ of $Y_n = h(X_{n\wedge\tau})$ exists and is finite, $P_i$-almost surely. By bounded convergence, $E_i[Y] = \lim_{n\uparrow\infty} E_i[Y_n]$, and since $E_i[Y_n] \leq E_i[Y_0] = h(i)$ (supermartingale property), we have $E_i[Y] \leq h(i)$.

If $\tau$ were $P_i$-a.s. finite, then $Y_n$ would eventually be frozen at a value $h(j)$ for $j \in F$, and therefore by (2.11), $E_i[Y] > h(i)$, a contradiction with the last inequality.

Therefore, $P_i(\tau < \infty) < 1$, which means that with a strictly positive probability, the chain starting from $i \notin F$ will not return to $F$. This is incompatible with irreducibility and recurrence. □

### The maximum principle

Let $\{X_n\}_{n\geq 0}$ be an HMC with countable state space $E$ and transition matrix $\mathbf{P}$. Let $D$ be an arbitrary subset of $E$, called the *domain*. The complement $\overline{D}$ of $D$ in $E$ will be called the *boundary*. Let $c : D \to \mathbb{R}$ and $\varphi : \overline{D} \to \mathbb{R}$ be non-negative functions called the *unit time cost* and the *final cost*, respectively. Let $T$ be the hitting time of $\overline{D}$.

For each state $i \in E$, define

$$h(i) = E_i\left[\sum_{0\leq k<T} c(X_k) + \varphi(X_T)1_{\{T<\infty\}}\right]. \tag{2.12}$$

The function $h : E \to \overline{\mathbb{R}}$ so defined is non-negative and possibly infinite. It is called the *average cost*. Note that $T$ is not required to be finite, and that $\overline{D}$ may be empty.

**Theorem 2.2.7** *The function $h : E \to \overline{\mathbb{R}}_+$ defined by (2.12)*

*(i) is non-negative and satisfies*

$$h = \begin{cases} \mathbf{P}h + c & \text{on } D, \\ \varphi & \text{on } \overline{D}, \end{cases} \tag{2.13}$$

*(ii) and is majored by any non-negative function $u : E \to \overline{\mathbb{R}}$ such that*

$$u \geq \begin{cases} \mathbf{P}u + c & \text{on } D, \\ \varphi & \text{on } \overline{D}. \end{cases} \tag{2.14}$$

*(iii) Moreover, if for all $i \in E$, $P_i(T < \infty) = 1$, then (2.13) has at most one non-negative bounded solution.*

**Proof.** (i) Properties $h \geq 0$ and $h = \varphi$ on $\overline{D}$ are satisfied by definition. First-step analysis gives, for $i \in D$,

$$h(i) = c(i) + \sum_{j \in E} p_{ij} h(j) \tag{2.15}$$

(rely on intuitive arguments or see the details after the proof).

(ii) Define for $n \geq 0$ the non-negative function $h_n : E \to \mathbb{R}$ by

$$h_n(i) = E_i \left[ \sum_{k=0}^{n-1} c(X_k) 1_{\{k<T\}} + \varphi(X_T) 1_{\{T<n\}} \right]. \tag{2.16}$$

Observe that $h_0 \equiv 0$ and, by monotone convergence, $\lim_{n \uparrow \infty} \uparrow h_n = h$. Also, with a proof similar to that of (i),

$$h_{n+1} = \begin{cases} \mathbf{P}h_n + c & \text{on } D, \\ \varphi & \text{on } \overline{D}. \end{cases} \tag{2.17}$$

With $u$ as in (2.14), we have $u \geq h_0$. By induction, $u \geq h_n$ (this is true for $n = 0$, and if this true for some $n$, it is true for $n+1$. Indeed $u \geq \mathbf{P}u + c \geq \mathbf{P}h_n + c = h_{n+1}$ on $D$, and $u \geq \varphi = h_{n+1}$ on $\overline{D}$). Therefore, $u \geq \lim_{n \to \infty} h_n = h$.

(iii) If $u$ is bounded and non-negative, then by Example 2.2,

$$M_n = u(X_n) - u(X_0) - \sum_{k=0}^{n-1} (\mathbf{P} - I) u(X_k) \tag{2.18}$$

is a Lévy martingale with respect to $\{X_n\}_{n\geq 0}$. By the optional sampling theorem, for all integers $K \geq 0$, $E_i[M_{T\wedge K}] = E_i[M_0] = 0$, and therefore

$$u(i) = E_i[u(X_{T\wedge K})] - E_i[\sum_{k=0}^{T\wedge K-1} (\mathbf{P} - I)u(X_k)] = E_i[u(X_{T\wedge K}) + \sum_{k=0}^{T\wedge K-1} c(X_k)],$$

since by hypothesis $(I-\mathbf{P})u = c$ on $D$. Since $P_i(T < \infty) = 1$, $\lim_{K\uparrow\infty} E_i[u(X_{T\wedge K}] = E_i[u(X_T)]$ by dominated convergence. But $u(X_T) = \varphi(X_T)$ because $u = \varphi$ on $\overline{D}$. Also, $\lim_{K\uparrow\infty} E_i[\sum_{k=0}^{T\wedge K-1} c(X_k)] = E_i[\sum_{k=0}^{T-1} c(X_k)]$ by monotone convergence. Finally,

$$u(i) = E_i[\sum_{k=0}^{T-1} c(X_k) + \varphi(X_T)] = h(i).$$

*Proof of (2.15)*

Write for $i \in D$,

$$\begin{aligned} v(i) &= E_i[c(X_0) + \sum_{1\leq n<T} c(X_n) + \varphi(X_T)1_{\{T<\infty\}}] \\ &= c(i) + E_i[\sum_{1\leq n<T} c(X_n) + \varphi(X_T)1_{\{T<\infty\}}], \end{aligned}$$

that is,

$$v(i) = c(i) + \sum_{j\in E} E_i[Z1_{\{X_1=j\}}],$$

where

$$Z = \sum_{1\leq n<T} c(X_n) + \varphi(X_T)1_{\{T<\infty\}}.$$

Since $X_0 = i \in D$ implies that $T \geq 1$ on $\{X_0 = i\}$, the random variable $Z$ is a function of $X_1$, $X_2, \ldots$, and therefore, by the Markov property,

$$E_i[Z1_{\{X_1=j\}}] = E[Z \,|\, X_1 = j]p_{ij}.$$

Now, since $T \geq 1$ on $\{X_0 = i\}$ when $i \in D$, the quantity $Z$ in the above calculations can be rewritten as

$$Z = \sum_{0\leq n<T-1} c(Y_n) + \varphi(Y_{T-1})1_{\{T-1<\infty\}},$$

where $Y_n = X_{n+1}$. Also, for the HMC $\{Y_n\}_{n\geq 0}$, $T' = T - 1$ is the hitting time of $\overline{D}$, and therefore

$$E[Z \,|\, X_1 = j] = E[\sum_{0\leq n<T'} c(Y_n) + \varphi(Y_{T'})1_{\{T'<\infty\}} \,|\, Y_0 = j],$$

and this quantity is just $v(j)$, since $\{X_n\}_{n\geq 0}$ and $\{Y_n\}_{n\geq 0}$ have the same transition matrix, and therefore have the same distribution when their initial states are the same. □

Theorem 2.2.7 can be rephrased as follows. The function $h$ given by (2.12) is a minorant of all non-negative solutions of (2.14), and for $u = h$, the inequalities in (2.14) become equalities. Moreover, if $h$ is bounded and $P_i(T < \infty) = 1$ for all $i \in E$, then $h$ is the *unique* solution of (2.13).

## 2.3 The electrical network analogy

Let $G = (V, \mathcal{E})$ be a finite graph, that is, $V$ is a finite collection of *vertices*, or *nodes*, and $\mathcal{E}$ is a subset of (*unordered*) pairs of vertices, denoted by $e = \langle i, j\rangle$ and one then notes $i \sim j$ (or equivalently $j \sim i$) the fact that $i$ and $j$ are the end vertices of a the edge $\langle i, j\rangle$. This graph is assumed *connected*. The edge/branch $e = \langle i, j\rangle$ has a positive number $c_e = c_{ij}$ $(= c_{ji})$ attached to it. In preparation for the electrical network analogy, call $c_e$ the *conductance* of edge $e$, and call its inverse $R_e = \frac{1}{c_e}$ the *resistance* of $e$.

Define a HMC on $E := V$ with transition matrix $\mathbf{P}$

$$p_{ij} = \frac{c_{ij}}{C_i},$$

where

$$C_i = \sum_{j\in V} c_{ij}.$$

The homogeneous Markov chain introduced in this way is called a *random walk* on the graph $G = (V, \mathcal{E})$. We shall occasionally call it the *network Markov chain*. The state $X_n$ at time $n$ is interpreted as the position on the set of vertices of a particle at time $n$. When on vertex $i$ the particle chooses to move to an adjacent vertex $j$ with a probability proportional to the conductance of the corresponding edge, that is with probability $p_{ij} = \frac{c_{ij}}{C_i}$. Note that this HMC is irreducible since the graph $G$ is assumed connected and the conductances are positive. Moreover, it is reversible with stationary probability

$$\pi(i) = \frac{C_i}{\sum_{j\in V} C_j}.$$

To see this, it suffices to check the reversibility equations

$$\pi(i)\frac{c_{ij}}{C_i} = \pi(j)\frac{c_{ji}}{C_j},$$

using the hypothesis that $c_{ij} = c_{ji} = c_e$. A *symmetric random walk* on the graph $G = (V, \mathcal{E})$ is a particular random walk for which $c_e \equiv 1$ (or any constant). In this case, at any given time, the particle being in a given site chooses at random the adjacent site where it will move. The corresponding stationary probability then takes the form

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

where $d_i$ is the degree of node $i$ (the number of nodes to which it is connected) and $|E|$ is the number of edges.

