PROBABILIDADE 2

S. Friedli

Contents

Beaucoup d'analystes, en effet, mettent au premier rang la notion du continu; c'est elle qui intervient d'une manière plus ou moins explicite dans leurs raisonnements. J'ai indiqué récemment [...] en quoi cette notion du continu, considéré comme ayant une puissance supérieure à celle du dénombrable, me paraît être une notion purement négative, la puissance des ensembles dénombrables étant la seule qui nous soit connue d'une manière positive, la seule qui intervienne effectivement dans nos raisonnements. Il est clair, en effet, que l'ensemble des éléments analytiques susceptibles d'être réellement définis et considérés ne peut être qu'un ensemble dénombrable; je crois que ce point de vue s'imposera chaque jour davantage aux mathématiciens et que le continu n'aura été qu'un instrument transitoire, dont l'utilité actuelle n'est pas négligeable [...], mais qui devra être regardé seulement comme un moyen d'étudier les ensembles dénombrables, lesquels constituent la seule réalité que nous puissions atteindre.

É. Borel, 1909 [**Bor09**].

CHAPTER 1

Conditional Expectation

1.1. Conventions

Throughout the text, we denote the fundamental probability space by $(\Omega, \mathcal{F}, P)$. Random variables, i.e. extended $\mathcal{F}$-measurable mappings from Ω to $\mathbf{R}$ (the extended real line), are usually denoted by X, Y, Z. We will often use the representation $X = X^+ - X^-$, where $X^+ := \sup\{X, 0\}$, $X^- := \sup\{-X, 0\}$. The set of random variables for which $\int |X| \, dP < \infty$, are called simply integrable. The expectation of an integrable random variable X is denoted

$$E[X] = \int X \, dP = \int X(\omega) P(d\omega) \, .$$

We denote $a \wedge b := \min\{a, b\}$, $a \vee b := \max\{a, b\}$. If the set $A \subset \mathbb{R}$ is empty, then $\inf A := +\infty$. We use the symbol $a_n \sim b_n$ to indicate that $\frac{a_n}{b_n} \to 1$ when $n \to \infty$, and We use the symbol $a_n \approx b_n$ to indicate that there exists two constants $c_1, c_2 > 0$ such that $c_1 \leq \frac{a_n}{b_n} \leq c_2$ when n is large enough. Our basic references are [**Bau88, Bil95, Bil65, Chu01, Shi84, Var00, Wil91, Str93, Str05, R.88, GS05, GS06, Rév05, Nev70, Pet00, Wal75**].

1.2. Conditionning with respect to an Event

In this section, we introduce the notion of conditional expectation with respect to a sub-σ-algebra, which is a fundamental notion in probability, especially for the definition of martingales. From now on, we call a collection $\mathcal{G}$ of subsets of Ω a sub-σ-algebra if it is a σ-algebra (in particular it must contain $\emptyset$ and Ω itself) and if $\mathcal{G} \subset \mathcal{F}$ (i.e. $A \in \mathcal{G}$ implies $A \in \mathcal{F}$). But before giving the general definition of conditional expectation we start by simpler considerations.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Any event $B \in \mathcal{F}$ with strictly positive probability allows to define a new probability measure $P(\cdot|B)$

on $(\Omega, \mathcal{F})$, conditionned on the event B:

$$P(A|B) := \frac{P(A \cap B)}{P(B)} . \tag{1.2.1}$$

In practice, in particular in statistics, one is interested in determining the probability of some event $A \in \mathcal{F}$, $P(A)$, but some practical restrictions are such that we only have access to some conditional probabilities $P(A|B_1)$, $P(A|B_2)$, ... where the B_n form a partition of the space Ω. In order to reconstruct the wanted probability $P(A)$, one then needs to know each $P(A|B_n)$, but also each $P(B_n)$. Once these informations are known, the formula of total probability gives (we assume for the time being that $P(B_n) > 0$ for each n)

$$P(A) = \sum_{n \geq 1} P(A|B_n) P(B_n)$$

If it reasonable to believe that $P(B_n)$ is known a priori, the true determination of $P(A)$ depends on the family $\{P(A|B_n)\}_{n \geq 1}$. Therefore, it is natural to encode these numbers into a simple function $f : \Omega \to [0, 1]$, constant on each B_n:

$$f(\omega) := \sum_{n \geq 1} P(A|B_n) 1_{B_n}(\omega) .$$

This random variable gives, in some sense, the best estimation of $P(A)$ when a point ω is chosen at random according to P: if $\omega \in B_n$, then $f(\omega) = P(A|B_n)$. If the experience is repeated a large number of times, the Law of Large Numbers

says that the empirical average of f will converge to

$$E[f] = P(A) .$$

This last equality is a particular case of the following. As a simple computation shows, we have, for all B which is a union of sets B_n,

$$P(A \cap B) = \int_B f \, dP .$$

The random variable f is called conditional probability of A with respect to the partition $\{B_n\}_{n \geq 1}$.

1.3. Conditionning with respect to a countable measurable partition

The same considerations hold for expectations. Assume $P(B) > 0$. If X is integrable, one can consider the expectation of X with respect to the measure $P(\cdot|B)$, called conditional expectation with respect to B,

$$E[X|B] := \int X(\omega)P(d\omega|B)\,. \tag{1.3.1}$$

As can be seen easily,

$$E[X|B] = \frac{1}{P(B)}\int_B X\,dP\,. \tag{1.3.2}$$

Observe that if (1.3.2) is taken as a definition for any integrable random variable X, then (1.2.1) can be obtained by choosing $X = 1_A$. We therefore consider (1.3.2) as a fundamental definition, which will also be seen to be more natural.

Let $\{B_n\}_{n\geq 1}$ be a (countable measurable) partition of Ω, that is a family of sets $B_n \in \mathcal{F}$ with $\bigcup_n B_n = \Omega$, and $B_n \cap B_m = \emptyset$ when $n \neq m$. Such a partition generates a sub-σ-field $\mathcal{B} \subset \mathcal{F}$, containing all unions of sets B_n. If $P(B_n) > 0$, we define $E[X|B_n]$ as in (1.3.1). If $P(B_n) = 0$, we define $E[X|B_n]$ in an arbitrary way, for example $E[X|B_n] := 0$. Then, for any integrable random variable X, define a new variable $E[X|\mathcal{B}] : \Omega \to \mathbb{R}$ by

$$E[X|\mathcal{B}](\omega) := \sum_{n\geq 1} E[X|B_n]1_{B_n}(\omega)\,, \tag{1.3.3}$$

called a version of the conditional expectation of X with respect to $\mathcal{B}$. The name "version" is used since we have made a specific choice on the sets B_n with zero probability.

Clearly $\omega \mapsto E[X|\mathcal{B}](\omega)$ is $\mathcal{B}$-measurable. Moreover we have, for all measurable set B,

$$\int_B E[X|\mathcal{B}]\,dP = \sum_{n\geq 1}\int_{B\cap B_n} E[X|B_n]\,dP = \sum_{n\geq 1} P(B\cap B_n)E[X|B_n]\,. \tag{1.3.4}$$

Now, if we assume that $B \in \mathcal{B}$ then B is a union of elements B_k, $k \in S$, which implies $P(B \cap B_n) = P(B_n)1_S(n)$, where $1_S(n) = 1$ if $n \in S$, 0 otherwise. Using the definition of $E[X|B_n]$, the last term in (1.3.4) equals

$$\sum_{n\geq 1} 1_S(n)\int_{B_n} X\,dP = \int_B X\,dP\,.$$

Therefore, the random variable $E[X|\mathcal{B}]$ satisfies

$$\int_B E[X|\mathcal{B}]\, dP = \int_B X\, dP\,, \quad \forall B \in \mathcal{B}\,. \tag{1.3.5}$$

Of course, any version of $E[X|\mathcal{B}]$ satisfies (1.3.5). The following shows that the versions of $E[X|\mathcal{B}]$ are the only random variables with this property.

LEMMA 1.3.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{B} \subset \mathcal{F}$ a sub-σ-algebra. Let Y_1, Y_2 be two $\mathcal{B}$-measurable integrable random variables. Then $Y_1 = Y_2$ a.e. if and only if*

$$\int_B Y_1\, dP = \int_B Y_2\, dP \quad \forall B \in \mathcal{B}\,. \tag{1.3.6}$$

PROOF. If $Y_1 = Y_2$ a.e. then (1.3.6) clearly holds [1]. Define $B := \{\omega : Y_1(\omega) > Y_2(\omega)\}$. We have $B \in \mathcal{B}$ and therefore (1.3.6) implies [2] $P(B) = 0$, i.e. $P(Y_1 \le Y_2) = 1$. Doing the same by interverting Y_1 and Y_2, one gets $P(Y_1 \ge Y_2) = 1$, which gives $P(Y_1 = Y_2) = 1$. □

1.4. Conditionning with Respect to a σ-algebra

In the previous section we have defined a version of the conditional expectation $E[X|\mathcal{B}]$ with respect to a sub-σ-algebra $\mathcal{B} \subset \mathcal{F}$ generated by a countable measurable partition, via the expression (1.3.3). We have then seen that two of its main properties were

(1) $E[X|\mathcal{B}]$ is $\mathcal{B}$-measurable.
(2) The family of relations (1.3.5) is satisfied.

We have seen in Lemma 1.3.1 that any other random variable satisfying these two properties is almost everywhere equal to $E[X|\mathcal{B}]$.

To extend the notion of conditional expectation to a general sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, the lack of countability does not allow a straight definition as in (1.3.3). In particular, a problem arises when considering measurable sets $B \in \mathcal{G}$ with zero probability, in which case (1.3.3) cannot be used to define $E[X|\mathcal{G}]$. Nevertheless, we shall use

[1] Namely, let N be a measurable set with $P(N) = 0$ so that $Y_1(\omega) = Y_2(\omega)$ for all $\omega \in N^c$. Let $B \in \mathcal{B}$. Then $\int_B (Y_1 - Y_2)dP = \int_{B\cap N}(Y_1 - Y_2)dP$. This last integral is zero since $1_N = 0$ a.e. and therefore for $i = 1, 2$, $\int_{B\cap N} |Y_i|\, dP \le \int |Y_i| 1_N\, dP = 0$.

[2] Here we make use of the following fact, which will be used often in this text: if $X \ge 0$ and $\int X\, dP = 0$, then $X = 0$ a.e.

the two conditions given above as a *definition* of $E[X|\mathcal{G}]$, and then verify existence, almost-everywhere uniqueness, and finally compatibility with the definition (1.3.3) in the case of sub-σ-algebras generated by countable partitions.

