Part I

Preliminaries, Examples and Motivations

1
Finite Markov chains

1.1 Preliminaries and notation

Let X be a finite set and denote by $L(X) = \{f : X \to \mathbb{C}\}$ the vector space of all complex-valued functions defined on X. Clearly $\dim L(X) = |X|$, where $|\cdot|$ denotes cardinality.

For $x \in X$ we denote by δ_x the *Dirac function* centered at x, that is

$$\delta_x(y) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{if } y \neq x. \end{cases}$$

The set $\{\delta_x : x \in X\}$ is a natural basis for $L(X)$ and if $f \in L(X)$ then $f = \sum_{x \in X} f(x)\delta_x$.

The space $L(X)$ is endowed with the scalar product defined by setting

$$\langle f_1, f_2 \rangle = \sum_{x \in X} f_1(x)\overline{f_2(x)}$$

for $f_1, f_2 \in L(X)$, and we set $\|f\|^2 = \langle f, f \rangle$. Note that the basis $\{\delta_x : x \in X\}$ is orthonormal with respect to $\langle \cdot, \cdot \rangle$. Sometimes we shall write $\langle \cdot, \cdot \rangle_{L(X)}$ to emphasize the space where the scalar product is defined if other spaces are also considered.

If $Y \subseteq X$, the symbol $\mathbf{1}_Y$ denotes the *characteristic function* of Y:

$$\mathbf{1}_Y(x) = \begin{cases} 1 & \text{if } x \in Y \\ 0 & \text{if } x \notin Y; \end{cases}$$

in particular, if $Y = X$ we write $\mathbf{1}$ instead of $\mathbf{1}_X$.

For $Y_1, Y_2, \ldots, Y_m \subseteq X$ we write $X = Y_1 \coprod Y_2 \coprod \cdots \coprod Y_m$ to indicate that the Y_j's constitute a *partition* of X, that is $X = Y_1 \cup Y_2 \cup \cdots \cup Y_m$ and $Y_i \cap Y_j = \emptyset$ whenever $i \neq j$. In other words the symbol $\coprod$ denotes a *disjoint union*. In particular, if we write $Y \coprod Y'$ we implicitly assume that $Y \cap Y' = \emptyset$.

If $A : L(X) \to L(X)$ is a linear operator, setting $a(x, y) = [A\delta_y](x)$ for all $x, y \in X$, we have that

$$[Af](x) = \sum_{y \in X} a(x, y) f(y) \tag{1.1}$$

for all $f \in L(X)$ and we say that the *matrix* $a = (a(x, y))_{x,y \in X}$, *indexed* by X, *represents* the operator A.

If the linear operators $A_1, A_2 : L(X) \to L(X)$ are represented by the matrices a_1 and a_2, respectively, then the composition $A_1 \circ A_2$ is represented by the corresponding product of matrices $a = a_1 \cdot a_2$ that is

$$a(x, y) = \sum_{z \in X} a_1(x, z) a_2(z, y).$$

For $k \in \mathbb{N}$ we denote by $a^k = \left(a^{(k)}(x, y)\right)_{x,y \in X}$ the product of k copies of a, namely

$$a^{(k)}(x, y) = \sum_{z \in X} a^{(k-1)}(x, z) a(z, y).$$

We remark that (1.1) can be also interpreted as the product of the matrix a with the column vector $f = (f(x))_{x \in X}$.

Given a matrix a and a column or, respectively, a row vector f, we denote by a^T and by f^T the transposed matrix (i.e. $a^T(x, y) = a(y, x)$ for all $x, y \in X$) and the row, respectively column transposed vector. This way we also denote by $f^T A$ the function given by

$$[f^T A](y) = \sum_{x \in X} f(x) a(x, y). \tag{1.2}$$

With our notation, the identity operator is represented by the identity matrix which may be expressed as $I = (\delta_x(y))_{x,y \in X}$. If X is a set of cardinality $|X| = n$ and $k \leq n$, then a k-subset of X is a subset $A \subseteq X$ such that $|A| = k$.