The connection between random walks and reversible HMC's is in fact both ways. Given a reversible irreducible positive recurrent transition matrix $\mathbf{P} = \{p_{ij}\}_{i,j\in V}$ on $V$ with stationary probability $\pi$, we may define the conductance of edge $e = \langle i, j\rangle$ by $c_{ij} = \pi(i)p_{ij}$ $(= c_{ji}$ by reversibility) and define in this way a random walk with the same transition matrix. In particular $C_i = \pi(i)$ and $p_{ij} = \frac{c_{ij}}{C_i}$.

The graph $G$ will now be interpreted as an electric *network* where electricity flows along the edges of the graph (the "branches" of the electrical network). By convention, if $i \not\sim j$, $c_{ij} = 0$. To each *oriented* pair $(i, j)$ there is associated a potential difference $\Phi_{ij}$ and a current $I_{ij}$ which are real numbers and satisfy the antisymmetry conditions

$$I_{ji} = -I_{ij} \text{ and } \Phi_{ji} = -\Phi_{ij}$$

for all edges $\langle i, j\rangle$. Two distinct nodes will play a particular role: the *source* $a$ and the *sink* $b$.

### Kirchoff's laws and Ohm's law

The potential differences follow *Kirchoff's potential law*: For any sequence of vertices $i_1, i_2, \ldots, i_{n+1}$ such that $i_{n+1} = i_1$ and $i_k \sim i_{k+1}$ for all $1 \le k \le n$,

$$\sum_{\ell=1}^{n} \Phi_{i_\ell, i_{\ell+1}} = 0\,.$$

They also follow *Kirchoff's current law*: For all nodes $i \in V$, $i \neq a, b$,

$$\sum_{j\in V} I_{ij} = 0\,.$$

Finally, the currents and potentials are linked by *Ohm's law*: For all edges $e = \langle i, j\rangle$

$$I_{ij} = c_e \Phi_{ij}\,.$$

It readily follows from Kirchoff's potential law that there exists a function $\Phi : V \to \mathbb{R}$ determined up to an additive constant such that

$$\Phi_{ij} = \Phi(j) - \Phi(i)\,.$$

Note that, by Ohm's law, the current $I_{ij}$ and the potential difference $\Phi(j) - \Phi(i)$ have the same sign ("currents flow in the direction of increasing potential").

Using Ohm's law, Kirchoff's potential law can be expressed in terms of currents:

$$\sum_{\ell=1}^{n} \frac{I_{i_\ell, i_{\ell+1}}}{c_{i_\ell, i_{\ell+1}}} = 0\,.$$

In particular both Kirchoff's laws are linear in the currents. Therefore we have the superposition principle:

**Theorem 2.3.1** *If $I$ and $I'$ are solutions of Kirchoff's laws with the same source $a$ and sink $b$, then so is their sum $I + I'$.*

From Kirchoff's current law and Ohm's law, we have that for all $i \neq a, b$,

$$\sum_{i \in V} c_{ij}(\Phi(j) - \Phi(i)) = 0\,,$$

or equivalently

$$\Phi(i) = \sum_{j \in V} \frac{c_{ij}}{C_i} \Phi(j)\,.$$

Therefore the potential function $\Phi$ is, with respect to this transition matrix, harmonic on $V \backslash \{a, b\}$. In particular, by Theorem 2.2.7, it is uniquely determined by its boundary values $\Phi(a)$ and $\Phi(b) = 0$.

### Probabilistic interpretation of voltage and current

We start with the voltage. Recall that $\Phi$ is harmonic on $D = V \backslash \{a, b\}$ and that if we fix its values on $\{a, b\}$, it is then uniquely determined. We call $\Phi_1$ the solution corresponding to a unit voltage at $a$ and a null voltage at $b$:

$$\Phi_1(a) = 1\,, \qquad \Phi_1(b) = 0\,.$$

The function $h$ given by

$$h(i) = P_i\,(T_a < T_b)$$

(the probability that starting from $i$, $a$ is reached before $b$) is harmonic on $D = V \backslash \{a, b\}$ and that $h(a) = 1$ and $h(b) = 0$. By unicity, $\Phi_1 \equiv h$.

We now interpret the current. A particle performing the random walk associated to the network moves from $a$ to $b$, but now it is supposed to leave the network once it has reached $b$. We show that the current $I_{ij}$ from $i$ to $j$ is proportional to the expected number of passages of this particle from $i$ to $j$ minus the expected number of passages in the opposite direction, from $j$ to $i$.

**Proof.** Let $u(i)$ be the expected number of visits to node $i$ before it reaches $b$ and leaves the network. Clearly $u(b) = 0$. Also for $i \neq a, b$, $u(i) = \sum_{j\in V} u(j)p_{ji}$. But $C_i p_{ij} = C_j p_{ji}$ so that $u(i) = \sum_{j\in V} u(j)p_{ij}\frac{C_i}{C_j}$ and finally

$$\frac{u(i)}{C_i} = \sum_{j\in V} p_{ij}\frac{u(j)}{C_j}\,.$$

Therefore the function $\Phi$ given by

$$\Phi(i) = \frac{u(i)}{C_i}\,.$$

is harmonic on $D = V\backslash\{a, b\}$. It is the unique such function whose values at $a$ and at $b$ are specified by

$$\Phi(a) = \frac{u(a)}{C_a}\,, \qquad \Phi(b) = 0\,. \qquad \star$$

Whith such voltage function,

$$\begin{aligned} I_{ij} &= (\Phi(i) - \Phi(j))c_{ij} \\ &= \left(\frac{u(i)}{C_i} - \frac{u(j)}{C_j}\right) c_{ij} \\ &= u(i)\frac{c_{ij}}{C_i} - u(j)\frac{c_{ji}}{C_j} = u(i)p_{ij} - u(j)p_{ji}\,. \end{aligned}$$

But $u(i)p_{ij}$ is the expected number of crossings from $i$ to $j$ and $u(j)p_{ji}$ is the expected number of crossings in the opposite direction. □

The above interpretation of currents in terms of edge crossings avails for the specific voltage $\Phi$ considered, not for the "standardized" voltage $\Phi_1$. Both currents are proportional. To determine the factor of proportionality, we may use the fact that, under voltage $\Phi$ determined by $(\star)$,

$$\sum_{j\in V} I_{aj} = 1$$

because, in view of the probabilistic interpretation of current in this case, the sum is equal to the expected value of the difference between the number of times the particle leaves $a$ and the number of times it enters $a$, that is 1 (each time the particle enters $a$ it leaves it immediately after, except for the one time when it leaves $a$ forever to be eventually absorbed in $b$).

### Effective resistance and escape probability

When a voltage $\Phi(a)$ is applied at $a$ and a null voltage at $b$, let $I_a$ be the current flowing out of the source $a$ (and therefore into the sink $b$). The effective resistance between $a$ and $b$ is defined by

$$R_{eff} = \frac{\Phi(a)}{I_a} \tag{2.19}$$

This quantity does not depend on the value $\Phi(a)$ since multiplication the voltage by a factor implies multiplication of the current by the same factor. When $\Phi(a) = \Phi_1(a) = 1$, the effective conductance equals the current $I_a$ flowing from $a$. But in this case

$$\begin{aligned} I_a &= \sum_{j\in V} (\Phi_1(a) - \Phi_1(j)) c_{aj} \\ &= \sum_{j\in V} (1 - \Phi_1(j)) c_{aj} \\ &= C_a \left( 1 - \sum_{j\in V} \Phi_1(j) \frac{c_{aj}}{C_a} \right) \\ &= C_a \left( 1 - \sum_{j\in V} p_{aj} \Phi_1(j) \right) . \end{aligned}$$

But the quantity $\left(1 - \sum_{j\in V} p_{aj}\Phi_1(j)\right)$ is the "escape probability"

$$P_{esc} := P(T_b < T_a)$$

that is the probability that the particle starting from $a$ reaches $b$ before returning to $a$. Therefore

$$C_{eff} = C_a P_{esc} .$$

### Dissipated energy and Thomson's principle

We start with a definition: a *flow* on the network (or, more precisely, on the graph) with source $a$ and sink $b$ is be a collection of real numbers $J = \{J_{ij}\}_{i,j\in V}$, such that

(a) $J_{ij} = -J_{ji}$,

(b) $J_{ij} = 0$ if $i \not\sim j$,

(c) $\sum_{j\in V} J_{ij} = 0$ for all $i \neq a, b$

Denote by $J_i = \sum_{j\in V} J_{ij}$ the *flow out of* $i$. In particular $J_i = 0$ for all $i \neq a, b$. Also

$$J_a = -J_b\,,$$

as the following calculation shows

$$\begin{aligned} J_a + J_b &= \sum_{I\in V} J_i \\ &= \sum_{i,j\in V} J_{ij} = \frac{1}{2}\sum_{i,j\in V}(J_{ij} + J_{ji}) = 0\,. \end{aligned}$$

Finally, for any function $w : V \to \mathbb{R}$,

$$(w(a) - w(b))J_a = \frac{1}{2}\sum_{i,j\in V}(w(j) - w(i))J_{ij}\,, \qquad (\star)$$

Indeed, from the properties of flows

$$\begin{aligned} \sum_{i,j\in V}(w(i) - w(j))J_{ij} &= \sum_{i,j\in V} w(i)J_{ij} - \sum_{i,j\in V} w(j)J_{ij} \\ &= \sum_{i,j\in V} w(i)J_{ij} + \sum_{i,j\in V} w(j)J_{ji} \\ &= \sum_{i\in V} w(i)J_i + \sum_{j\in V} w(j)J_j \\ &= w(a)J_a + w(b)J(b) + w(a)J_a + w(b)J(b) \\ &= w(a)J_a - w(b)J(a) + w(a)J_a - w(b)J(a) = 2(w(a) - w(b))J_a\,. \end{aligned}$$

A *unit flow* $J$ is one for which $J_a = 1$.