DEFINITION 1.4.1. *Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra and $X : \Omega \to \mathbb{R}$ a positive integrable random variable. Any random variable $Y : \Omega \to \mathbb{R}$ which is $\mathcal{G}$-measurable and satisfies*

$$\int_B Y\,dP = \int_B X\,dP\,, \quad \forall B \in \mathcal{G}\,, \tag{1.4.1}$$

is called a **version of the conditional expectation of X with respect to $\mathcal{G}$**. *We denote any version of the conditional expectation by $E[X|\mathcal{G}]$. For an arbitrary integrable random variable X, define conditional expectation as follows: $E[X|\mathcal{G}] := E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}]$.*

The main property of conditional expectation is therefore

$$\int_B E[X|\mathcal{G}]\,dP = \int_B X\,dP\,, \quad \forall B \in \mathcal{G}\,, \tag{1.4.2}$$

Observe that when it exists, $E[X|\mathcal{G}]$ is integrable. Namely, applying (1.4.2) with $B = \Omega$ gives

$$E[E[X|\mathcal{G}]] = \int E[X|\mathcal{G}]\,dP = \int X\,dP = E[X]\,. \tag{1.4.3}$$

We now verify that conditional expectation always exists. Define, for all $B \in \mathcal{G}$,

$$\nu(B) := \int_B X\,dP\,.$$

Since we assume $X \geq 0$ to be integrable, ν is a finite measure on $(\Omega, \mathcal{G})$. Moreover, it is absolutely continuous with respect to P (more precisely, to the restriction of P to $\mathcal{G}$). Therefore, the Radon-Nikodým Theorem guarantees the existence of a positive $\mathcal{G}$-measurable function Y such that

$$\nu(B) = \int_B Y\,dP\,, \quad \forall B \in \mathcal{G}\,.$$

Therefore, Y is nothing but the Radon-Nikodým density $\frac{d\nu}{dP}$, and is determined uniquely up to sets of measure zero (Lemma 1.3.1). We will give, in Section 1.6, another way of proving the existence of Y, not using the Radon-Nikodým Theorem.

EXAMPLE 1.4.1. To verify that the abstract definition coincides with the one of Section 1.3, let $\mathcal{B}$ be the σ-algebra generated by a measurable countable partition $\{B_n\}_{n\geq 1}$ and let $E[X|\mathcal{B}]$ denote any version of the conditional expectation. (Observe that $E[X|\mathcal{B}]$ is now defined *without* using (1.2.1).) Since $E[X|\mathcal{B}]$ is $\mathcal{B}$-measurable, it is constant on each B_n, i.e. $E[X|\mathcal{B}](\omega) = b_n$ for all $\omega \in B_n$. Now for all n we have

$$b_n P(B_n) = b_n \int_{B_n} dP = \int_{B_n} E[X|\mathcal{B}]\, dP = \int_{B_n} X\, dP\,.$$

Therefore, when $P(B_n) > 0$ one has

$$b_n = \frac{1}{P(B_n)} \int_{B_n} X\, dP\,,$$

which coincides with (1.3.3) on all sets B_n of positive measure. Since the other sets have measure zero and that we only need consider a countable number of them, we have therefore constructed a version of (1.3.3).

Before starting the study of general properties of $E[X|\mathcal{G}]$, we discuss the two extreme cases which give some insight into the dependence of $E[X|\mathcal{G}]$ on $\mathcal{G}$. First, let $\mathcal{G}$ be the largest possible sub-σ-algebra of $\mathcal{F}$, i.e. $\mathcal{G} = \mathcal{F}$, then clearly $E[X|\mathcal{G}] = X$ a.e. This follows from the fact that X is $\mathcal{F}$-measurable and from the trivial identity

$$\int_B X\, dP = \int_B X\, dP\,, \quad \forall B \in \mathcal{F}\,.$$

On the other extreme, if $\mathcal{G}$ is the smallest possible algebra, i.e. $\mathcal{G} = \{\emptyset, \Omega\}$, then $E[X|\mathcal{G}] = E[X]$ a.e. Namely, $E[X|\mathcal{G}]$ must be constant on Ω, and this constant is fixed by the only condition (take $B = \Omega$ in (1.4.2))

$$E[X|\mathcal{G}] = \int E[X|\mathcal{G}]\, dP = \int X\, dP = E[X]\,.$$

These two particular cases show that $E[X|\mathcal{G}]$ also gives an approximation of X; the finer $\mathcal{G}$, the better the approximation.

Basic Properties. From now on and until the end of the section, $\mathcal{G}$ will denote a sub-σ-algebra of $\mathcal{F}$. The following property can be easily verified

$$E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}] \quad a.e. \tag{1.4.4}$$

The first important property of conditional expectation is the following monotonicity result.

LEMMA 1.4.1. *If X, Y are two integrable random variables such that $X \le Y$ a.e., then $E[X|\mathcal{G}] \le E[Y|\mathcal{G}]$ a.e.*

PROOF. Assume $X \le Y$ a.e. Define $Z := E[Y|\mathcal{G}] - E[X|\mathcal{G}]$. For all $A \in \mathcal{G}$ we have

$$\int_A Z\,dP = \int_A (Y - X)\,dP \ge 0$$

This implies, for $A = \{Z < 0\}$, that $Z1_A = 0$ a.e., which implies $1_A = 0$ a.e. Let N be a set so that $P(N) = 0$ and $1_A(\omega) = 1$ for all $\omega \in N^c$. We have

$$P(A) = \int 1_A\,dP = \int 1_{A\cap N} dP \le P(N) = 0\,.$$

Therefore, $P(Z \ge 0) = 1$. □

Applying this lemma once with $Y = |X|$ and once with $Y = -|X|$ yields, using the linearity in (1.4.4),

COROLLARY 1.4.1. *If X is integrable then $|E[X|\mathcal{G}]| \le E[|X||\mathcal{G}]$ a.e.*

More fundamental are the following identities, which show what happens when the operation of conditionning is iterated, by conditionning successively with respect to two sub-σ-algebras.

LEMMA 1.4.2. *Let X be integrable and $\mathcal{G}, \mathcal{H}$ be two sub-σ-algebras, such that $\mathcal{G} \subset \mathcal{H}$. Then*

$$E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{G}] = E[E[X|\mathcal{H}]|\mathcal{G}] \quad a.e. \tag{1.4.5}$$

PROOF. Observe first that $E[E[X|\mathcal{H}]|\mathcal{G}]$ and $E[E[X|\mathcal{G}]|\mathcal{H}]$ are well-defined since $E[X|\mathcal{H}]$ and $E[X|\mathcal{G}]$ are integrable. For the first equality, observe that by definition of $E[\cdot|\mathcal{H}]$,

$$\int_H E[E[X|\mathcal{G}]|\mathcal{H}]\,dP = \int_H E[X|\mathcal{G}]\,dP\,, \quad \forall H \in \mathcal{H}\,.$$

In particular, since $\mathcal{G} \subset \mathcal{H}$,

$$\int_G E[E[X|\mathcal{G}]|\mathcal{H}]\,dP = \int_G E[X|\mathcal{G}]\,dP\,, \quad \forall G \in \mathcal{G}\,.$$

Now by the definition of $E[X|\mathcal{G}]$,

$$\int_G E[X|\mathcal{G}]\,dP = \int_G X\,dP\,, \quad \forall G \in \mathcal{G}\,.$$

This shows the first equality in (1.4.5). Similarly, using $\mathcal{G} \subset \mathcal{H}$ gives, for all $G \in \mathcal{G}$,

$$\int_G E[E[X|\mathcal{H}]|\mathcal{G}]\,dP = \int_G E[X|\mathcal{H}]\,dP = \int_G X\,dP\,,$$

which shows the second equality in (1.4.5). □

The identities (1.4.5) are known as the Tower Property of conditional expectation.

1.5. Basic Convergence Theorems

In this paragraph we give the conditional versions of the classical convergence theorems of Integration Theory.

THEOREM 1.5.1 (Monotone Convergence Theorem, Conditional Version). *Let X_n be an increasing sequence of integrable random variables such that $X_n \nearrow X$ a.e., where X is integrable. Then*

$$E[X_n|\mathcal{G}] \nearrow E[X|\mathcal{G}] \quad a.e.$$

PROOF. Since $X_n \le X_{n+1}$ a.e. we have $E[X_n|\mathcal{G}] \le E[X_{n+1}|\mathcal{G}]$ a.e. by Lemma 1.4.1. Therefore, the limit $Z := \lim_{n\to\infty} E[X_n|\mathcal{G}]$ exists a.e. (!) Using twice the Monotone Convergence Theorem,

$$\int_A Z\,dP = \lim_{n\to\infty}\int_A E[X_n|\mathcal{G}]\,dP = \lim_{n\to\infty}\int_A X_n\,dP = \int_A X\,dP\,, \quad \forall A \in \mathcal{G}\,.$$

This shows that $Z = E[X|\mathcal{G}]$ a.e. □

COROLLARY 1.5.1. *If Y_n is a sequence of positive integrable random variables, then*

$$E\Big(\sum_{n\ge1} Y_n|\mathcal{G}\Big) = \sum_{n\ge1} E[Y_n|\mathcal{G}] \quad a.e. \tag{1.5.1}$$

THEOREM 1.5.2 (Dominated Convergence Theorem, Conditional Version). *Let X_n be a sequence of integrable random variables such that $X_n \to X$ a.e., and such that $|X_n| \le Y$ where Y is integrable. Then X is integrable and*

$$E[X_n|\mathcal{G}] \to E[X|\mathcal{G}] \quad a.e.$$

PROOF. Clearly, X is integrable since $|X| \le Y$ a.e. Using Corollary 1.4.1,

$$|E[X_n|\mathcal{G}] - E[X|\mathcal{G}]| = |E[X_n - X|\mathcal{G}]| \le E[|X_n - X||\mathcal{G}] \le E[Z_n|\mathcal{G}]\,,$$

where $Z_n := \sup_{m\geq n} |X_m - X|$. But $Z_n \searrow 0$ a.e. and since $|Z_n| \leq 2Y$, Z_n is integrable and we have

$$\lim_{n\to\infty} \int Z_n \, dP = 0 \tag{1.5.2}$$

by the Dominated Convergence Theorem. Since $0 \leq Z_{n+1} \leq Z_n$, we have $E[Z_{n+1}|\mathcal{G}] \leq E[Z_n|\mathcal{G}]$ a.e. (Lemma 1.4.1). Therefore $Z := \lim_{n\to\infty} E[Z_n|\mathcal{G}]$ exists a.e. (!), and we have, for all n,

$$0 \leq \int Z \, dP \leq \int E[Z_n|\mathcal{G}] \, dP = \int Z_n \, dP \,, \tag{1.5.3}$$

which by (1.5.2) converges to 0 when $n \to \infty$. This implies $Z = 0$ a.e. and proves the theorem. □