If $v_1, v_2, \ldots, v_m$ are vectors in a vector space V, then $\langle v_1, v_2, \ldots, v_m \rangle$ will denote their linear span.

1.2 Four basic examples

This section is an informal description of four examples of finite diffusion processes. Their common feature is that their structure is rich in symmetries so that one can treat them by methods and techniques from finite harmonic analysis.

The main scope of this book is to study these examples in full detail and present all the mathematical background.

Example 1.2.1 (The simple random walk on the discrete circle) Let C_n denote a regular n-gon. We number the vertices consecutively from 0 to $n-1$: we regard these numbers as elements of $\mathbb{Z}/n\mathbb{Z}$, i.e. mod n. This way, there is an edge connecting the jth with the $(j+1)$st for all $j = 0, 1, \ldots, n-1$. This is the *discrete circle*.

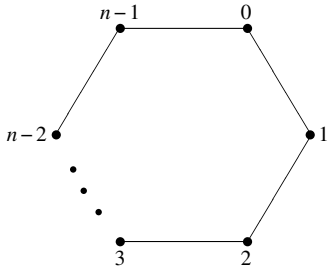

Figure 1.1. The discrete circle C_n

Suppose that a person is randomly moving on the vertices of C_n according to the following rule. The time is discrete ($t = 0, 1, 2, \ldots$) and at each instant of time there is a move.

At the beginning, that is at time $t = 0$, the random walker is in 0.

Suppose that at time t he is at position j. Then he tosses a fair coin and he moves from j to either $j+1$ or to $j-1$ if the result is a head or a tail, respectively. In other words, given that at time t he is in j, at time $t+1$ he can be at position $j+1$ or $j-1$ with the same probability $1/2$.

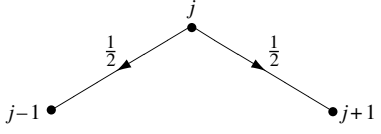

Figure 1.2. The equiprobable moves $j \mapsto j-1$ and $j \mapsto j+1$

Example 1.2.2 (The Ehrenfest diffusion model) The Ehrenfest diffusion model consists of two urns numbered 0, 1 and n balls, numbered $1, 2, \ldots, n$. A *configuration* is a placement of the balls into the urns. Therefore there are 2^n configurations and each of them can be identified with the subset A of $\{1, 2, \ldots, n\}$ corresponding to the set of balls contained in the urn 0. Note that there is no ordering inside the urns.

The *initial configuration* corresponds to the situation when all balls are inside urn 0.

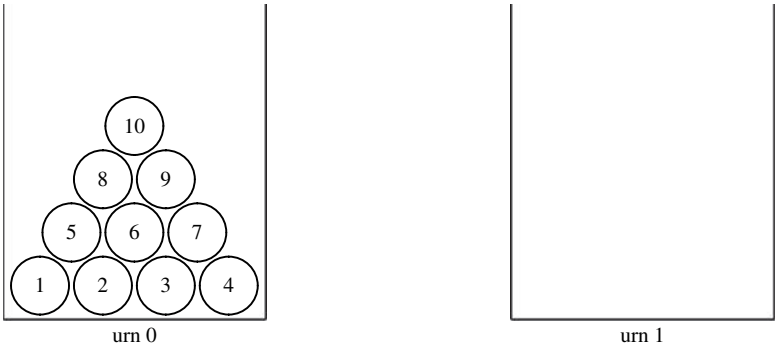

Figure 1.3. The initial configuration for the Ehrenfest urn model ($n = 10$)

Then, at each step, a ball is randomly chosen (each ball might be chosen with probability $1/n$) and it is moved to the other urn.