These preliminaries given, we introduce the *energy dissipated in the network by the flow*$J$:

$$E(J) = \frac{1}{2}\sum_{i,j\in V} J_{ij}^2 R_{ij}.$$

This is a meaningful electrical quantity for the special case where the flow is a current $I$ corresponding to a potential $\Phi$, in which case, by Ohm's law:

$$E(I) = \frac{1}{2}\sum_{i,j\in V} I_{ij}^2 R_{ij} = \frac{1}{2}\sum_{i,j\in V} I_{ij}(\Phi(j) - \Phi(i)).$$

**Theorem 2.3.2** *The effective resistance between the source and the sink is equal to the energy dissipated in the network when a unit current passes from the source to the sink.*

**Proof.** By $(\star)$,

$$E(I) = (\Phi(a) - \Phi(b))I_a = \Phi(a)I_a\,,$$

and by definition (2.19) of the effective resistance $R_{eff}$ between $a$ and $b$,

$$E(I) = I_a^2 R_{eff}\,,$$

so that, if we adjust the input current to be $I_a = 1$, we have that □

The following result is known as Thomson's principle.

**Theorem 2.3.3** *The energy dissipation $E(J)$ is minimized among all unit flows $J$ by the unit current flow $I$.*

**Proof.** Let $J$ be a unit flow from $a$ to $b$ and let $I$ be a unit current flow from $a$ to $b$. Define $D = J - I$. This is a flow from $a$ to $b$ with $D_a = 0$. We have that

$$\begin{aligned}\sum_{i,j\in V} J_{ij}^2 R_{ij} &= \sum_{i,j\in V} (I_{ij} + D_{ij})^2 R_{ij} \\ &= \sum_{i,j\in V} I_{ij}^2 R_{ij} + 2\sum_{i,j\in V} I_{ij} D_{ij} R_{ij} + \sum_{i,j\in V} D_{ij}^2 R_{ij} \\ &= \sum_{i,j\in V} I_{ij}^2 R_{ij} + 2\sum_{i,j\in V} (\Phi(j) - \Phi(i)) D_{ij} + \sum_{i,j\in V} D_{ij}^2 R_{ij}.\end{aligned}$$

From (**??**) with $w = \Phi$ and $J = D$, the middle term equals $4(\Phi(a) - \Phi(b))D_a = 0$, so that

$$\sum_{i,j\in V} J_{ij}^2 R_{ij} = \sum_{i,j\in V} I_{ij}^2 R_{ij} + \sum_{i,j\in V} D_{ij}^2 R_{ij} \geq \sum_{i,j\in V} I_{ij}^2 R_{ij}\,.$$

□

### Rayleigh's monotonicity law

We can now state and prove Rayleigh's principle:

**Theorem 2.3.4** *The effective resistance between two points can only increase as any resistance in the circuit increases.*

**Proof.** Change the resistances $R_{ij}$ to $\overline{R}_{ij} \geq R_{ij}$ and let $I$ and $\overline{I}$ be the corresponding unit current flows. Then

$$\overline{R}_{eff} = \frac{1}{2}\sum_{i,j\in V} \overline{I}_{ij}^2 \overline{R}_{ij} \geq \frac{1}{2}\sum_{i,j\in V} \overline{I}_{ij}^2 R_{ij}\,.$$

But by Thomson's principle,

$$\frac{1}{2}\sum_{i,j\in V}\overline{I}_{ij}^2 R_{ij} \geq \frac{1}{2}\sum_{i,j\in V} I_{ij}^2 R_{ij} = R_{eff}\,.$$

□

**Recurrence via effective resistance**

Let now $G = (V, \mathcal{E})$ be an *infinite* connected graph with finite-degree vertices, and distinguish some arbitrary vertex, henceforth called 0. Recall that the graph distance $d(i, j)$ between two vertices is the smallest number of edges to be crossed when going from $i$ to $j$. For $N \geq 0$, let

$$K_N = \{i \in V;\, d(0, i) \leq N\}$$

and

$$\partial K_N = K_N - K_{N-1} = \{i \in V;\, d(0, i) = N\}$$

Let $G_N$ be the restriction of $G$ to $K_N$. A graph $\overline{G}_N$ is obtained from $G_N$ by merging the vertices of $\partial K_N$ in a single vertex named $b_N$. Let $R_{eff}(n)$ be the effective resistance between 0 and $b_N$ of the network $\overline{G}_N$. Since $\overline{G}_N$ is obtained from $\overline{G}_{N+1}$ by merging the vertices of $\partial K_N \cup \{b_{N+1}\}$, $R_{eff}(N) \leq R_{eff}(N+1)$. In particular the limit

$$R_{eff} = \lim_{N\uparrow\infty} R_{eff}(N)$$

exists. But it may be infinite. In fact, this gives a criterion of recurrence for the network HMC $\{X_n\}_{n\geq o}$. More precisely:

**Theorem 2.3.5** *The probability of return to* 0 *of the network* HMC *is*

$$P(X_n = 0 \text{ for some } n \geq 1) = 1 - \frac{1}{C_0 R_{eff}}\,.$$

*In particular this chain is recurrent if and only if* $R_{eff} = \infty$.

**Proof.** The formula

$$h_N(i) = P(X_n \text{ hits } K_N \text{ before } 0)$$

defines an harmonic function on $V_N \backslash \{\{0\} \cup K_N\}$ with boundary conditions $h_N(0) = 0$ and $h_N(i) = 1$ for all $i \in K_N$. Therefore, the function $g_N$ defined

$$g_N(i) = h_N(i) \text{ on } K_{N-1} \cup \{b_N\}$$

and $g_N(b_N) = 1$ is a potential function for the network $\overline{G}_N$ with source 0 and sink $b_N$. Therefore

$$\begin{aligned} P_0\left(X_n \text{ returns to 0 befor reaching } \partial K_N\right) &= 1 - \sum_{j\sim 0} p_{0j} g_N(j) \\ &= 1 - \sum_{j\sim 0} \frac{c_{0j}}{C_0}(g_N(j) - g_N(0)). \end{aligned}$$

By Ohm's law, $\sum_{j\sim 0} c_{0j}(g_N(j) - g_N(0))$ is the the total current $I_N(0)$ out of 0, and therefefore since the potential difference between $b_N$ and 0 is 1, $I_N(0) = \frac{1}{R_{eff}(N)}$. Therefore

$$P_0\left(X_n \text{ returns to 0 before reaching } \partial K_N\right) = 1 - \frac{1}{C_0 R_{eff}(N)}$$

and the result follows since by the sequential continuity of probability

$$P(X_n = 0 \text{ for some } n \geq 1) = \lim_{N\uparrow\infty} P_0\left(X_n \text{ returns to 0 before reaching } \partial K_N\right) .$$

□

## 2.4 Gibbs fields

Markov fields are so called Gibbs fields in honour of the Josiah Willard Gibbs, the father of Statistical Mechanics. The physicist Ernst Ising used them as a simplified model of ferromagnetism, and yet another physicist, Rudolf Peierls, showed that such mathematical model could predict the phase transition phenomenon.

**Definition 2.4.1** *Let $G = (V, \mathcal{E})$ be a finite graph, and let $v_1 \sim v_2$ denote the fact that $(v_1, v_2)$ is an edge of the graph. We shall also refer to elements of $V$ as* sites. *Let $\Lambda$ be a finite set, the* phase space. *A* random field *on $V$ with phases in $\Lambda$ is a collection $X = \{X(v)\}_{v\in V}$ of random variables $X(v)$ with values in $\Lambda$.*

A random field can be regarded as a random variable taking its values in the *configuration space* $E := \Lambda^V$. A configuration $x \in \Lambda^V$ is of the form $x = (x(v), v \in V)$, where $x(v) \in \Lambda$ for all $v \in V$. For a given configuration $x$ and a given subset $A \subseteq S$, define

$$x(A) = (x(v), v \in A),$$

the restriction of $x$ to $A$. If $S\backslash A$ denotes the complement of $A$ in $V$, one writes $x = (x(A), x(V\backslash A))$. In particular, for fixed $v \in V$, $x = (x(v), x(V\backslash v))$, where $S\backslash v$ is a shorter way of writing $V\backslash\{v\}$.

Of special interest are the random fields characterized by *local interactions*. This leads to the notion of Markov random field. The "locality" is in terms of the neighbourhood structure inherited from the graph structure. More precisely, for any $v \in V$, $N_v := \{w \in V; w \sim v\}$ is the neighborhhod of $v$. In the following, $\widetilde{\mathcal{N}}_v$ denotes the set $\mathcal{N}_v \cup \{v\}$.

**Definition 2.4.2** *The random field $X$ is called a* Markov random field (MRF) *with respect to the neighborhood system $N$ if for all sites $v \in V$, the random variables $X(v)$ and $X(V \backslash \widetilde{\mathcal{N}}_v)$ are independent given $X(\mathcal{N}_v)$.*

In mathematical symbols:

$$P(X(v) = x(v) \mid X(V \backslash v) = x(S \backslash v)) = P(X(v) = x(v) \mid X(\mathcal{N}_v) = x(\mathcal{N}_v)) \quad (2.20)$$

for all $x \in \Lambda^V$ and all $v \in V$. Property (2.20) is clearly of the Markov type: the distribution of the phase at a site is directly influenced only by the phases of the neighboring sites.

Note that any random field is Markovian with respect to the trivial topology, where the neighborhood of any site is the whole set $V$. However, the interesting Markov fields (from the point of view of modeling, simulation, and optimization) are those with relatively small neighborhoods.