Corollary 1.5.2. *If X is integrable, Y is $\mathcal{G}$-measurable and XY is integrable, then*

$$E[XY|\mathcal{G}] = YE[X|\mathcal{G}] \quad a.e. \tag{1.5.4}$$

Proof. We first verify (1.5.4) for $Y = 1_A$, $A \in \mathcal{G}$. Then for all $B \in \mathcal{G}$,

$$\begin{aligned}\int_B YE[X|\mathcal{G}] \, dP = \int_{B\cap A} E[X|\mathcal{G}] \, dP &= \int_{B\cap A} X \, dP \\ &= \int_B XY \, dP = \int_B E[XY|\mathcal{G}] \, dP \,.\end{aligned}$$

By linearity, this extends to any finite linear combination of indicator functions, i.e. to any simple function. Therefore, assume first that Y is positive and let Y_n be a sequence of simple functions $Y_n \nearrow Y$. Since $|XY_n| \leq |XY|$, which is integrable, and since $XY_n \to XY$, Theorem 1.5.2 gives

$$E[XY|\mathcal{G}] = \lim_{n\to\infty} E[XY_n|\mathcal{G}] = \lim_{n\to\infty} Y_n E[X|\mathcal{G}] = YE[X|\mathcal{G}] \quad a.e.$$

In the last inequality we used the fact that $E[X|\mathcal{G}] < \infty$ a.e., which follows from the fact that it is integrable. For the general case simply use the decomposition $Y = Y^+ - Y^-$. □

Theorem 1.5.3 (Jensen's Inequality, Conditional Version). *Let X be integrable and $\phi = \phi(x)$ be convex. Assume $\phi(x)$ is finite for all $x \in \mathbb{R}$ and that $\phi(X)$ is integrable. Then*

$$\phi(E[X|\mathcal{G}]) \leq E[\phi(X)|\mathcal{G}] \quad a.e. \tag{1.5.5}$$

PROOF. Since ϕ is convex and finite, it is continuous, and there exists, for all $x_0 \in \mathbb{R}$, a finite constant A_0 such that

$$\phi(x) \geq \phi(x_0) + A_0(x - x_0), \quad \forall x \in \mathbb{R}.$$

We will first assume that $E[X|\mathcal{G}](\omega) < \infty$ for all $\omega \in \Omega$. Applying the previous inequality with $x = X$, $x_0 = E[X|\mathcal{G}]$, and taking $E[\cdot|\mathcal{G}]$ on both sides gives

$$E[\phi(E[X|\mathcal{G}])|\mathcal{G}] \leq E[\phi(X)|\mathcal{G}].$$

Since $E[X|\mathcal{G}]$ is $\mathcal{G}$-measurable, $\phi(E[X|\mathcal{G}])$ is too, and by Corollary 1.5.2,

$$E[\phi(E[X|\mathcal{G}])|\mathcal{G}] = \phi(E[X|\mathcal{G}])E[1|\mathcal{G}] = \phi(E[X|\mathcal{G}]) \quad a.e.,$$

which proves (1.5.5). When $E[X|\mathcal{G}]$ is not bounded, define the sets $A_n := \{\omega : |E[X|\mathcal{G}](\omega)| \leq n\}$. By the preceding result we have

$$\phi(E[1_{A_n}X|\mathcal{G}]) \leq E[\phi(1_{A_n}X)|\mathcal{G}] \quad a.e. \tag{1.5.6}$$

By Theorem 1.5.2, $E[1_{A_n}X|\mathcal{G}] \to E[X|\mathcal{G}]$ a.e. Therefore, since ϕ is continuous, $\phi(E[1_{A_n}X|\mathcal{G})] \to \phi(E[X|\mathcal{G}])$ a.e. For the right-hand side of (1.5.6), write

$$\begin{aligned} E[\phi(1_{A_n}X)|\mathcal{G}] &= E[1_{A_n}\phi(X)|\mathcal{G}] + E[1_{A_n^c}\phi(0)|\mathcal{G}] && a.e. \\ &= 1_{A_n}E[\phi(X)|\mathcal{G}] + 1_{A_n^c}\phi(0) && a.e. \end{aligned}$$

Since $E[\phi(X)|\mathcal{G}]$ is integrable, it is finite a.e. The same reasoning applies to $E[X|\mathcal{G}]$, which leads to $1_{A_n} \to 1$ a.e. Therefore the first term converges to $E[\phi(X)|\mathcal{G}]$ a.e. The second converges to 0 a.e., which finishes the proof. □

Observe that Corollary 1.4.1 can be obtained by Jensen's Inequality with $\phi(x) = |x|$.

1.6. A Geometric Interpretation of $E[X|\mathcal{G}]$

There exists an enlighting geometric interpretation of conditional expectation. For convenience, let us abreviate $L_p(\Omega, \mathcal{F}, P)$ by $L_p(\mathcal{F})$. As we know, the conditional expectation transforms a random variable $X \in L_1(\mathcal{F})$ into a random variable $\mathcal{L}_\mathcal{G}X := E[X|\mathcal{G}] \in L_1(\mathcal{G})$. We have seen in (1.4.4) that $\mathcal{L}_\mathcal{G}$ is linear and since $L_1(\mathcal{G}) \subset L_1(\mathcal{F})$, we want to interpret $\mathcal{L}_\mathcal{G}$ as a projection.

This can be done by resctricting ourselves to $L_2(\mathcal{F})$. Observe that $\mathcal{L}_{\mathcal{G}}$ maps $L_2(\mathcal{F})$ into $L_2(\mathcal{G})$. Namely, let $X \in L_2(\mathcal{F})$. Then [3] $X \in L_1(\mathcal{F})$ and therefore $E[X|\mathcal{G}]$ is well defined. Using Jensen's Inequality with $\phi(x) = x^2$, we get

$$\int |E[X|\mathcal{G}]|^2 \, dP \leq \int E[|X|^2|\mathcal{G}] \, dP = \int |X|^2 \, dP < \infty \,,$$

and therefore $E[X|\mathcal{G}] \in L_2(\mathcal{G})$.

Now, we use the fact that $L_2(\mathcal{F})$ is a Hilbert space, complete with respect to the norm $\|\cdot\|_2 = \langle \cdot, \cdot \rangle$ inherited from the scalar product

$$\langle X, Y \rangle := \int XY \, dP \,.$$

Since $L_2(\mathcal{G})$ is obviously a closed (with respect to the topology induced by $\|\cdot\|_2$) subspace of $L_2(\mathcal{F})$, one can consider the orthogonal projection $\pi_{\mathcal{G}} : L_2(\mathcal{F}) \to L_2(\mathcal{G})$. $\pi_{\mathcal{G}}$ is self-adjoint and satisfies $\pi_{\mathcal{G}}^2 = \pi_{\mathcal{G}}$. Now $\pi_{\mathcal{G}} X$ is $\mathcal{G}$-measurable, by definition, and for all $A \in \mathcal{G}$,

$$\int_A \pi_{\mathcal{G}} X \, dP = \int (\pi_{\mathcal{G}} X) 1_A \, dP = \int X (\pi_{\mathcal{G}} 1_A) \, dP = \int X 1_A \, dP = \int_A X \, dP \,,$$

where we have used self-adjointness and $\pi_{\mathcal{G}} 1_A = 1_A$, since $1_A \in L_2(\mathcal{G})$. Therefore, $\pi_{\mathcal{G}} X$ is the equivalence class of almost everywhere equal random variables which contains $E[X|\mathcal{G}]$. That is, $\pi_{\mathcal{G}} X$ represents all the possible versions of $E[X|\mathcal{G}]$.

To extend this construction to any integrable random variable, let first $X \in L_1(\mathcal{F})$ be positive, and let $X_n := \inf\{X, n\}$. Then obviously $X_n \in L_2(\mathcal{F})$ and $Y_n := \pi_{\mathcal{G}} X_n = E[X_n|\mathcal{G}]$ can be constructed as above. We have $Y_n \leq Y_{n+1}$ since $X_n \leq X_{n+1}$, and therefore $Y_n \nearrow Y := \lim_n Y_n$ exists. Using two times the dominated convergence theorem then shows that $Y = E[X|\mathcal{G}]$. Namely, for any $B \in \mathcal{G}$,

$$\int_B Y \, dP = \lim_{n\to\infty} \int_B Y_n \, dP = \lim_{n\to\infty} \int_B E[X_n|\mathcal{G}] \, dP = \lim_{n\to\infty} \int_B X_n \, dP = \int_B X \, dP \,.$$

[3] Namely, if $X \in L_2(\mathcal{F})$ then

$$\int |X| \, dP = \int_{|X|\leq 1} |X| \, dP + \int_{|X|>1} |X| \, dP \leq P(|X| \leq 1) + \int |X|^2 \, dP < \infty \,.$$

This method gives a construction of $E[X|\mathcal{G}]$ which doesn't rely on the Radon-Nikodým Theorem [4]. Using this point of view, various results obtained before for conditional expectation can be obtained by using the properties of the orthogonal projection $\pi_\mathcal{G}$. For example, if $\mathcal{G} \subset \mathcal{H}$, then $L_1(\mathcal{G}) \subset L_1(\mathcal{H})$ and therefore the identities $\pi_\mathcal{G}\pi_\mathcal{H} = \pi_\mathcal{G} = \pi_\mathcal{H}\pi_\mathcal{G}$, which are obvious from the geometric point of view, directly imply Lemma 1.4.2.

1.7. Conditional Probability

When applied to the special case where the random variable X is an indicator function, conditional expectation leads to the definition of conditional probability.