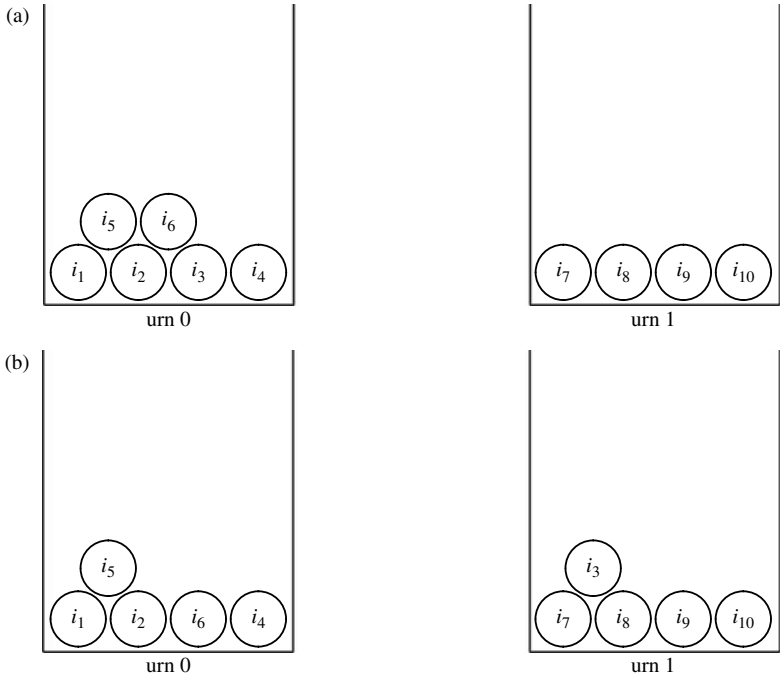

Figure 1.4. (**a**) A configuration at time t. (**b**) The configuration at time $t + 1$ if the chosen ball is i_3

Denoting by Q_n the set of all subsets of $\{1, 2, \dots, n\}$, we can describe this random process as follows. If we are in a configuration $A \in Q_n$ then, at the next step, we are in a new configuration $B \in Q_n$ of the form $A \coprod \{j\}$ for some $j \notin A$ or $A \setminus \{j'\}$ for some $j' \in A$ and each of these configurations is chosen with probability $1/n$ (actually, to avoid a parity problem, in our study of this model we shall make a slight change in the definition of the mixing process).

Example 1.2.3 (The Bernoulli–Laplace diffusion model) The Bernoulli–Laplace diffusion model consists of two urns numbered $0, 1$ and $2n$ balls, numbered $1, 2, \dots, 2n$. A *configuration* is a placement of the balls into the two urns, n balls each. Therefore there are $\binom{2n}{n}$ configurations, each of them can be identified with an n-subset A of $\{1, 2, \dots, 2n\}$ corresponding to the set of balls contained in the urn 0. The *initial configuration* corresponds to the situation when the balls contained in urn 0 are $1, 2, \dots, n$ (clearly the balls $n+1, n+2, \dots, 2n$ are contained in the urn 1).

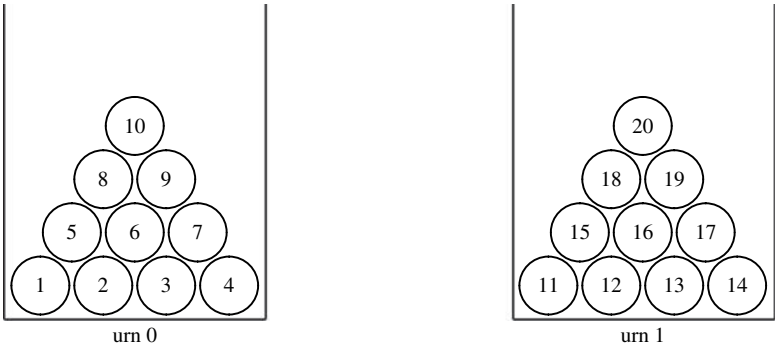

Figure 1.5. The initial configuration for the Bernoulli–Laplace urn model ($n = 10$)

Then at each step two balls are randomly chosen, one in urn 0, the other one in urn 1 and switched.