**Definition 2.4.3** *The* local characteristic *of the* MRF *at site $v$ is the function $\pi^s : \Lambda^V \to [0, 1]$ defined by*

$$\pi^v(x) = P(X(v) = x(v) \mid X(\mathcal{N}_v) = x(\mathcal{N}_v)).$$

*The family $\{\pi^v\}_{v \in V}$ is called the* local specification *of the* MRF.

One sometimes writes

$$\pi^v(x) = \pi(x(v) \mid x(\mathcal{N}_v))$$

in order to stress the role of the neighborhood system.

We say that a MRF satisfies the *positivity condition* if its probability distribution is strictly positive.

**Theorem 2.4.1** *Two distributions of an* MRF *with a* finite *configuration space $\Lambda^V$ that satisfy the positivity condition and have the same local specification are identical.*

**Proof.** Enumerate $V$ as $\{1, 2, \ldots, K\}$. Therefore $x = (x_1, \ldots, x_{K-1}, x_K) \in \Lambda^K$. The following identity

$$\pi(x) = \prod_{i=1}^{K} \frac{\pi(x_i \mid x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_K)}{\pi(y_i \mid x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_K)} \pi(y) \qquad (\star)$$

holds for any $x, y \in \Lambda^K$. For the proof, one checks that

$$\pi(x) = \frac{\pi(x_K \mid x_1, \ldots, x_{K-1})}{\pi(y_K \mid x_1, \ldots, x_{K-1})} \pi(x_1, \ldots, x_{K-1}, y_K)$$

by simply rewriting the conditional probabilities thereof using Bayes's definition. Similarly,

$$\pi(x_1, \ldots, x_{K-1}, y_K) = \frac{\pi(x_{K-1} \mid x_1, \ldots, x_{K-2}, y_K)}{\pi(y_{K-1} \mid x_1, \ldots, x_{K-2}, y_K)} \pi(x_1, \ldots, x_{K-2}, y_{K-1}, y_K)$$

and so forth. The above calculations make sense because of the positivity condition.

Let now $\pi$ and $\pi'$ be probability distributions with the same local specifications and satisfying the positivity condition. Choose any $y$. Identity $(\star)$ shows that for all $x$,

$$\frac{\pi'(x)}{\pi(x)} = \frac{\pi'(y)}{\pi(y)}$$

a constant, necessarily equal to 1 since $\pi$ and $\pi'$ are probability distributions. □

### Cliques, potential, and Gibbs distributions

Consider the probability distribution

$$\pi_T(x) = \frac{1}{Z_T} e^{-\frac{1}{T}U(x)} \tag{2.21}$$

on the configuration space $\Lambda^V$, where $T > 0$ is the *temperature*, $U(x)$ is the *energy* of configuration $x$, and $Z_T$ is the normalizing constant, called the *partition function*. Since $\pi_T(x)$ takes its values in $[0, 1]$, necessarily $-\infty < U(x) \leq +\infty$. Note that $U(x) < +\infty$ if and only if $\pi_T(x) > 0$. One of the challenges associated with Gibbs models is obtaining explicit formulas for averages, considering that it is generally hard to compute the partition function.

Such distributions are interesting for physicists when the energy is expressed in terms of a potential function describing the local interactions. The notion of clique then plays a central role.

**Definition 2.4.4** *Any singleton $\{v\}$ is a clique. A subset $C \subseteq V$ with more than one element is called a* clique *if and only if any two distinct sites of $C$ are mutual neighbors. A clique $C$ is called* maximal *if for any site $v$, $C \cup \{v\}$ is* not *a clique.*

**Definition 2.4.5** *A* Gibbs potential *on* $\Lambda^V$ *relative to the neighborhood system* $N$ *is a collection* $\{V_C\}_{C\subseteq V}$ *of functions* $V_C : \Lambda^V \to \mathbb{R}\cup\{+\infty\}$ *such that*

*(i)* $V_C \equiv 0$ *if* $C$ *is not a clique,*

*(ii) for all* $x, x' \in \Lambda^V$ *and all* $C \subseteq V$,

$$(x(C) = x'(C)) \Rightarrow (V_C(x) = V_C(x'))\,.$$

*The energy function* $U : \Lambda^V \to \mathbb{R}\cup\{+\infty\}$ *is said to* derive from the potential *$\{V_C\}_{C\subseteq V}$ if*

$$U(x) = \sum_C V_C(x)\,.$$

The function $V_C$ depends only on the phases at the sites inside subset $C$. One could write more explicitly $V_C(x(C))$ instead of $V_C(x)$, but this notation will not be used.

In this context, the distribution in (2.21) is called a *Gibbs distribution*.

---

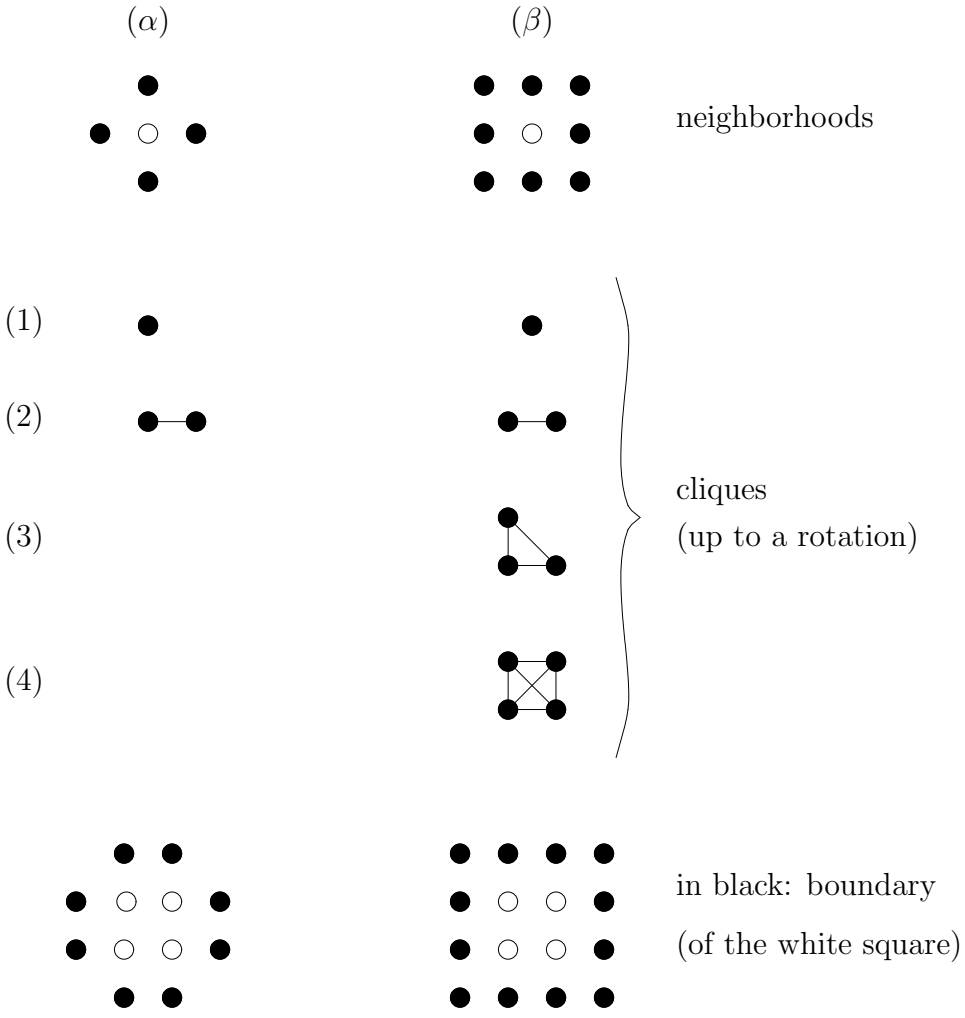
EXAMPLE 2.4.1: ISING MODEL, TAKE 1. In statistical physics, the following model is regarded as a qualitatively correct idealization of a piece of ferromagnetic material. Here $V = \mathbb{Z}_m^2 = \{(i,j) \in \mathbb{Z}^2, i,j \in [1,m]\}$ and $\Lambda = \{+1,-1\}$, where $\pm 1$ is the orientation of the magnetic spin at a given site. The figure below depicts two particular neighborhood systems, their respective cliques, and the boundary of a $2\times 2$ square for both cases.
In Ising's *finite* model, $V = \mathbb{Z}_m^2$, $\Lambda = \{+1,-1\}$, and the neighborhood system is as in $(\alpha)$ of Fig. 7.1.1. The Gibbs potential is

$$\begin{aligned} V_{\{v\}}(x) &= -\frac{H}{k}x(v), \\ V_{\langle v,w\rangle}(x) &= -\frac{J}{k}x(w)x(v), \end{aligned}$$

where $\langle v,w\rangle$ is the 2-element clique $\{v,w\}$, where $w \in \mathcal{N}_v$. Here, $k$ is the Boltzmann constant, $H$ is the external magnetic field, and $J$ is the internal energy of an elementary magnetic dipole. The energy function corresponding to this potential is therefore

$$U(x) = -\frac{J}{k}\sum_{\langle v,w\rangle} x(v)x(w) - \frac{H}{k}\sum_{v\in V} x(v).$$

---




Two examples of neighborhoods, cliques, and boundaries

### The Markov–Gibbs equivalence

Gibbs distributions with an energy deriving from a Gibbs potential relative to a neighborhood system are distributions of Markov fields relative to the same neighborhood system. This result admits a (partial) converse, Theorem 2.4.3 below.