Definition 1.7.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $A \in \mathcal{F}$. The conditional probability of A with respect to a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ is defined by*

$$P(A|\mathcal{G}) := E[1_A|\mathcal{G}] . \tag{1.7.1}$$

Again, $P(A|\mathcal{G})$ is only defined uniquely up to sets of measure zero. Being any version of $E[1_A|\mathcal{G}]$, it satisfies

$$P(A \cap B) = \int_B P(A|\mathcal{G})\, dP , \quad \forall B \in \mathcal{G} . \tag{1.7.2}$$

A simple example is when $\mathcal{G}$ is the σ-algebra generated by a partition $\{B, B^c\}$, $B \in \mathcal{F}$. In this case, $P(A|\mathcal{G})$ is constant on B and B^c (since it is $\mathcal{G}$-measurable) and if $0 < P(B) < 1$,

$$P(A|\mathcal{G})(\omega) = \begin{cases} P(A|B) & \text{if } \omega \in B , \\ P(A|B^c) & \text{if } \omega \in B^c , \end{cases} \tag{1.7.3}$$

where $P(A|B)$ and $P(A|B^c)$ are defined as in (1.2.1). Observe, nevertheless, that our definition makes sense even when $P(B)$ or $P(B^c)$ equals zero.

For each *fixed* A, the properties of the function $\omega \mapsto P(A|\mathcal{G})(\omega)$ are known almost-everywhere. In particular, we have $P(\emptyset|\mathcal{G}) = 1$ and $P(\Omega|\mathcal{G}) = 1$ almost everywhere. Moreover, if $(A_n)_{n\geq 1}$ is any sequence

[4] In fact, as we will see later, the Radon-Nikodým Theorem can be obtained as a consequence of the convergence results on martingales that will be described later.

of pairwise disjoint events, then by Corollary 1.5.1

$$P\Big(\bigcup_{n\geq 1} A_n|\mathcal{G}\Big) = \sum_{n\geq 1} P(A_n|\mathcal{G})\,, \quad a.e. \tag{1.7.4}$$

Nevertheless, this does *not* imply that $P(\cdot|\mathcal{G})$ is a probability measure on $(\Omega, \mathcal{F})$. Indeed, the properties mentionned above hold one sets of measure one, but in each case this set depends on the event observed. For example, for a given sequence $(A_n)_{n\geq 1}$, the set of ω's for which (1.7.4) holds depends on the whole sequence $(A_n)_{n\geq 1}$. This leads to the following question: does there exist a version of $P(\cdot|\mathcal{G})$ for which $A \mapsto P(A|\mathcal{G})(\omega)$ is a probability measure for each ω belonging to a set of probability one? The answer to this question will be affirmative when more structure is given to the set Ω. We discuss this in the following section.

1.7.1. Regular Conditional Probabilities. In the preceding section we associated to each event $A \in \mathcal{F}$ the random variable $\omega \mapsto P(A|\mathcal{G})(\omega)$. Due to the previous discussion, we now wish to consider $P(A|\mathcal{G})(\omega)$, for each fixed $\omega \in \Omega$, as a function of A. This leads to the following natural definition.

Definition 1.7.2. *Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. A map $(\omega, A) \mapsto P^*(A|\mathcal{G})(\omega)$ is called a* regular conditional probability with respect to $\mathcal{G}$ *if the following holds:*

1. *For all $\omega \in \Omega$, $A \mapsto P^*(A|\mathcal{G})(\omega)$ is a probability measure on $(\Omega, \mathcal{F})$.*
2. *For all $A \in \mathcal{F}$, $\omega \mapsto P^*(A|\mathcal{G})(\omega)$ is a version of $P(A|\mathcal{G})$.*

The element $P^*(\cdot|\mathcal{G})$ is sometimes called expectation kernel (see Bauer [**Bau88**]). This terminology is made clear in the following proposition, which shows that the conditional expectation of any integrable variable can be derived from $P^*(\cdot|\mathcal{G})$.

Proposition 1.7.1. *Let $P^*(\cdot|\mathcal{G})$ be an expectation kernel. Then for all integrable random variable X,*

$$E[X|\mathcal{G}] = \int X(\omega) P^*(d\omega|\mathcal{G})\,, \quad a.e. \tag{1.7.5}$$

The role of $P^*(\cdot|\mathcal{G})$, in the construction of $E[\cdot|\mathcal{G}]$, should therefore be seen as the analog of the role played by $P(\cdot)$ in the construction of

$E[\cdot]$, via the common expression

$$E[X] = \int X(\omega) P(d\omega)\,.$$

Proof of Proposition 1.7.1: Let $X = 1_A$. Then (1.7.5) is immediate. If $X \geq 0$ then by taking any increasing sequence of simple random variables $X_n \nearrow X$, the Monotone Convergence Theorem (conditional and standard) gives

$$E[X|\mathcal{G}] = \lim_{n\to\infty} E[X_n|\mathcal{G}] = \lim_{n\to\infty} \int X_n(\omega) P^*(d\omega|\mathcal{G}) = \int X(\omega) P^*(d\omega|\mathcal{G})\,, \quad a.e.$$

For a general integrable random variable, use $X = X^+ - X^-$. □

The existence of a regular conditional probability is not guaranteed in general. The following result, which we give without proof, gives the existence of a regular conditional probability under some assumption on the measurable space $(\Omega, \mathcal{F})$. See [**Bau88**] for details. A measurable space $(\Omega, \mathcal{F})$ is standard Borel if Ω is a complete separable metric space and $\mathcal{F}$ is its Borel σ-algebra.

Theorem 1.7.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, where $(\Omega, \mathcal{F})$ is Borel standard. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Then there exists a regular conditional probability with respect to $\mathcal{G}$.*

CHAPTER 2

The Simple Random Walk

In this section we study a few basic properties of the simple random walk on $\mathbb{Z}$, such as recurrence, the Reflection Principle, a Large Deviation estimate, and the Law of the Iterated Logarithm. The simple random walk will be encountered in many places in subsequent sections. For example, we will come back to random walks in Section 4.3.2, after having developped the general theory of recurrence, in the framework of Markov chains. The simple random walk will be the first example of martingale, treated in Section 5.

Consider a sequence $(Y_n)_{n\geq 1}$ of independent identically distributed random variables with $P(Y_k = +1) = p$, $P(Y_k = -1) = q$ with $p + q = 1$. These can be constructed using Theorem 3.1.1. Define $S_0 := 0$, and for all $n \geq 1$, $S_n := \sum_{k=1}^n Y_k$. The sequence $(S_n)_{n\geq 0}$ is called the simple random walk on $\mathbb{Z}$. When $p = \frac{1}{2}$, the random walk is called symmetric.

The purpose of this chapter is to describe the long-time behaviour of the walk, i.e. describe the typical behaviour of S_n when n becomes large. Observe first that, by the Strong Law of Large Numbers,

$$\frac{S_n}{n} \to E[Y_1] \equiv 2p - 1 \quad \text{a.s.} \tag{2.0.6}$$

This implies, in particular, that $P(\lim_{n\to\infty} S_n = +\infty) = 1$ when $p > \frac{1}{2}$, and that $P(\lim_{n\to\infty} S_n = -\infty) = 1$ when $p < \frac{1}{2}$: when $p \neq \frac{1}{2}$, this means that almost surely, the walk visits the origin a finite number of times and then travels towards $\pm\infty$ with an asymptotic speed $v = 2p - 1 \neq 0$. A first natural set of questions is thus: when $p = \frac{1}{2}$, does the walk come back to the origin an infinite number of times? Does it visit any $k \in \mathbb{Z}$ an infinite number of times? What can we say about the random variable giving the first time at which the walk visits the origin? Does the walk, on the whole, spend an equal amount of time on each side of the origin?

The basic combinatorial identity is the following: if $n \geq 1$ and $k \in \mathbb{Z}$ are admissible, in the sense that $\{S_n = k\} \neq \emptyset$, then

$$P(S_n = k) = \binom{n}{\frac{k+n}{2}} p^{\frac{n+k}{2}} q^{\frac{n-k}{2}}. \tag{2.0.7}$$

This allows a few comments. First, let us consider the events $\{S_{2n} = 0\}$. By (2.0.7) and the Stirling Formula (see Exercise 2.1),

$$P(S_{2n} = 0) = \binom{2n}{n} (pq)^n \sim \frac{1}{\sqrt{\pi n}} (4pq)^n. \tag{2.0.8}$$

Therefore, when $p \neq q$, i.e $p \neq \frac{1}{2}$, we have $4pq < 1$ and so $\sum_n P(S_{2n} = 0) < \infty$. By Borel-Cantelli, this implies that $P(S_{2n} = 0 \text{ i.o.}) = 0$, i.e. the walk visits the origin a finite number of times almost surely, which we already knew. On the other hand, when $p = \frac{1}{2}$ then $\sum_n P(S_{2n} = 0) = \infty$ but nothing can be said about the visits at the origin since the events $\{S_{2n} = 0\}$ are not independent.

The asymptotic formula given in (2.0.7) is just a particular case of the following limit theorem, which will be used at a few places in this section. The proof, based entirely on (2.0.7) and the Stirling Formula, is left as an exercise (it can also be found in [**Shi84**], p. 56).

THEOREM 2.0.1 (Local Limit Theorem). *Assume* $p = \frac{1}{2}$. *If* $l_n = o(n^{\frac{2}{3}})$, *then* [1]

$$P(S_n = l_n) \sim \sqrt{\frac{2}{\pi n}} e^{-\frac{l_n^2}{2n}} \tag{2.0.9}$$

To prove the Central Limit Theorem, one needs only $l_n = o(n^{\frac{1}{2}})$. When proving the Law of the Iterated Logarithm, we will see that an exponent larger than $\frac{1}{2}$ is necessary.

2.1. Recurrence

Recurrence poses the problem of knowing if and how does the walk come back to its starting point. Therefore, define the time of first return to the origin,

$$T_0 := \inf\{n \geq 1 : S_n = 0\}, \tag{2.1.1}$$

where we remind that we make the convention that the infimum over an emptyset is $+\infty$. T_0 is an $\mathbb{N} \cup \{\infty\}$-valued random variable. The

[1] Remember that $a_n \sim b_n$ means that $\frac{a_n}{b_n} \to 1$ when $n \to \infty$.

random walk is called recurrent if $P(T_0 < \infty) = 1$, and transient if $P(T_0 < \infty) < 1$. The basic recurrence result is the following.