Denoting by Ω_n the set of all n-subsets of $\{1, 2, \dots, 2n\}$, we can describe this random process as follows. If we are in a configuration $A \in \Omega_n$ then at the next step we are in a new configuration $B \in \Omega_n$ of the form $A \coprod \{i\} \setminus \{j\}$ for some $i \notin A$ and $j \in A$ and each of these configurations is chosen with probability $1/n^2$.

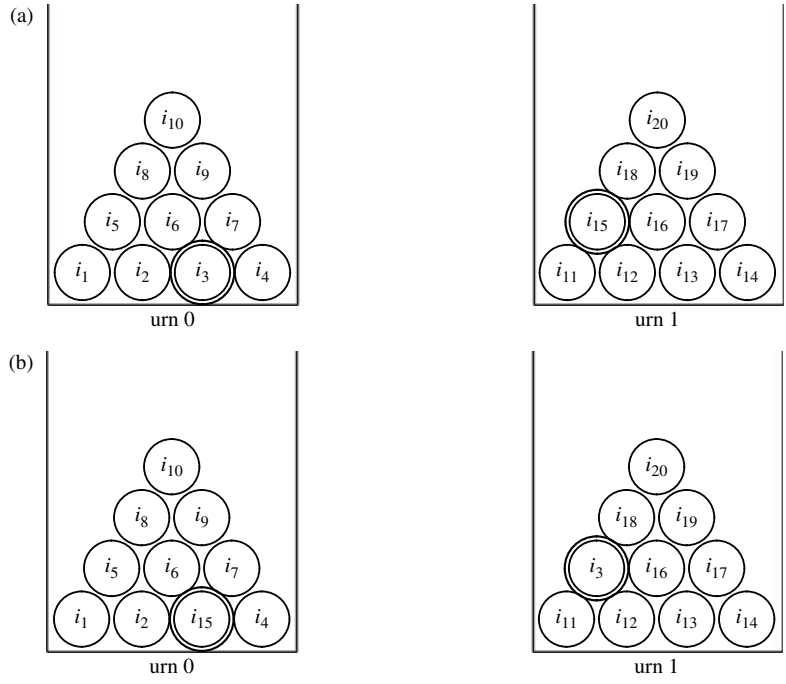

Figure 1.6. (**a**) A configuration at time t. (**b**) The configuration at time $t+1$ if the chosen balls are i_3 and i_{15}

Example 1.2.4 (Random transpositions) Consider a deck of n cards numbered from 1 to n. A *configuration* is a placement (permutation) of the cards in a row on a table. Therefore there are $n!$ configurations. The *initial configuration* corresponds to the placement $1, 2, \ldots, n$.

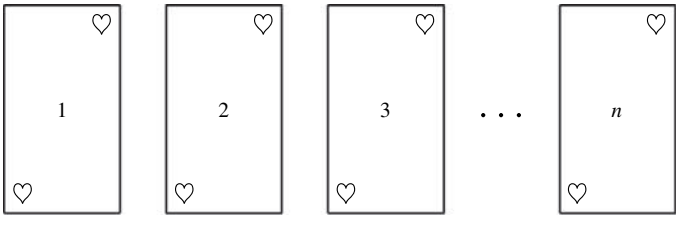

Figure 1.7. The initial configuration of the deck

Then at each step both the left and the right hand independently choose a random card. Note that the possibility is not excluded that the same card is chosen by both hands: such an event may occur with probability $1/n$.

If we have chosen two *distinct* cards we switch them, otherwise we leave the configuration unchanged.