**Theorem 2.4.2** *If $X$ is a random field with a distribution $\pi$, of the form (2.21), where the energy $U(x)$ derives from a Gibbs potential $\{V_C\}_{C\subset V}$ relative to the graph structure $G=(V,\mathcal{E})$, then $X$ is Markovian relatively to the same graph structure. Moreover, its local specification is given by the formula*

$$\pi^v(x)=\frac{e^{-\sum_{C\ni v}V_C(x)}}{\sum_{\lambda\in\Lambda}e^{-\sum_{C\ni v}V_C(\lambda,x(V\backslash v))}}, \tag{2.22}$$

*where the notation $\sum_{C\ni v}$ means that the sum extends over the sets $C$ that contain the site $v$.*

**Proof.** First observe that the right-hand side of (2.22) depends on $x$ only through $x(v)$ and $x(\mathcal{N}_v)$. Indeed, $V_C(x)$ depends only on $(x(w), w\in C)$, and for a clique $C$, if $w\in C$ and $v\in C$, then either $w=v$ or $w\sim v$. Therefore, if one can show that $P(X(v)=x(v)|X(V\backslash v)=x(V\backslash v))$ equals the right-hand side of (2.22), and in particular is a function of $x(v)$ and $x(\mathcal{N}_v)$ only, then the Markov property (2.20) and equality (2.22) will be proven. By definition of conditional probability,

$$P(X(v)=x(v)\mid X(V\backslash v)=x(S\backslash v))=\frac{\pi(x)}{\sum_{\lambda\in\Lambda}\pi(\lambda,x(V\backslash v))}\,. \tag{$\dagger$}$$

But

$$\pi(x)=\frac{1}{Z}e^{-\sum_{C\ni v}V_C(x)+\sum_{C\not\ni v}V_C(x)},$$

and similarly,

$$\pi(\lambda,x(V\backslash v))=\frac{1}{Z}e^{-\sum_{C\ni v}V_C(\lambda,x(V\backslash v))-\sum_{C\not\ni v}V_C(\lambda,x(S\backslash v))}.$$

If $C$ is a clique and $v$ is not in $C$, then $V_C(\lambda,x(V\backslash v))=V_C(x)$ and is therefore independent of $\lambda\in\Lambda$. Therefore, the righthand side of ($\dagger$) is found, after factoring out $\exp x(w)\left\{-\sum_{C\not\ni v}V_C(x)\right\}$, to be equal to the righthand side of (2.22). □

The *local energy* at site $v$ of configuration $x$ is

$$\mathcal{E}_v(x)=\sum_{C\ni v}V_C(x).$$

With this notation, (2.22) becomes

$$\pi^v(x) = \frac{e^{-\mathcal{E}_v(x)}}{\sum_{\lambda\in\Lambda} e^{-\mathcal{E}_v(\lambda, x(V\setminus v))}}.$$

EXAMPLE 2.4.2: ISING MODEL, TAKE 2. The local characteristics in the Ising model are

$$\pi_T^v(x) = \frac{e^{\frac{1}{kT}\left\{J\sum_{w;|w-v|=1} x(w)+H\right\}x(v)}}{e^{+\frac{1}{kT}\left\{J\sum_{w\sim v} x(w)+H\right\}} + e^{-\frac{1}{kT}\left\{J\sum_{w\sim v} x(w)+H\right\}}}.$$

Theorem 2.4.2 above is the direct part of the *Gibbs–Markov equivalence* theorem: A Gibbs distribution relative to a neighborhood system is the distribution of a Markov field with respect to the same neighborhood system. The converse part (Hammersley–Clifford theorem) is important from a theoretical point of view, since together with the direct part it concludes that Gibbs distributions and MRF's are essentially the same objects.

**Theorem 2.4.3** *Let $\pi$ be the distribution of a Markov random field with respect to the graph structure $G = (V, \mathcal{E})$ and satisfying the positivity condition. Then*

$$\pi(x) = \frac{1}{Z} e^{-U(x)}$$

*for some energy function $U$ deriving from a Gibbs potential $\{V_C\}_{C\subseteq V}$ associated with the same graph structure $G = (V, \mathcal{E})$.*

The proof is omitted. However, quite often in practice, the potential as well as the topology of $V$ can be obtained directly from the expression of the energy.

EXAMPLE 2.4.3: MARKOV CHAINS AS MARKOV FIELDS. Let $V = \{0, 1, \dots N\}$ and $\Lambda = E$, a finite space. A random field $X$ on $V$ with phase space $\Lambda$ is therefore a vector $X$ with values in $E^{N+1}$. Suppose that $X_0, \dots, X_N$ is a homogeneous Markov chain with transition matrix $\mathbf{P} = \{p_{ij}\}_{i,j\in E}$ and initial distribution $\nu = \{\nu_i\}_{i\in E}$. In particular, with $x = (x_0, \dots, x_N)$,

$$\pi(x) = \nu_{x_0} p_{x_0 x_1} \cdots p_{x_{N-1} x_N},$$

that is,

$$\pi(x) = e^{-U(x)},$$

where

$$U(x) = -\ln \nu_{x_0} - \sum_{n=0}^{N-1} (\ln p_{x_i x_{i+1}}).$$

Clearly, this energy derives from a Gibbs potential associated with the nearest-neighbor topology for which the cliques are, besides the singletons, the pairs of adjacent sites. The potential functions are:

$$V_{\{0\}}(x) = -\ln \nu_{x_0}, \quad V_{\{n,n+1\}}(x) = -\ln p_{x_n x_{n+1}}.$$

The local characteristic at site $n$, $2 \leq n \leq N-1$, can be computed from formula (2.22), which gives

$$\pi^n(x) = \frac{\exp(\ln p_{x_{n-1}x_n} + \ln p_{x_n x_{n+1}})}{\sum_{y \in E} \exp(\ln p_{x_{n-1}y} + \ln p_{y x_{n+1}})},$$

that is,

$$\pi^n(x) = \frac{p_{x_{n-1}x_n} p_{x_n x_{n+1}}}{p^{(2)}_{x_{n-1}x_{n+1}}},$$

where $p^{(2)}_{ij}$ is the general term of the two-step transition matrix $\mathbf{P}^2$. Similar computations give $\pi^0(x)$ and $\pi^N(x)$. We note that, in view of the neighborhood structure, for $2 \leq n \leq N-1$, $X_n$ is independent of $X_0, \ldots, X_{n-2}, X_{n+2}, \ldots, X_N$ given $X_{n-1}$ and $X_{n+1}$.

---

**Natural sampling in Physics**

We shall see that, according to the theory of Statistical Physics, Nature performs in a natural (of course!) way Monte Carlo method. As a matter of fact, the basic ideas of MMC was first proposed by physicists. Although only the "static" aspects of Gibbs fields were presented so far, in Nature, the dynamical aspects are sometimes essential. For instance, the orientation of the magnetic spins in a ferromagnet vary constantly in time, and the macroscopic properties are in fact those of the statistical equilibrium. The following mathematical development captures the essential aspects of Statistical Physics.

Consider a random field that changes randomly with time. We then have a stochastic process $\{X_n\}_{n \geq 0}$, where

$$X_n = (X_n(v), v \in V)$$

and $X_n(v) \in \Lambda$. The state at time $n$ of this process is a random field on $V$ with phases in $\Lambda$, or equivalently, a random variable with values in the state space $E = \Lambda^V$, which for simplicity we assume finite. The stochastic process $\{X_n\}_{n \geq 0}$ will be called a *dynamical random field*.

The purpose of the current subsection is to show how a given random field probability distribution

$$\pi(x) = \frac{1}{Z} e^{-U(x)} \tag{2.23}$$

can arise as the stationary distribution of a field-valued Markov chain. According to the general principles of the Monte Carlo Markov chain method, we just need to exhibit an *irreducible aperiodic* Markov chain $\{X_n\}_{n\geq 0}$ with state space $E = \Lambda^V$ and stationary distribution (2.23).

The *Gibbs sampler* uses a strictly positive probability distribution $(q_v, v \in V)$ on $V$, and the transition from $X_n = x$ to $X_{n+1} = y$ is made according to the following rule. The new state $y$ is obtained from the old state $x$ by changing (or not) the value of the phase at *one site only*. The site $s$ to be changed (or not) at time $n$ is chosen independently of the past with probability $q_v$. When site $v$ has been selected, the current configuration $x$ is changed into $y$ as follows: $y(V\backslash v) = x(V\backslash v)$, and the new phase $y(v)$ at site $v$ is selected with probability $\pi(y(v) \mid x(V\backslash v))$. Thus, configuration $x$ is changed into $y = (y(v), x(V\backslash v))$ with probability $q_v\pi(y(v) \mid x(V\backslash v))$, according to the local specification at site $v$. This gives for the nonzero entries of the transition matrix

$$P(X_{n+1} = y \mid X_n = x) = q_v\pi(y(v) \mid x(V\backslash v))1_{y(V\backslash v)=x(V\backslash v)}. \tag{2.24}$$

Suppose that the corresponding chain is irreducible and aperiodic. To prove that $\pi$ is the stationary distribution, we use the detailed balance test (Theorem 1.2.9). This test suggests to check that the detailed balance equations do hold true. For this to be true, we must have for all $x, y \in \Lambda^V$,

$$\pi(x)P(X_{n+1} = y \mid X_n = x) = \pi(y)P(X_{n+1} = x \mid X_n = y),$$

that is, in view of (2.24), for all $v \in V$,

$$\pi(x)q_v\pi(y(v) \mid x(V\backslash v)) = \pi(y)q_v\pi(x(v) \mid x(V\backslash v)).$$

This is indeed so, since the last equality reduces to the identity

$$\pi(x)q_v\frac{\pi(y(v), x(S\backslash v)}{P(X(V\backslash v) = x(V\backslash v))} = \pi(y(v), x(V\backslash v))q_v\frac{\pi(x)}{P(X(V\backslash v) = x(V\backslash v))}.$$

---

EXAMPLE 2.4.4: ISING MODEL, TAKE 3: WHAT MAGNETS DO. In the Ising model (Examples 2.4.1 and 2.4.1), the local specification at site $s$ depends only on the local configuration $x(\mathcal{N}_v)$. Note that small neighborhoods speed up computations. Note also that the Gibbs sampler is a natural sampler, in the sense that in a piece

of ferromagnetic material, for instance, the spins are randomly changed according to the local specification. When nature decides to update the orientation of a dipole, it does so according to the law of statistical mechanics. It computes the local energy

$$\mathcal{E}(x(v), x(\mathcal{N}_v)) = x(v)\left(\frac{J}{k}\sum_{w\sim v} x(w) + \frac{H}{k}\right)$$

for each of the two possible spins, that is $\mathcal{E}_+ = \mathcal{E}(+1, x(\mathcal{N}_v))$ and $\mathcal{E}_- = \mathcal{E}(-1, x(\mathcal{N}_v))$, and takes the corresponding orientation with a probability proportional to $e^{-\mathcal{E}_+}$ and $e^{-\mathcal{E}_-}$, respectively, according to the fundamental law of statistical mechanics.