THEOREM 2.1.1. *For the simple random walk on* $\mathbb{Z}$,

$$P(T_0 < \infty) = 1 - |p - q| . \tag{2.1.2}$$

In particular, the walk is recurrent if and only if $p = q = \frac{1}{2}$. *Moreover,*

$$E[T_0] \begin{cases} < \infty & \text{if } p \neq \frac{1}{2} , \\ = \infty & \text{if } p = \frac{1}{2} . \end{cases} \tag{2.1.3}$$

PROOF. Let f denote the generating function for the distribution of T_0:

$$f(s) := E[s^{T_0}] = \sum_{n \geq 1} P(T_0 = 2n) s^{2n} \quad -1 < s < 1 . \tag{2.1.4}$$

We are interested in computing

$$P(T_0 < \infty) = P\Big(\bigcup_{n \geq 1} \{T_0 = 2n\} \Big) = \sum_{n \geq 1} P(T_0 = 2n) = \lim_{s \to 1^-} f(s) .$$

We used Abel's Theorem (see Exercise 2.2) for the interchange of the limit and the sum. To compute $f(s)$, it is useful to introduce also the generating function for the distribution $P(S_{2n} = 0)$:

$$g(s) := \sum_{n \geq 0} P(S_{2n} = 0) s^{2n} .$$

Since we have an explicit expression for $P(S_{2n} = 0)$ in (2.0.7), g is easy to compute:

$$g(s) = \sum_{n \geq 0} \binom{2n}{n} (pqs^2)^n \equiv \frac{1}{\sqrt{1 - 4pqs^2}} .$$

This last identity is a simple Taylor expansion (Exercise 2.5). Now, observe that by the Markov Property (see Exercise 2.3)

$$\begin{aligned} P(S_{2n} = 0) &= \sum_{k=1}^{n} P(S_{2n} = 0 | T_0 = 2k) P(T_0 = 2k) \\ &= \sum_{k=1}^{n} P(S_{2n-2k} = 0) P(T_0 = 2k) , \end{aligned}$$

which is the nth coefficient of the power series representing $f(s)g(s)$. Therefore, multiplying by s^n and summing over $n \geq 0$, we get $g(s) = 1 + g(s)f(s)$. Therefore, $f(s) = 1 - \sqrt{1 - 4pqs^2}$, and so

$$P(T_0 < \infty) = \lim_{s \to 1^-} f(s) = 1 - \sqrt{1 - 4pq}\,. \tag{2.1.5}$$

Since $1 = (p+q)^2$, we have shown (2.1.2). For the second part, observe that, again by Abel's Theorem,

$$E[T_0] = \sum_{n \geq 1} 2nP(T_0 = 2n) = \lim_{s \to 1^-} f'(s)\,.$$

(2.1.3) follows easily by explicit computation of $\lim_{s \to 1^-} f'(s)$. □

2.2. The Reflection Principle and The Arcsine Law

We know that when $p = q$, the walk comes back to the origin almost surely. In the present section we study the time spent by the walk on either side of the origin.

It will be convenient to consider the trajectory of the random walk as a two dimensional spacetime line on $\mathbb{N} \times \mathbb{Z}$, in which the point (n, x) corresponds to the event $\{S_n = x\}$. We start by a key combinatorial result.

Lemma 2.2.1 (Reflection Principle). *Let $x, y > 0$, $n \geq 1$. The number of paths joining $(0, x)$ to (n, y) which visit at least once the origin is equal to the number of paths joining $(0, -x)$ to (n, y).*

Proof. Let $\mathcal{C}(0, -x; n, y)$ denote the set of paths joining $(0, -x)$ to (n, y), and $\mathcal{C}(0, x; n, y)^*$ denote the set of paths joining $(0, x)$ to (n, y) which hit the origin at least once. We construct a bijection φ between these two sets. Assuming $(n, y-x)$ is admissible, let $\pi \in \mathcal{C}(0, x; n, y)^*$. Define n' as the first time π hits the origin. Define a path $\varphi(\pi)$ by

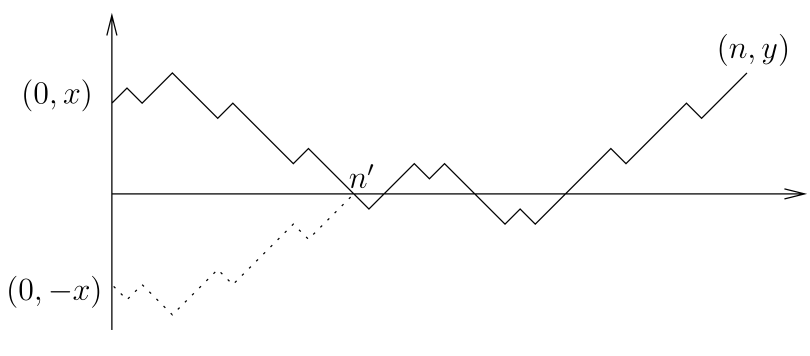


Figure 1. The Reflection Principle.

reflecting π through the origin on the interval $[0, n']$, and leaving it unchanged on the interval $[n', n]$. Then $\varphi(\pi) \in \mathcal{C}(0, -x; n, y)$ and φ is clearly invertible, proving the claim. □

Let $N_{n,x}$ be the number of paths which start at the origin and reach the point x by time n: $N_{n,x} := |\{S_n = x\}|$. In order to reach x by time n, a path must contain $n_+ := \frac{n+x}{2}$ upward-steps and $n_- := \frac{n-x}{2}$ downward-steps. Therefore,

$$N_{n,x} = \binom{n}{n_+} . \tag{2.2.1}$$

THEOREM 2.2.1. *Let $y > 0$. Conditionnally on the event that the walk reached y by time n, the probability that it never visited the origin at times $1, 2, \dots, n-1$ equals $\frac{y}{n}$.*

PROOF. We assume (n, y) is admissible. If the walk starts at the origin and never visits it again, its position at time 1 must be 1. The total number of paths which start at $(1, 1)$ and reach (n, y) is $N_{n-1,y-1}$, and the number of those which visit at least once the origin equals, by the Reflection Principle, $N_{n-1,y+1}$. The number of paths which reach y by time n and don't visit the origin is therefore, with $\alpha := \frac{n+y}{2}$,

$$N_{n-1,y-1} - N_{n-1,y+1} = \binom{n-1}{\alpha-1} - \binom{n-1}{\alpha} = \frac{2\alpha - n}{n} N_{n,y} \equiv \frac{y}{n} N_{n,y} , \tag{2.2.2}$$

which gives the result. □

This last theorem is usually called the **Ballot Theorem**. Assume that two individuals A and B were the unique candidates to an election. One starts looking at the votes, one after the other. At the end, one gets $\alpha > \beta$, where α and β are the respective total numbers of votes of each candidate. That is, A won the election. What is the probability that A stayed ahead of B all through the counting of the votes? If one considers the counting of the votes as a symmetric random walk starting at the origin, in which an upward-step is a vote for A and a downward-step is a vote for B, then the total number of steps is $\alpha + \beta$ and the position to be reached is $\alpha - \beta > 0$. The probability that A stays ahead of B through the counting of the votes equals the probability that this random walk never visits the origin up to time $\alpha + \beta$, which equals $\frac{\alpha-\beta}{\alpha+\beta}$ by Theorem 2.2.1.

One can use (2.2.2) in another way. Since the number of paths starting at the origin, ending at $(2n, 2y)$ and not going through the origin for times $1, 2, \dots, 2n$, equals $N_{2n-1,2y-1} - N_{2n-1,2y+1}$, dividing by the total number of paths up to time $2n$ gives

$$P(S_1 > 0, \dots, S_{2n} > 0, S_{2n} = 2y) = \frac{1}{2}P(S_{2n-1} = 2y-1) - \frac{1}{2}P(S_{2n-1} = 2y+1)\,.$$

The $\frac{1}{2}$s appear because of the choice on the last step. By summing over $y = 1, \dots, n$, this gives

$$P(S_1 > 0, \dots, S_{2n} > 0) = \frac{1}{2}P(S_{2n-1} = 1)\,.$$

In the same way, $P(S_1 < 0, \dots, S_{2n} < 0) = \frac{1}{2}P(S_{2n-1} = -1)$. We have shown

Theorem 2.2.2. *For the simple random walk starting at the origin,*

$$P(S_1 \neq 0, \dots, S_{2n} \neq 0) = P(S_{2n} = 0)\,. \tag{2.2.3}$$

We can use this identity in many ways. For example, consider the first return to the origin, $T_0 := \inf\{k \geq 1 : S_k = 0\}$; we can express its distribution as follows

$$P(T_0 > 2n) = P(S_1 \neq 0, \dots, S_{2n} \neq 0) = P(S_{2n} = 0) \sim \frac{1}{\sqrt{\pi n}}\,.$$

In particular, $P(T_0 = \infty) = \lim_{n\to\infty} P(T_0 > 2n) = 0$, and $E[T_0] = \infty$, which we already knew from Theorem 2.1.1.

As a second application of (2.2.3), we will now show that on a given time interval, the walk spends most of its time on only one side of the origin. Define the time of the last visit to the origin before time $2n$:

$$L_{2n} := \sup\{k \leq 2n : S_k = 0\}\,. \tag{2.2.4}$$

Theorem 2.2.3 (Arsine Law). *Let $0 < a < b < 1$. Then, as $n \to \infty$,*

$$P\Big(a \leq \frac{L_{2n}}{2n} < b\Big) \longrightarrow \frac{1}{\pi}\int_a^b \frac{1}{\sqrt{x(1-x)}}dx\,. \tag{2.2.5}$$

The name of this theorem stems from the fact that by a change of variable $\sqrt{x} \equiv y$,

$$\frac{1}{\pi}\int_a^b \frac{1}{\sqrt{x(1-x)}}dx = \frac{2}{\pi}\int_{\sqrt{a}}^{\sqrt{b}} \frac{1}{\sqrt{1-y^2}}dy = \frac{2}{\pi}\big(\arcsin\sqrt{b} - \arcsin\sqrt{a}\big)\,.$$

The counter-intuitiveness of the Arsine Law is that the asymptotic distribution of $\frac{L_{2n}}{2n}$ is symmetric around $\frac{1}{2}$. In particular,

$$P\Big(\frac{L_{2n}}{2n} < \frac{1}{2}\Big) \longrightarrow \frac{1}{2}\,.$$

Quoting Durrett: “in gambling terms, if two people were to bet 1$ on a coin flip every day of the year, then with probability $\frac{1}{2}$ one of the players would be ahead from July 1st to the end of the year, an event that would undoubtedly cause the other player to complain about his bad luck”.

The Arcsine Law follows from the following lemma. Let $u_{2k} := P(S_{2k} = 0)$.

Lemma 2.2.2. $P(L_{2n} = 2k) = u_{2k}u_{2n-2k}$.