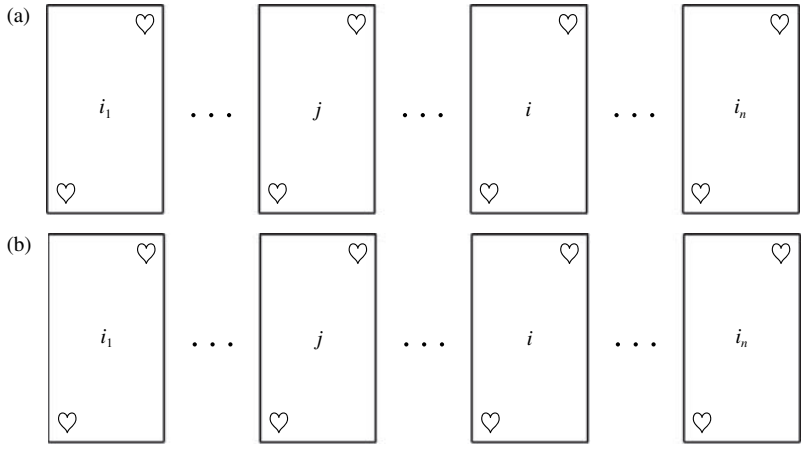

Figure 1.8. (**a**) A card configuration at time t. (**b**) The card configuration at time $t+1$ if the chosen cards are $i \neq j$

In other words, denoting by $T_n = \{\{i,j\} : i,j \in \{1,2,\dots,n\}, i \neq j\}$ the set of all unordered pairs of cards, with probability $1/n$ we leave the configuration unchanged, while, with probability $(n-1)/n$ we randomly pick one of the $n(n-1)/2$ elements in T_n. Altogether, at each step each element $t \in T_n$ can be chosen with probability $2/n^2$.

The following description is equivalent. Denote by S_n the *symmetric group* of degree n and by $T_n = \{g \in S_n : \exists i \neq j \text{ s.t. } g(k) = k, \forall k \neq i,j \text{ and } g(i) = j\}$ the set of all *transpositions*. Denote by $1_{S_n} \in S_n$ the identity element ($1_{S_n}(k) = k,\ \forall k \in \{1,2,\dots,n\}$). Each configuration, say $\mathbf{i} = (i_1, i_2, \dots, i_n)$ corresponds to the element $g_{\mathbf{i}} \in S_n$ defined by $g_{\mathbf{i}}(j) = i_j$ for all $j = 1,2,\dots,n$. We start at the identity 1_{S_n} and, at each step, we choose either a random element $t \in T_n$ with probability $2/n^2$ or the identity $t = 1_{S_n}$ with probability $1/n$. This way, if at the kth step we are in the configuration $\mathbf{i}$ then, at the $(k+1)$st step we are in the configuration $t \cdot \mathbf{i}$ corresponding to the group element $tg_{\mathbf{i}}$.

In each of the above examples, we have described a diffusion model with a deterministic initial configuration. For the discrete circle the random walker is at position (vertex) 0; in the Ehrenfest model all balls are in urn 0 (Figure 1.3); in the Bernoulli-Laplace model the balls $1,2,\dots,n$ are in urn 0 and the balls $n+1, n+2, \dots, 2n$ in urn 1 (Figure 1.5);

finally for the random transpositions the cards $1, 2, \ldots, n$ are placed in increasing order (Figure 1.7).

For each model we address the following natural question: *how long (how many steps) does it take to mix things up?* For instance, how many random transpositions (Example 1.2.4) are needed to generate a random permutation? It is indeed clear, even at an intuitive level, that repeated random transpositions do mix the cards, that is that after sufficiently many random transpositions, all configurations (i.e. permutations) become equiprobable.

In 1981 Diaconis and Shahshahani gave a striking answer to the above question: $\frac{1}{2}n \log n$ transposition are necessary and sufficient to generate a random permutation. Moreover they discovered that this model presents what is now called a *cutoff phenomenon*: the transition from *order* to *chaos* is concentrated in a small neighborhood of $\frac{1}{2}n \log n$.

Subsequently Diaconis and Shahshahani studied the models in Examples 1.2.1, 1.2.2 and 1.2.3. They showed that in the last two examples there is a cutoff phenomenon after $k = \frac{1}{4}n \log n$ steps, while for the discrete circle, the position of the random walker becomes random after $k = n^2$ steps, but no cutoff phenomenon occurs.