## 2.5 Phase transition in the Ising model

Consider the slightly generalized Ising model of a piece of ferromagnetic material, with spins distributed according to

$$\pi_T(x) = \frac{1}{Z_T} e^{\frac{-U(x)}{T}}. \tag{2.25}$$

The finite site space is enumerated as $S = \{1, 2, \ldots, N\}$, and therefore a configuration $x$ is denoted by $(x(1), x(2), \ldots, x(N))$. The energy function has two terms,

$$U(x) = \mathcal{E}_0(x) - \frac{H}{k}\sum_{i=1}^{N} x(i),$$

where the term $\mathcal{E}_0(x)$ is assumed symmetric, in the sense that for any configuration $x$,

$$\mathcal{E}_0(x) = \mathcal{E}_0(-x).$$

The constant $H$ is the external magnetic field. The *magnetic moment* of configuration $x$ is

$$m(x) = \sum_{i=1}^{N} x(i),$$

and the *magnetization* is the average *magnetic moment* per site

$$M(H, T) = \frac{1}{N}\sum_{x\in E} \pi_T(x) m(x).$$

In Exercise **??**, you are invited to check that

$$\frac{\partial M(H, T)}{\partial H} \geq 0.$$

Also, it is clear that

$$M(-H,T) = -M(H,T)$$

and

$$-1 \leq M(H,T) \leq +1.$$

In summary, at fixed temperature $T$, the magnetization $M(H,T)$ is a nondecreasing odd function of $H$ with values in $[-1,+1]$. Also,

$$M(0,T) = 0\,, \qquad (\diamond)$$

since for any configuration $x$, $m(-x) = -m(x)$, and therefore $\pi_T(-x) = \pi_T(x)$ when $H = 0$. Moreover, the magnetization is an analytic function of $H$.

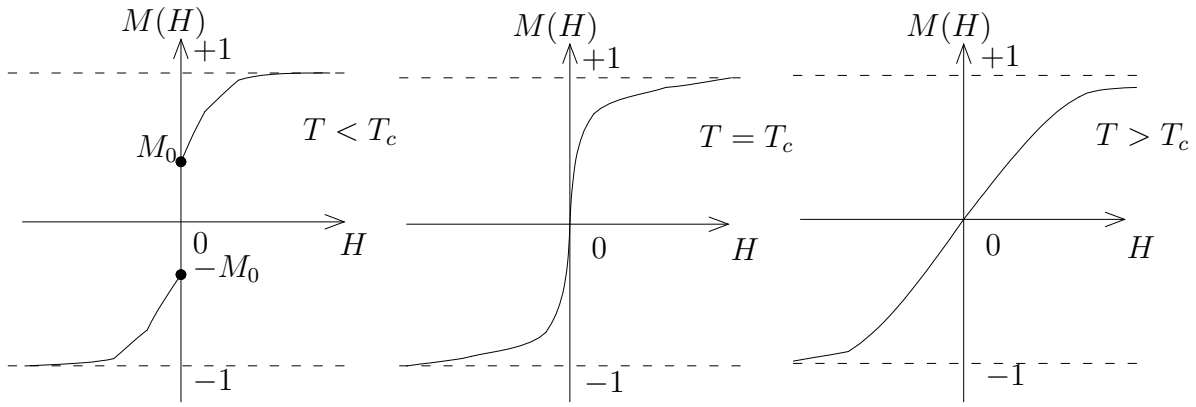
However, the experimental results seem to contradict the last two assertions. Indeed, if an iron bar is placed in a strong magnetic field $H$ parallel to the axis, it is completely magnetized with magnetization $M(H,T) = +1$, and if the magnetic field is slowly decreased to 0, the magnetization decreases, but tends to a limit $M(0,T) = M_0 > 0$, in disagreement with $(\diamond)$. By symmetry, we therefore have a discontinuity of the magnetization at $H = 0$ (see the figure below (a)), in contradiction to the theoretical analyticity of the magnetization as a function of $H$.

This discontinuity is called a *phase transition* by physicists, by analogy with the discontinuity in density at a liquid–gas phase transition. It occurs at room temperature, and if the temperature is increased, the residual, or *spontaneous*, magnetization $M_0$ decreases until it reaches the value 0 at a certain temperature $T_c$, called the *critical temperature*. Then for $T > T_c$, the discontinuity at 0 disappears, and the magnetization curve is smooth (see the figure below (c)). At $T = T_c$, the slope at $H = 0$ is infinite, i.e., the *magnetic susceptibility* is infinite (see the figure below (b)).

The discrepancy with experience and theory observed below the critical temperature is due to the fact that the experimental results describe a situation at the *thermodynamical limit* $N = \infty$. For fixed but large $N$ the theoretical magnetization curve is analytic, but it presents for all practical purposes the same aspect as in Figure a below.

To summarize the experimental results, it seems that below the critical temperature, the spontaneous magnetization has, when no external magnetic field is applied, two "choices." We shall now explain this phenomenon within the classical Ising model.

Consider the Ising model in the absence of external field ($H = 0$). The energy of




Magnetization and critical temperature

a configuration $x$ is of the form

$$U(x) = -J \sum_{\langle v,w \rangle} x(v)x(w),$$

where $\langle v, w \rangle$ represents an unordered pair of neighbors. When the cardinal of the site space $V$ is infinite, the sum in the expression of the energy is not defined for all configurations, and therefore one cannot define the Gibbs distribution $\pi_T$ on $\Lambda^V$ by formula (2.25). However, the local specification

$$\pi_T^v(x) = \frac{e^{\beta \sum_{\langle v,w \rangle} x(v)x(w)}}{e^{\beta \sum_{\langle v,w \rangle} x(w)} + e^{-\beta \sum_{\langle v,w \rangle} x(w)}}, \tag{2.26}$$

where $\beta$ is, up to a factor, the inverse temperature, is well-defined for all configurations and all sites.

In the sequel, we shall repeatedly use abbreviated notation. For instance, if $\pi$ is the distribution of a random field $X$ under probability $P$, then $\pi((x(A))$ denotes $P(X(A) = x(A))$, $\pi(x(0) = +1)$ denotes $P(X(0) = +1)$, etc.

A probability distribution $\pi_T$ on $\Lambda^V$ is called a solution of the DLR problem if it admits the local specification (2.26)[2].

When $V = K_N = \mathbb{Z}^2 \cap [-N, +N]^2$, we know that there exists a unique solution, given by (2.25). When $S = \mathbb{Z}^2$, Dobrushin has proven the existence of at least one solution of the DLR problem. One way of constructing a solution is to select an

[2]DLR stands for Dobrushin (1965), and Lanford and Ruelle (1969).

arbitrary configuration $z$, to construct the unique probability distribution $\pi_T^{(N)}$ on $\Lambda^V$ such that

$$\pi_T^{(N)}(z(V\backslash K_{N-1})) = 1$$

(the field is frozen at the configuration $z$ outside $K_{N-1}$) and such that the restriction of $\pi_T^{(N)}$ to $K_{N-1}$ has the required local characteristics, given by (2.26), and then let $N$ tend to infinity. A solution $\pi_T$ of the DLR problem corresponding to the configuration $z$ selected is obtained as follows. For all configurations $x$ and all *finite* subsets $A \subset V$, the following limit exists:

$$\pi_T(x(A)) = \lim_{N\uparrow\infty} \pi_T^{(N)}(x(A)), \tag{2.27}$$

and moreover, there exists a unique random field $X$ with the local specification (2.26) and such that, for all configurations $x$ and all *finite* subsets $A \subset V$,

$$P(X(A) = x(A)) = \pi_T(x(A)).$$

Note that $\pi_T^{(N)}$ depends on the configuration $z$ only through the restriction of $z$ to the boundary $K_N\backslash K_{N-1}$.

If the DLR problem has more than one solution, we say that a *phase transition* occurs. The method given by Dobrushin to construct a solution suggests a way of proving phase transition when it occurs. It suffices to select two configurations $z_1$ and $z_2$, and to show that for a given finite subset $A \subset S$, the right-hand side of (2.27) is different for $z = z_1$ and $z = z_2$. It has been proven by Peierls (1936) that for sufficiently small values of the temperature, phase transition occurs. Peierls applied the above program with $z_1$ being the configuration with all spins positive and $z_2$ the all negative configuration, and with $A = \{0\}$, where 0 denotes the central site of $\mathbb{Z}^2$.