Proof. We have, by the Markov Property,

$$\begin{aligned} P(L_{2n} = 2k) &= P(S_{2k} = 0, S_{2k+1} \neq 0, \dots, S_{2n} \neq 0) \\ &= P(S_{2k} = 0)P(S_{2k+1} \neq 0, \dots, S_{2n} \neq 0 | S_{2k} = 0) \\ &= P(S_{2k} = 0)P(S_1 \neq 0, \dots, S_{2n-2k} \neq 0)\,, \end{aligned}$$

which, by (2.2.3), proves the lemma. □

Sketch of the proof of Theorem 2.2.3: Consider the sequence a_n (resp. b_n) defined such that $2na_n$ (resp. $2nb_n$) is the smallest (resp. largest) even integer larger (resp. smaller) than $2na$ (resp. $2nb$). Then,

$$\begin{aligned} P\Big(a \le \frac{L_{2n}}{2n} < b\Big) = P\Big(a_n \le \frac{L_{2n}}{2n} < b_n\Big) &= \sum_{k: a_n \le \frac{2k}{2n} \le b_n} P(L_{2n=2k}) \\ &= \sum_{k: a_n \le \frac{2k}{2n} \le b_n} u_{2k}u_{2n-2k}\,. \end{aligned}$$

By the asymptotic behaviour of u_{2k} for large k, we have

$$\begin{aligned} \sum_{k: a_n \le \frac{2k}{2n} \le b_n} u_{2k}u_{2n-2k} &\sim \sum_{k: a_n \le \frac{2k}{2n} \le b_n} \frac{1}{\sqrt{\pi k}} \frac{1}{\sqrt{\pi(2n-2k)}} \\ &= \frac{1}{n} \sum_{k: a_n \le \frac{2k}{2n} \le b_n} \frac{1}{\sqrt{\pi \frac{k}{n}(1-\frac{k}{n})}} \to \int_a^b \frac{1}{\sqrt{\pi x(1-x)}} dx\,. \end{aligned}$$

□

2.3. Large Deviations: the Bernstein Estimate

What is the probability that the walk makes a large excursion, i.e. how can one estimate the probability of events such as $\{S_n \geq cn\}$? These happen to be exponentially small.

THEOREM 2.3.1. *For all $c > 0$,*

$$P(S_n \geq cn) \leq \exp\big(-\frac{c^2}{16}n\big) \quad \forall n \geq 1 \tag{2.3.1}$$

It is important to notice that the Bernstein Estimate (2.3.1) can also be used for a c which depends on n, for example $c = n^{-\frac{1}{3}}$.

PROOF. It is simpler to use the variables $L_n := \frac{S_n+n}{2} \in \{0, 1, \dots, n\}$:

$$P(S_n \geq cn) = P\Big(L_n \geq \frac{1+c}{2}n\Big) \equiv P\big(L_n \geq (p+\delta)n\big),$$

with $p = \frac{1}{2}$, $\delta = \frac{c}{2}$. Now

$$P\big(L_n \geq (p+\delta)n\big) = \sum_{(p+\delta)n \leq k \leq n} P(L_n = k) = \sum_{(p+\delta)n \leq k \leq n} \binom{n}{k} p^k q^{n-k}.$$

Introduce a parameter $\lambda > 0$ that will be chosen below, and observe that for each k of the sum, one has $1 \leq e^{\lambda(k-(p+\delta)n)}$. By rearranging $k - pn = qk - p(n-k)$, we get

$$\begin{aligned} P\big(L_n \geq (p+\delta)n\big) &\leq e^{-\lambda\delta n} \sum_{(p+\delta)n \leq k \leq n} \binom{n}{k} [pe^{\lambda q}]^k [qe^{-\lambda p}]^{n-k} \\ &\leq e^{-\lambda\delta n}\big(pe^{\lambda q} + qe^{-\lambda p}\big)^n. \end{aligned}$$

Since $e^x \leq x + e^{x^2}$, we get

$$P\big(L_n \geq (p+\delta)n\big) \leq e^{-\lambda\delta n}\big(pe^{\lambda^2 q^2} + qe^{\lambda^2 p^2}\big)^n \leq e^{-\lambda\delta n} e^{\lambda^2 n} \equiv e^{-\frac{\delta^2}{4}n},$$

once we choose $\lambda := \frac{\delta}{2}$. This proves the theorem. □

2.4. The Law of the Iterated Logarithm

In this section we consider only $p = \frac{1}{2}$. We have seen in Theorem 2.3.1 that when $c > 0$, $\{S_n \geq cn\}$ are very rare events. Nevertheless, the simple symmetric random walk *does* make excursions far from the origin, and we make this precise in the present section, following Révész [**Rév05**].

Loosely speaking, the aim is to find an increasing sequence of intervals centered at the origin $I_0 \subset I_1 \subset I_2 \subset \dots$ such that, almost surely,: 1) S_n belongs to I_n, unless maybe for a finite number of times; in this sense, the sequence of intervals I_n contains the whole trajectory of the walk, 2) the sequence I_n is the smallest possible, in the sense that S_n must hit the boundary of I_n infinitely many times.

The Law of the Iterated Logarithm below shows that the sequence I_n must grow like

$$|I_n| \approx \sqrt{n \log\log n}\,. \tag{2.4.1}$$

Let us argue in favor of the appearance of something of the form "log log". The details are left as an exercise (Exercise 2.7). First, from the Strong Law of Large Numbers (2.0.6),

$$\lim_{n\to\infty} \frac{S_n}{n} = 0 \quad \text{a.s.} \tag{2.4.2}$$

almost surely, and so clearly we need $|I_n| \ll n$. On the other hand, normal excursions, of order $n^{\frac{1}{2}}$, are described by the Central Limit Theorem: they are probable. Therefore, we need $|I_n| \gg n^{\frac{1}{2}}$. Then, as can be seen easily, $E[S_n^{2k}] = O(n^k)$ for all $k \geq 1$. By Chebycheff's Inequality, $P(|S_n| \geq n^{\frac{1}{2}+\epsilon}) = O(n^{-2\epsilon k})$ for all $\epsilon > 0$, which implies

$$\lim_{n\to\infty} \frac{S_n}{n^{\frac{1}{2}+\epsilon}} = 0 \quad \text{a.s.} \tag{2.4.3}$$

This shows that $|I_n| \ll n^{\frac{1}{2}+\epsilon}$ for any $\epsilon > 0$. The next candidates are therefore amplitudes of order $n^{\frac{1}{2}}(\log n)^\delta$, $\delta > 0$. Nevertheless, it can be shown (see [**R.88**] p.65) that for all $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{S_n}{n^{\frac{1}{2}}(\log n)^{\frac{1}{2}+\epsilon}} = 0 \quad \text{a.s.} \tag{2.4.4}$$

The sequence I_n must therefore satisfy $n^{\frac{1}{2}} \ll |I_n| \ll n^{\frac{1}{2}}(\log n)^{\frac{1}{2}+\epsilon}$ for all $\epsilon > 0$. The good rate happens to be the one given in (2.4.1) with a constant equal to $\sqrt{2}$, as shown in the following result, due to Khinchin (1923).

Theorem 2.4.1 (Law of the Iterated Logarithm). *Let $(S_n)_{n\geq 0}$ denote the simple symmetric random walk on $\mathbb{Z}$. Then, almost surely,*

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1 \quad \textit{and} \quad \liminf_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = -1\,. \tag{2.4.5}$$

The Law of the Iterated Logarithm implies, in particular, that $\limsup_n S_n = +\infty$, $\liminf_n S_n = -\infty$ almost surely, which implies that $P(S_n = x \text{ i.o.}) = 1$ for all $x \in \mathbb{Z}$. That is, the symmetric random walk visits any site an infinite number of times. In particular, it is recurrent.

Although Theorem 2.4.1 holds under more general assumptions on the increments (X_n), we will prove it in our simple setting, following Révész [**Rév05**]. The central ingredient is the following combinatorial lemma.

LEMMA 2.4.1. *Let $(S_n)_{n\geq 0}$ denote the simple symmetric random walk on $\mathbb{Z}$. Let $M_n := \max_{0\leq j\leq n} |S_j|$. There exist two constants $c_1, c_2 > 0$ such that for any sequence $0 < k_n = o(n^{\frac{1}{6}})$, the following holds for large n:*

$$\frac{c_1}{k_n} e^{-\frac{k_n^2}{2}} \leq P(S_n \geq n^{\frac{1}{2}} k_n) \leq \frac{c_2}{k_n} e^{-\frac{k_n^2}{2}}, \tag{2.4.6}$$

$$\frac{c_1}{k_n} e^{-\frac{k_n^2}{2}} \leq P(M_n \geq n^{\frac{1}{2}} k_n) \leq \frac{4c_2}{k_n} e^{-\frac{k_n^2}{2}}. \tag{2.4.7}$$

The events $\{S_n \geq n^{\frac{1}{2}} k_n\}$ describe fluctuations which are slightly larger than normal, which is exactly what we need since in our case $k_n \sim \sqrt{\log\log n} = o(n^{\frac{1}{6}})$.