We shall first treat Examples 1.2.1 and 1.2.2 because they are significantly simpler (they only require elementary finite abelian harmonic analysis) and together they provide instances of absence and presence of the cutoff phenomenon. They will be discussed in the next chapter, while in the present one we continue by providing some probabilistic background.

Example 1.2.3 will be discussed in the second part of the book. It is a paradigmatic, but sufficiently simple, example of a problem on a homogeneous space on which there is a treatable set of spherical functions.

Example 1.2.4, which, historically, was the first to be analyzed, requires a deep knowledge of the representation theory of the symmetric group, a beautiful subject in mathematics on its own, which is a cornerstone both in pure and in applied mathematics (see, for instance, the book of Sternberg [212] for the applications to physics).

1.3 Markov chains

In this section we give the formal definition of a Markov chain in a simplified setting. We begin with some basic notions of probability theory.

A *probability measure* (or *distribution*) on a finite set X is a function $\nu : X \to [0, 1]$ such that $\sum_{x \in X} \nu(x) = 1$. It is called *strict* if $\nu(x) > 0$

for all $x \in X$. For a subset $A \subseteq X$ the quantity $\nu(A) = \sum_{x \in A} \nu(x)$ is the probability of A. Clearly $\nu(A \cup B) = \nu(A) + \nu(B) - \nu(A \cap B)$ for all $A, B \subseteq X$, $\nu(\emptyset) = 0$ and $\nu(X) = 1$.

The subsets of X are usually called the *events*. We also say that (X, ν) is a *finite measure space*.

Example 1.3.1 Let h, k be two positive integers and $X = \{r, b\}$ be a two-set. Define $\nu : X \to [0, 1]$ by setting $\nu(r) = \frac{h}{h+k}$ and $\nu(b) = \frac{k}{h+k}$. It is immediate to check that (X, ν) is a probability space. It can be used to modelize a single drawing from an urn containing h red balls and k blue balls: $\nu(r)$ (resp. $\nu(b)$) is the probability that the chosen ball is red (resp. blue).

Given two events $A, B \subseteq X$, with $\nu(A) > 0$, the *conditional probability* of B assuming A is

$$\nu(B|A) = \frac{\nu(A \cap B)}{\nu(A)}.$$

It expresses the probability of the event B given that the event A has occurred.

The following properties are obvious:

- $\nu(B|A) = 1$ if $A \subseteq B$;
- $\nu(\emptyset|A) = 0$;
- $\nu(B_1 \cup B_2|A) = \nu(B_1|A) + \nu(B_2|A) - \nu(B_1 \cap B_2|A)$;

for all $A, B, B_1, B_2 \subseteq X$ with $\nu(A) > 0$, and clearly $\nu(\cdot|A)$ is a probability measure on A.

Moreover, given elements $A_1, A_2, \dots, A_n \subseteq X$ with $\nu(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$, we have the so-called *Bayes sequential formula*

$$\begin{aligned} \nu(A_1 \cap A_2 \cap \cdots \cap A_n) = {} & \nu(A_1)\nu(A_2|A_1)\nu(A_3|A_1 \cap A_2) \\ & \cdots \nu(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}). \end{aligned} \tag{1.3}$$

Indeed:

$$\begin{aligned} \nu(A_1 \cap A_2 \cap \cdots \cap A_n) &= \nu(A_n|A_1 \cap \cdots \cap A_{n-1})\nu(A_1 \cap \cdots \cap A_{n-1}) \\ &= \nu(A_n|A_1 \cap \cdots \cap A_{n-1})\nu(A_{n-1}|A_1 \cap \cdots \cap A_{n-2}) \\ &\qquad \cdot \nu(A_1 \cap \cdots \cap A_{n-2}) \\ &= \cdots = \\ &= \nu(A_n|A_1 \cap \cdots \cap A_{n-1})\nu(A_{n-1}|A_1 \cap \cdots \cap A_{n-2}) \\ &\qquad \cdot \nu(A_{n-2}|A_1 \cap \cdots \cap A_{n-3}) \cdots \nu(A_2|A_1)\nu(A_1). \end{aligned}$$