Denote then by $\pi_+^{(N)}$ (resp., $\pi_-^{(N)}$) the restriction to $K_N$ of $\pi_T^{(N)}$ when $z = z_1$ (resp., $z = z_2$). We shall prove that if $T$ is large enough, then $\pi_+^{(N)}(x(0) = -1) < \frac{1}{3}$ for all $N$. By symmetry, $\pi_-^{(N)}(x(0) = +1) < \frac{1}{3}$, and therefore $\pi_-^{(N)}(x(0) = -1) > \frac{2}{3}$. Passing to the limit $N \uparrow \infty$, we see that $\pi_+(x(0) = -1) < \frac{1}{3}$ and $\pi_-(x(0) = -1) > \frac{2}{3}$, and therefore, the limiting distributions are not identical.

We now proceed to the execution of the above program. For all $x \in \Lambda^{K_N}$,

$$\pi_+^{(N)}(x) = \frac{e^{-2\beta n_o(x)}}{Z_+^{(N)}}\ , \tag{2.28}$$

where $n_o(x)$ is the number of *odd bounds* in configuration $x$, that is, the number of cliques $\langle v, w\rangle$ such that $x(v) \neq x(w)$, and where $Z_+^{(N)}$ is the normalization factor.

To obtain (2.28), it suffices to observe that

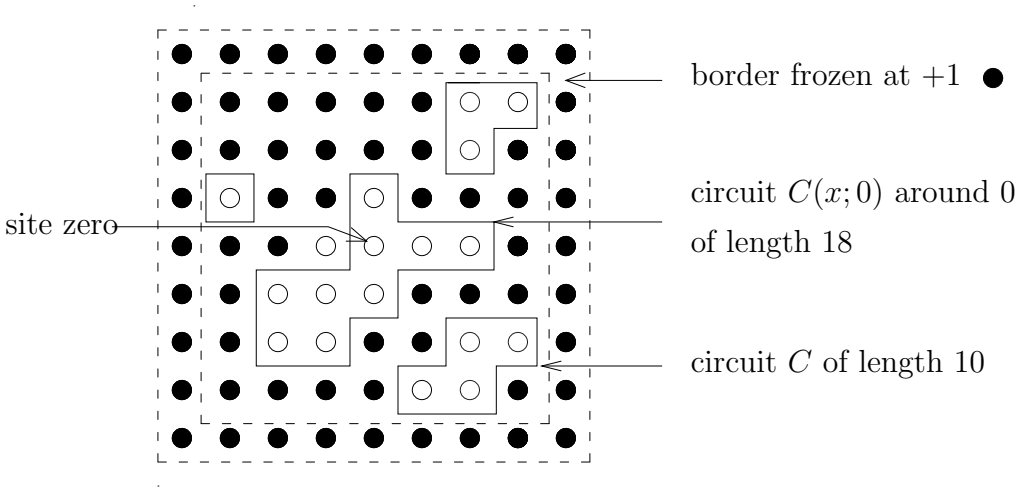
$$-\sum_{\langle v,w\rangle} x(v)x(w) = n_o(x) - n_e(x),$$

where $n_e(x)$ is the number of even bounds, and that $n_e(x) = M - n_o(x)$, where $M$ is the total number of pair cliques. Therefore,

$$U(x) = 2\beta n_o(x) - M,$$

from which (2.28) follows.

Before proceeding to the proof of the announced upper bound for $\pi_+^{(N)}(x(0) = -1)$, a few definitions are needed. Actually, no formal definition will be proposed; instead, the reader is referred to pictures. The figure below features *circuits* $C$ of various lengths.




Circuits in the Ising model

For a given configuration $x$, $C(x;0)$ denotes the circuit which is the boundary of the largest connected batch of sites with negative phases, containing site 0. It is a *circuit around* 0. If the phase at the central site is positive, then $C(x;0)$ is the empty set.

For a given configuration $x$, denote by $\widetilde{x}$ the configuration obtained by reversing all the phases inside circuit $C(x;0)$. For a given circuit $C$ around 0,

$$\pi_+^{(N)}(C(x;0) = C) = \frac{\sum_{x;C(x;0)=C} e^{-2\beta n_o(x)}}{\sum_y e^{-2\beta n_o(y)}}.$$

But

$$\sum_z e^{-2\beta n_o(z)} \geq \sum_{y;C(y;0)=C} e^{-2\beta n_o(\widetilde{y})}$$

(one can always associate to a configuration $y$ such that $C(y;0) = C$ the configuration $z = \widetilde{y}$, and therefore the sum on the right-hand side is a subsum of the left-hand side). Therefore,

$$\pi_+^{(N)}(C(x;0) = C) \leq \frac{\sum_{x;C(x;0)=C} e^{-2\beta n_o(x)}}{\sum_{x;C(x;0)=C} e^{-2\beta n_o(\widetilde{x})}}.$$

If $x$ is such that $C(x;0) = C$, then $n_0(\widetilde{x}) = n_0(x) - L$, where $L$ is the length of $C$, and therefore

$$\pi_+^{(N)}(C(x;0) = C) \leq e^{-2\beta L}.$$

In particular,

$$\pi_+^{(N)}(x(0) = -1) \leq \sum r(L) e^{-2\beta L},$$

where the latter summation is over all lengths $L$ of circuits around 0, and $r(L)$ is the number of nonempty circuits around 0 of length $L$. The possible lengths are $4, 6, \ldots, 2f(N)$, where $f(N) \uparrow \infty$ as $N \uparrow \infty$. In order to bound $r(L)$ from above, observe that a circuit around 0 of length $L$ must have at least one point at a distance smaller than or equal to $\frac{L}{2}$ of the central site 0. There are $L^2$ ways of selecting such a point, and then at most 4 ways of selecting the segment of $C$ starting from this point, and then at most 3 ways of selecting the next connected segment, and so on, so that

$$r(L) \leq 4L^2 3^L.$$

Therefore,

$$\pi_+^{(N)}(x(0) = -1) \leq \sum_{L=4,6,\ldots} 4L^2 (3e^{-2\beta})^L.$$

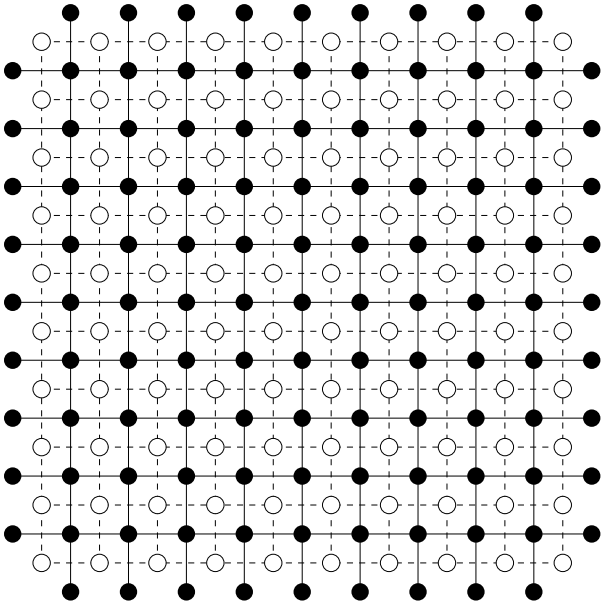
Now, the series $\sum_{L=4,6,\ldots} L^2 x^L$ has a radius of convergence not less than 1, and therefore, if $3e^{-\beta}$ is small enough, or equivalently if $T$ is large enough, $\pi_+^{(N)}(x(0) = -1) < \frac{1}{3}$ for all $N$.

## 2.6 Percolation

Consider the set $V = \mathbb{Z}^2$ (the *(infinite) grid* or *(infinite) lattice*) whose elements are called *nodes* or *vertices*. Let $E_{NN}$ the collection of *nearest-neighbour potential edges*, that is the collection of all unordered pairs $\langle u, v \rangle$ of mutually adjacent vertices[3]. A *percolation graph* on $\mathbb{Z}^2$ is, by definition, a graph $G = (V, E)$ where

[3]Vertex $u = (i, j)$ has 4 adjacent vertices $v = (i+1, j), (i-1, j), (i, j+1), (i, j-1)$.

$V = \mathbb{Z}^2$ and $E$ is a subset of $E_{NN}$. The graph $(V, E_{NN})$ is called the *fully connected* percolation graph. The *dual grid* (or lattice) in two dimensions $V'$ is the original grid $V = \mathbb{Z}^2$ shifted by $(\frac{1}{2}, \frac{1}{2})$ (its vertices are of the form $(i', j') = (i + \frac{1}{2}, j + \frac{1}{2})$).



In a given percolation graph $G$, a *path* from vertex $u$ to vertex $v$ is, by definition, a sequence of vertices $v_0, v_1, \ldots, v_m$ such that $u = v_0$ and $v = v_m \neq u$, and for all $0 \leq i \leq m - 1$, $\langle v_i, v_{i+1} \rangle$ is an edge of $G$. Note that in this definition the extremities $u$ and $v$ must be different. Such path is called *loop-free* if the vertices thereof are distinct. If in addition there is an edge linking $u$ and $v$, the sequence $v_0 = u, v_1, \ldots, v_m = v, u$ is called a *circuit* (we insist that it has to be loop-free to be so called).

The *random percolation graph* $G_p$ on $\mathbb{Z}^2$, where $p \in [0, 1]$, is a random element taking its values in the set of percolation graphs on $\mathbb{Z}^2$, and whose collection $E_p$ of edges is randomly selected according to the following procedure. Let be given a collection $\{U_{\langle u,v \rangle}\}_{\langle u,v \rangle \in E_{NN}}$ of IID random variables uniformly distributed on $[0, 1]$, called the *random generators*. Then the potential edge $\langle u, v \rangle$ is included in $E_p$ (becomes an edge of $G_p$) if and only if $U_{\langle u,v \rangle} \leq p$. Thus, a potential edge becomes an edge of $G_p$ with probability $p$ independently of all other potential edges. The specific procedure used to implement this selection allows to construct all the random percolation graphs simultaneously, using the same collection of random generators. In particular, if $0 \leq p_1 < p_2 \leq 1$, $G_{p_1} \subseteq G_{p_2}$, by which it is meant that $E_{p_1} \subseteq E_{p_2}$.