PROOF. The inequalities (2.4.6) rely on the Local Limit Theorem 2.0.1. Take $L > 0$ large enough, and consider the decomposition

$$P(S_n \geq n^{\frac{1}{2}} k_n) = P(n^{\frac{1}{2}} k_n \leq S_n < L n^{\frac{1}{2}} k_n) + P(S_n \geq L n^{\frac{1}{2}} k_n)$$

For the first term, we can use the Local Limit Theorem to obtain upper and lower bounds: with $k_n = o(n^{\frac{1}{6}})$ one has, for large enough n,

$$\begin{aligned} P(n^{\frac{1}{2}} k_n \leq S_n < L n^{\frac{1}{2}} k_n) &\sim \sqrt{\frac{2}{\pi n}} \sum_{k=n^{\frac{1}{2}} k_n}^{L n^{\frac{1}{2}} k_n} e^{-\frac{k^2}{2n}} \\ &\approx \frac{1}{\sqrt{n}} \int_{n^{\frac{1}{2}} k_n}^{L n^{\frac{1}{2}} k_n} e^{-\frac{x^2}{2n}} dx \\ &= \int_{k_n}^{L k_n} e^{-\frac{y^2}{2}} dy \approx \int_{k_n}^{\infty} e^{-\frac{y^2}{2}} dy \approx \frac{1}{k_n} e^{-\frac{k_n^2}{2}}. \end{aligned} \tag{2.4.8}$$

For the second term, we use the Bernstein Estimate (2.3.1) with $c = c_n = Ln^{-\frac{1}{2}}k_n$:

$$P(S_n \geq Ln^{\frac{1}{2}}k_n) = P(S_n \geq c_n n) \leq \exp\Big(-\frac{c_n^2}{16}n\Big) = \exp\Big(-\frac{L^2}{8}\frac{k_n^2}{2}\Big)$$

which, when L is large enough, becomes negligible compared to (2.4.8). This shows (2.4.6). Then, (2.4.7) follows from (2.4.6) and $P(M_n \geq a) \leq 4P(S_n \geq a)$. To see this, we first show that if $M_n^+ = \max_{0\leq j\leq n} S_j$, then

$$P(M_n^+ \geq a) = 2P(S_n > a) + P(S_n = a)\,. \tag{2.4.9}$$

For $a = 0$ the identity is trivial. For $a > 0$, write

$$\begin{aligned} P(M_n^+ \geq a) - P(S_n = a) &= P(M_n^+ \geq a, S_n < a) + P(M_n^+ \geq a, S_n > a) \\ &= P(M_n^+ \geq a, S_n < a) + P(S_n > a)\,. \end{aligned} \tag{2.4.10}$$

It therefore suffices to show that

$$P(M_n^+ \geq a, S_n < a) = P(M_n^+ \geq a, S_n > a)\,.$$

But this follows from a simple reflection argument analogous to what was done in the Reflection Principle: on $\{M_n^+ \geq a, S_n < a\}$, consider the first time n' at which the walk crosses the line $\mathfrak{L} = \{(x, y) : y = a\}$. On $[n', n]$ reflect the path across $\mathfrak{L}$, which transforms the constraint $S_n < a$ into $S_n > a$. □

Proof of Theorem 2.4.1: Define $\Lambda_n := \sqrt{2n \log\log n}$. We first show that for all $\epsilon > 0$,

$$\limsup_{n\to\infty} \frac{M_n}{\Lambda_n} \leq 1 + \epsilon \quad \text{a.s.} \tag{2.4.11}$$

Since $S_n \leq |S_n| \leq M_n$, this implies $\limsup_{n\to\infty} \frac{S_n}{\Lambda_n} \leq 1 + \epsilon$. By (2.4.7),

$$P(M_n \geq (1+\epsilon)\Lambda_n) \leq \frac{1}{(\log n)^{1+\epsilon}}$$

for large enough n. If we consider a subsequence $n_k = \lfloor \theta^k \rfloor$, where $\theta > 1$, then $(\log n)^{1+\epsilon} \sim k^{1+\epsilon}$ and by Borel-Cantelli we get $P(M_{n_k} \geq (1+\epsilon)\Lambda_{n_k} \text{ i.o.}) = 0$, which means

$$\limsup_{k\to\infty} \frac{M_{n_k}}{\Lambda_{n_k}} \leq 1 + \epsilon \quad \text{a.s.} \tag{2.4.12}$$

For any large n, consider the number $k \geq 1$ for which $n_k \leq n < n_{k+1}$. We have (here it is crucial that M_n increases with n)

$$\frac{M_n}{\Lambda_n} \leq \frac{M_{n_{k+1}}}{\Lambda_{n_k}} = \frac{\Lambda_{n_{k+1}}}{\Lambda_{n_k}} \frac{M_{n_{k+1}}}{\Lambda_{n_{k+1}}}$$

Since

$$\lim_{k\to\infty} \frac{\Lambda_{n_{k+1}}}{\Lambda_{n_k}} = \sqrt{\theta}\,, \tag{2.4.13}$$

and since θ can be taken arbitrarily close to 1, we have shown (2.4.11). For the lower bound, we use the same subsequence n_k and show that θ can be taken large enough so that almost surely, for infinitely many ks,

$$\frac{S_{n_k}}{\Lambda_{n_k}} \geq 1 - 2\epsilon\,. \tag{2.4.14}$$

This implies $\limsup_{n\to\infty} \frac{S_n}{\Lambda_n} \geq 1-\epsilon$. In order to use Borel-Cantelli for the lower bound, we consider the independent events $\{S_{n_{k+1}} - S_{n_k} \geq (1-\epsilon)\Lambda_{n_{k+1}}\}$. Since $S_{n_{k+1}} - S_{n_k}$ has the same distribution as $S_{n_{k+1}-n_k}$, we have by (2.4.6), for large enough k and θ,

$$\begin{aligned} P(S_{n_{k+1}} - S_{n_k} \geq (1-\epsilon)\Lambda_{n_{k+1}}) &= P\Big[\frac{S_{n_{k+1}-n_k}}{\sqrt{n_{k+1}-n_k}} \geq (1-\epsilon)\frac{\Lambda_{n_{k+1}}}{\sqrt{n_{k+1}-n_k}}\Big] \\ &\geq \frac{C}{(\log n_{k+1})^{1-\frac{\epsilon}{4}}} \approx \frac{1}{k^{1-\frac{\epsilon}{4}}} \end{aligned}$$

where $C = C(\theta) > 0$. This implies that $\{S_{n_{k+1}} - S_{n_k} \geq (1-\epsilon)\Lambda_{n_{k+1}} \text{ i.o.}\}$ has probability one. On this set, we write (this holds for infinitely many ks)

$$\begin{aligned} \frac{S_{n_{k+1}}}{\Lambda_{n_{k+1}}} = \frac{S_{n_{k+1}} - S_{n_k}}{\Lambda_{n_{k+1}}} + \frac{\Lambda_{n_k}}{\Lambda_{n_{k+1}}}\frac{S_{n_k}}{\Lambda_{n_k}} &\geq (1-\epsilon) + \frac{\Lambda_{n_k}}{\Lambda_{n_{k+1}}}\frac{S_{n_k}}{\Lambda_{n_k}} \\ &\geq (1-\epsilon) - \frac{\Lambda_{n_k}}{\Lambda_{n_{k+1}}}\frac{M_{n_k}}{\Lambda_{n_k}} \end{aligned}$$

By (2.4.13) and (2.4.12),

$$\limsup_{k\to\infty} \frac{\Lambda_{n_k}}{\Lambda_{n_{k+1}}}\frac{M_{n_k}}{\Lambda_{n_k}} \leq \frac{1+\epsilon}{\sqrt{\theta}} \leq \epsilon \qquad \text{a.s.}$$

for large enough θ. This proves (2.4.14). □

The almost sure information provided by the Law of the Iterated Logarithm for the excursions of the walk are as follows: for all $\epsilon > 0$, we have

$$S_n \leq (1+\epsilon)\sqrt{2n \log\log n}$$

for all large enough n, and

$$S_n \geq (1-\epsilon)\sqrt{2n \log\log n}$$

for an infinite number of ns. This shows that the largest typical excursions are of order $\sqrt{2n\log\log n}$. Various refinements of this asymptotic behaviour can be found in [**Rév05**]. For example, for all $\epsilon > 0$,

$$S_n \leq \sqrt{n(2\log\log n + (3+\epsilon)\log\log\log n)}$$

for all large enough n, and

$$S_n \geq \sqrt{n(2\log\log n + 3\log\log\log n)}$$

for an infinite number of ns.

2.5. Exercises

Exercise 2.1. ([**Str05**] p.18) Let $Z_1, \ldots, Z_n$ be independent, identically distributed with $Z_1 \sim \exp(1)$. Show that for all $0 < R \leq \sqrt{n}$,

$$1-\frac{1}{R^2} \leq P\Big[\Big|\frac{Z_1+\cdots+Z_n-n}{\sqrt{n}}\Big| \leq R\Big] = \frac{1}{(n-1)!}\int_{-\sqrt{n}R+n}^{+\sqrt{n}R+n} t^{n-1}e^{-t}dt \leq 1$$

Make a change of variables to obtain

$$\int_{-\sqrt{n}R+n}^{+\sqrt{n}R+n} t^{n-1}e^{-t}dt = n^{n-\frac{1}{2}}e^{-n}\int_{-R}^{R}\exp\big[-\frac{\sigma^2}{2}+E_n(\sigma)\big]d\sigma\,,$$

where

$$E_n(\sigma) = (n-1)\log\big(1+\frac{\sigma}{\sqrt{n}}\big) - \sqrt{n}\sigma + \frac{\sigma^2}{2}\,.$$

Use a Taylor expansion for $\log(1+x)$ to show that $E_n(\sigma) \to 0$ uniformly for $|\sigma| < R$. Combine to obtain

$$1-\frac{1}{R^2} \leq \liminf_{n\to\infty}\frac{n^{n+\frac{1}{2}}e^{-n}}{n!}\int_{-R}^{+R} e^{-\frac{\sigma^2}{2}}d\sigma \leq \limsup_{n\to\infty}\frac{n^{n+\frac{1}{2}}e^{-n}}{n!}\int_{-R}^{+R} e^{-\frac{\sigma^2}{2}}d\sigma \leq 1\,.$$

This proves Stirling's Formula: $n! \sim \sqrt{2\pi n}n^n e^{-n}$.

Exercise 2.2. Show Abel's Theorem: if $a_n \geq 0$ and $\sum_n a_n s^n$ is convergent for all $0 < s < 1$, then $\lim_{s\to 1^-}\sum_n a_n s^n = \sum_n a_n$ (whether both are finite or infinite).

Exercise 2.3. Show the Markov Property of the simple random walk: for all $x_1, \ldots, x_{n+1} \in \mathbb{Z}$,

$$P(S_{n+1} = x_{n+1}|S_n = x_n, \ldots, S_1 = x_1) = P(S_{n+1} = x_{n+1}|S_n = x_n)\,.$$

Use this to prove that $P(S_{2n}=0|T_0=2k)=P(S_{2n-2k}=0)$, where T_0 is the time of first return to the origin.