Two vertices $u$ and $v$ of a given percolation graph are said to be *in the same component*, or to be *connected*, if there exists a path of this graph connecting them. A *component* of the percolation graph is a set $C$ of mutually connected vertices such that no vertex outside $C$ is connected to a vertex in $C$. Its cardinal is denoted by $|C|$. Denote by $C(G,u)$ the component of the percolation graph containing vertex $u$.

We now introduce the notion of a *dual percolation graph*. The dual percolation graph of a given percolation graph $G$ on $V=\mathbb{Z}^2$ is a percolation graph on the dual grid $V'$ which has an edge linking two adjacent vertices $u'$ and $v'$ if and only if this edge does not cross an edge of $G$. We call $G'$ such a graph. In particular $G'_p$ is the dual random percolation graph of the random percolation graph $G_p$.

### Subcritical and supercritical percolation

*Percolation* is said to occur in a given percolation graph if there exists an infinite component. The fundamental result of this section is:

**Theorem 2.6.1** *There exists a* critical value $p_c \in [\frac{1}{3},\frac{2}{3}]$ *such that the probability that $G_p$ percolates is null if $p<p_c$ (*the subcritical case*), and equal to 1 if $p>p_c$ (*the supercritical case*).*

The proof will be given after some preliminaries. We start with a trivial observation concerning $C(G_p,0)$ (0 stands for $(0,0)$, the origin of $\mathbb{Z}^2$). Defining

$$\theta(p)=P(|C(G_p,0)|=\infty),$$

the probability that the origin is in an unbounded component of the random percolation graph $G_p$, we have that $\theta(0)=0$ and $\theta(1)=1$. Next, $\theta:[0,1]\to[0,1]$ is non-decreasing. Indeed if $0\le p_1<p_2\le 1$, $G_{p_1}\subseteq G_{p_2}$, and therefore $|C(G_{p_1},0|)=\infty$ implies $|C(G_{p_2},0|)=\infty$. This remark provides an opportunity to introduce the notions of *increasing set* and *increasing function* defined on the set of percolation graphs.

**Definition 2.6.1** *A set $A$ of percolation graphs is called non-decreasing if for all percolation graphs $G^{(1)},G^{(2)}$ such that $G^{(1)}\subseteq G^{(2)}$, $G^{(1)}\in A$ implies that $G^{(2)}\in A$. A function $f$ taking its values in the set of percolation graphs on $\mathbb{Z}^2$ is called non-decreasing if $G^{(1)}\subseteq G^{(2)}$, $G^{(1)}\in A$ implies that $f(G^{(1)})\le f(G^{(2)})$.*

In particular $1_A$ is a non-decreasing function whenever $A$ is a non-decreasing set.

EXAMPLE 2.6.1: The event $\{|C(G,0)|=+\infty\}$ is a non-decreasing event. So is the event "there is a path in $G$ fom a given vertex $u$ to a given vertex $v$".

In very much the same way as we proved the non-decreasingness of the function $\theta$, one can prove the following result.

**Lemma 2.6.1** *If $A$ is a non-decreasing event, then the function $p \to P(G_p \in A)$ is non-decreasing. If $f$ is a non-decreasing function, then the function $p \to E[f(G_p)]$ is non-decreasing.*

Theorem 2.6.1 will be obtained as a consequence of the following lemma.

**Lemma 2.6.2** *There exists a* critical value *$p_c \in [\frac{1}{3}, \frac{2}{3}]$ such that $\theta(p) = 0$ if $p < p_c$, and $\theta(p) > 0$ if $p > p_c$.*

We now show how Lemma 2.6.2 has Theorem 2.6.1 for consequence. Let $A$ be the non-decreasing event "there exists an infinite component". The random variable $1_A(G_p)$ does not depend on any finite subset of the collection of independent variables $\{U_{\langle u,v\rangle}\}_{\langle u,v\rangle \in E_{NN}}$. By Kolmogorov's 0–1-law, $P(G_p \in A)$ can take only one of the values 0 or 1. Since on the other hand $P(G_p \in A) \geq \theta(p)$, $\theta(p) > 0$ implies $P(G_p \in A) = 1$. Also, by the union bound,

$$\begin{aligned} P(G_p \in A) &\leq \sum_{u \in \mathbb{Z}^2} P(|C(G_p, u)| = +\infty) \\ &= \sum_{u \in \mathbb{Z}^2} P(|C(G_p, 0)| = +\infty) = \sum_{u \in \mathbb{Z}^2} \theta(p)\,, \end{aligned}$$

and therefore, $\theta(p) = 0$ implies that $P(G_p \in A) = 0$.

What remains to be done is the proof of Lemma 2.6.2.

**Part 1**. We show that for $p < \frac{1}{3}$, $\theta(p) = 0$. Call $\sigma(n)$ the number of loop-free paths starting from 0 and of length $n$. Such path can be constructed progressively edge by edge starting from the origin. For the first edge (from 0) there are 4 choices, and for each of the $n-1$ remaining edges there are at most 3 choices. Hence the bound

$$\sigma(n) \leq 4 \times 3^{n-1}\,.$$

We order these $\sigma(n)$ paths arbitrarily.

Let $N(n, G)$ be the number of paths of length $n$ starting from 0 in a percolation graph $G$. If there exists in $G_p$ an infinite path starting from 0 (or equivalently, if there exists an infinite component of $G_p$ containing the origin) then, for each $n$ there exists at least one path of length $n$ starting from 0, that is,

$$\{|C(G_p, 0)| = \infty\} = \cap_{n=1}^{\infty} \{N(n, G_p) \geq 1\}$$

and therefore, for all $n \geq 1$,

$$\theta(p) \leq P(N(n, G_p) \geq 1) = P\left(\cup_{i=1}^{\sigma(n)} \{Y_i(G_p) = 1\}\right)$$

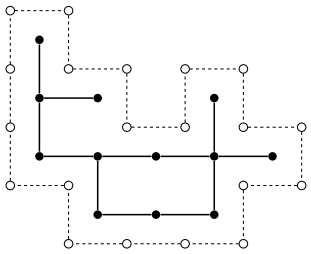
where $Y_i(G_p)$ is the indicator function for the presence in $G_p$ of the $i$-th loop-free path of length $n$ starting from 0 in the fully connected percolation graph. Therefore, by the union bound,

$$\theta(p) \leq \sum_{i=1}^{\sigma(n)} P\left(Y_i(G_p) = 1\right) \leq \sigma(n) p^n = 4p(3p)^{n-1}\,.$$

If $p < \frac{1}{3}$, this quantity tends to 0 as $n \uparrow +\infty$.

**Part 2**. We show that for $p > \frac{2}{3}$, $\theta(p) > 0$.

The statement that $|C(G_p, 0)| < \infty$ is equivalent to saying that 0 is surrounded by a circuit of $G'_p$.



Call $\rho(n)$ the number of circuits of length $n$ of the fully connected dual graph that surround the origin of the original grid. We have that

$$\rho(n) \leq n\sigma(n-1)\,,$$

which accounts for the fact that such circuits contain *at most* a path of length $n-1$ that passes through a dual vertex of the form $(\frac{1}{2}, \frac{1}{2} + i)$ for some $0 \leq i < n$.

The set $\mathcal{C}$ of circuits of the fully connected dual percolation graph that surround the origin 0 of the original grid is countable. Denote by $\mathcal{C}_k \subset \mathcal{C}$ the subset of such circuits that surround the box $B_k \subset S = \mathbb{Z}^2$ of side length $k$ centered at the origin 0. Call $\Delta(B_k)$ the boundary of $B_k$. The two following statements are equivalent:

(i) There is no circuit of $\mathcal{C}_k$ that is a circuit of $G'_p$,

(ii) There is at least one vertex $u \in \Delta(B_k)$ with $|C(G_p, u)| = \infty$.

Therefore

$$\begin{aligned} P\left(\cup_{u\in\Delta(B_k)}\{|C(G_p,u)|=\infty\}\right) &= P\left(\cap_{c\in\mathcal{C}_k}\{c \text{ is not a circuit of } G'_p\}\right) \\ &= 1-P\left(\cup_{c\in\mathcal{C}_k}\{c \text{ is a circuit of } G'_p\}\right) \\ &\geq 1-\sum_{c\in\mathcal{C}_k} P\left(\{c \text{ is a circuit of } G'_p\}\right) . \end{aligned} \tag{2.29}$$

A given circuit of length $n$ occurs in the dual random percolation graph $G'_p$ with probability $(1-p)^n$ and therefore

$$\sum_{c\in\mathcal{C}_k} P\left(\{c \text{ is a circuit of } G'_p\}\right) \leq \sum_{n=4k}^{\infty} n\sigma(n-1)(1-p)^n \leq \frac{4}{9}\sum_{n=4k}^{\infty}(3(1-p))^n n\,. \tag{2.30}$$

If $p>\frac{2}{3}$, the series $\sum_{n=1}^{\infty}(3(1-p))^n n$ converges, and therefore, for $k$ large enough, $\frac{4}{9}\sum_{n=4k}^{\infty}(3(1-p))^n n<1$. From this and (2.29), we obtain that for large enough $k$,

$$P\left(\cup_{u\in\Delta(B_k)}\{|C(G_p,u)|=\infty\}\right)>0$$

which implies that $P(|C(G_p,0)|=\infty)>0$ since there is a positive probability that there exists in $G_p$ a path from the origin to any point of the boundary of $B_k$.

# Index