EXERCISE 2.4. Show the Martingale Property of the simple symmetric random walk $(S_n)_{n\geq 0}$: 1) $E[|S_n|]<\infty$ for all n, 2) if $\mathcal{F}_n:=\sigma(X_1,\dots,X_n)$ denotes the σ-algebra generated by $X_1,\dots,X_n$, then

$$E[S_{n+1}|\mathcal{F}_n]=S_n \quad \text{a.s.}$$

EXERCISE 2.5. Show that

$$\sum_{n\geq 0}\binom{2n}{n}(pqs^2)^n=\frac{1}{\sqrt{1-4pqs^2}}\,.$$

EXERCISE 2.6. [**GS05**] p. 27. Consider the random walk defined by $S_0:=0$, $S_n:=X_1+\dots+X_n$, where the X_k are integer valued, i.i.d. random variables. The range of the walk up to time n is the number of distinct integers $k\in\mathbb{Z}$ for which there exists $m\leq n$ such that $S_m=k$. Show that

$$P(R_n=R_{n-1}+1)=P(S_1\neq 0,\dots,S_n\neq 0)\,,$$

and conclude that when $n\to\infty$,

$$\frac{E[R_n]}{n}\to P(S_k\neq 0\,\forall k\geq 1)\,.$$

What is the value of this limit in the case of the simple symmetric random walk?

EXERCISE 2.7. (Towards the Law of the Iterated Logarithm, [**Rév05**] p29-30.) Consider the simple random walk on $\mathbb{Z}$, denoted $(S_n)_{n\geq 1}$.

(1) Verify that $E[|S_n|^2]=n$, and estimate $P(|S_{n^2}|\geq \epsilon n^2)$ for all $\epsilon>0$. Conclude that $\frac{|S_{n^2}|}{n^2}\to 0$ a.s. From this, deduce:

$$\frac{S_n}{n}\to 0 \;\text{ a.s.} \qquad \text{(Borel, 1909)}\,.$$

(2) Show that $E[S_n^{2k}]=O(n^k)$ when $n\to\infty$. Conclude that $P(|S_n|\geq n^{\frac{1}{2}+\epsilon})=O(n^{-2\epsilon k})$ for all $\epsilon>0$, which implies

$$\frac{S_n}{n^{\frac{1}{2}+\epsilon}}\to 0 \;\text{ a.s.} \qquad \text{(Hausdorff, 1913)}$$

(3) Show that for all t, $E[e^{tS_n}]=(\cosh t)^n$. Compute $\lim_n E[e^{\frac{S_n}{\sqrt{n}}}]$. Use this to bound $P[S_n\geq (1+\epsilon)n^{\frac{1}{2}}\log n]$ for large n. Conclude

that

$$\limsup_{n\to\infty} \frac{S_n}{n^{\frac{1}{2}} \log n} \leq 1 \text{ a.s.}$$

EXERCISE 2.8. [**GS05**] p. 332. Let $(X_n)_{n\geq 1}$ be independent, $S_n := X_1 + \cdots + X_n$. A function $\varphi = \varphi(x) \geq 0$ is said to belong to the the upper class if $P(S_n > \varphi(n) \text{ i.o.}) = 0$. Consider the case where $X_k \sim \mathcal{N}(0,1)$, and study functions of the form $\varphi(x) = \sqrt{\lambda x \log x}$, $\lambda > 0$. For which values of λ does φ belong to the upper class?
Hint: use

$$\int_x^\infty e^{-\frac{t^2}{2}} dt \leq \frac{1}{x} e^{-\frac{x^2}{2}} .$$

EXERCISE 2.9. Consider $\Lambda_N := \{0, 1, 2, \dots, N\}$, and let $\partial\Lambda_N := \{0, N\}$, $\operatorname{int}\Lambda_N := \Lambda_N \backslash \partial\Lambda_N = \{1, 2, \dots, N-1\}$.

(1) A function $f : \Lambda_N \to \mathbb{R}$ is called harmonic if

$$\frac{f(x-1) + f(x+1)}{2} = f(x) \quad \text{for all } x \in \operatorname{int}\Lambda_N .$$

  (a) Show the Maximum Principle: a harmonic function attains its maximum and minumum on $\partial\Lambda_N$.
  (b) Consider the Dirichlet Problem: find a harmonic function f such that $f(0) = a$, $f(N) = b$. Show that the solution to the Dirichlet Problem is unique. Hint: consider two solutions f, g and study $h := f - g$.

(2) For each $x \in \operatorname{int}\Lambda_N$, consider the simple symetric random walk starting at x. Let $p_N(x)$ denote the probability that the walk, starting at x, reaches 0 before N. Clearly, $p_N(0) = 1$, $p_N(N) = 0$. Show that $p_N : \Lambda_N \to [0,1]$ is harmonic. Make an ansatz for the solution of the Dirichlet Problem. Assuming N is even, compute $p_N(\frac{N}{2})$ and $N \to \infty$.

(3) ([**GS05**] p. 74) Generalize to the non-symmetric case: take $p \in (0,1)$, $p \neq \frac{1}{2}$, and let $q := 1 - p$, and define a function $f : \Lambda_N \to \mathbb{R}$ to be harmonic if

$$qf(x-1) + pf(x+1) = f(x) \quad \text{for all } x \in \operatorname{int}\Lambda_N .$$

Prove the same statements as in (1). Show that in (2), the solution is of the form $p_N(x) = a\theta_1^x + b\theta_2^x$, where $\theta_1 = 1$, $\theta_2 = \frac{q}{p}$. Use the boundary condition to find the constants a and b. Compute $\lim_{N\to\infty} p_N(x)$ for a fixed $x \geq 1$ and then $\lim_{N\to\infty} p_N(\frac{N}{2})$. Can you obtain recurrence?

(4) A gambler wins 1$ with probability p and loses 1$ with probability $1-p$. The game stops only when the gambler goes bankrupt. If he starts with an initial amount of k\$, what is the probability of him going bankrupt?

EXERCISE 2.10. [**Str05**] Let $B_1, B_2, \ldots$ be i.i.d. $\mathbb{Z}$-valued random variables with $0 < E[|B_1|] < \infty$. Let $[x]^+ := \max\{0, x\}$. Define the queue $(Q_n)_{n\geq 0}$ by $Q_0 := 0$,

$$Q_n := [Q_{n-1} + B_n]^+ .$$

(1) Show that (voir Landim p. 110)

$$Q_n = S_n - \min_{0\leq m\leq n} S_m = \max_{0\leq m\leq n} (S_n - S_m) ,$$

where $S_n := \sum_{k=1}^n B_k$, $S_0 := 0$. Conclude that for each $n \geq 0$, the distribution of Q_n is the same as that of $M_n = \max_{0\leq m\leq n} S_m$.

(2) Set $M_\infty := \lim_{n\to\infty} M_n \in \mathbb{N} \cup \{\infty\}$. Conclude that

$$\lim_{n\to\infty} P(Q_n = j) = P(M_\infty = j) .$$

(3) Set $\mu := E[B_1]$. Use the WLLN to show that when $\mu > 0$, $P(M_\infty = \infty) = 1$. Do the same when $\mu = 0$ (use a previous exercise). Conclude that when $E[B_1] \geq 0$, $P(Q_n = j) \to 0$ (the queue grows infinitely long).

(4) Assume now $\mu < 0$. Use the SLLN to conclude that $P(M_\infty < \infty) = 1$, and therefore Q_n has a limiting distribution $\nu_j := \lim_{n\to\infty} P(Q_n = j) = P(M_\infty = j)$, with $\sum_j \nu_j = 1$.

(5) Consider the case where the B_n are Bernoulli with parameter p. Proceed as in part (4) of Exercise 2.9 to compute the distribution of M_∞, and obtain

$$\lim_{n\to\infty} P(Q_n = j) = \begin{cases} 0 & \text{if } p \geq q \,, \\ \frac{q-p}{q}\left(\frac{p}{q}\right)^i & \text{if } p < q \,. \end{cases}$$

(6) Generalize the preceding to the case where $B_n \in \{\pm 1, 0\}$, $p = P(B_1 = +1)$, $q = P(B_1 = -1)$. Here the idea is that M_∞ is distributed in the same way as $\sup_n Y_n$ where Y_n is the random walk corresponding to Bernoulli variables with parameter $\frac{p}{p+q}$.

EXERCISE 2.11. The Branching Process. Let $Y \in \{0, 1, 2, \ldots\}$ have distribution $P(Y = k) = p_k$, $\sum_k p_k = 1$. Consider an array $(Y_i^{(j)}, i, j \geq 1)$ of independent random variables which all have the same distribution as Y. We define a process describing the evolution of a population in which all individuals can have children, and the number of the ith individual of the jth generation is $Y_i^{(j)}$. We consider the number of individuals of the population at generation n, denoted X_n and that there is exactly one individual at generation 0: $X_0 := 1$. Then, for $n \geq 1$,

$$X_{n+1} := \begin{cases} Y_1^{(n)} + \cdots + Y_{X_n}^{(n)} & \text{if } X_n > 0\,, \\ 0 & \text{if } X_n = 0\,. \end{cases}$$

We say the population survives if $X_n > 0$ for all $n \geq 1$, and dies if there exists n such that $X_n = 0$. Clearly, the mean number of children per individual, $\lambda := E[Y]$, is determinant for the survival of the population. The problem is interesting only if one assumes that $p_0 = P(Y = 0) > 0$. Let π denote the probability that the population dies out. We will show the following result:

$$\lambda \leq 1 \;\Rightarrow\; \pi = 1\,,$$
$$\lambda > 1 \;\Rightarrow\; \pi < 1\,.$$

(1) Show that $\pi = \lim_n \pi_n$, where $\pi_n := P(X_n = 0)$.
(2) Consider the generating function of Y, $f(t) := E[e^{tY}]$. Show that $f(0) > 0$, $f(1) = 1$.
(3) Show that f is differentiable at all $t \in (-1, +1)$ and that $\lim_{t\to 0^-} f'(t) = \lambda$. Show that f is convex.
(4) Let $f_n(t) := E[e^{tX_n}]$. Show that $f_n(0) = \pi_n$, and that

$$f_{n+1}(t) = f_n(f(t))\,, \quad \forall t \in [0, 1]\,.$$

Hint: to study X_{n+1}, condition on X_n.
(5) Show that $\pi_{n+1} = f(\pi_n)$. Take the limit $n \to \infty$ and study the solutions of the limiting equation in function of λ.

CHAPTER 3

Kolmogorov: Extension Theorem and 0-1 Laws

A great deal of Probability Theory is to state convergence results for sequences of i.i.d variables $X_1, X_2, \ldots$. We give here the construction that shows that such sequences do indeed exist, and then give general asymptotic features of these sequences, known as $0-1$ Laws.

3.1. The Extension Theorem

The main existence theorem for families of independent random variables with prescribed distributions is the following.

THEOREM 3.1.1. *Let $(\nu_n)_{n\geq 1}$ be a sequence of probability measures on the line $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a sequence of independent random variables $(X_n)_{n\geq 1}$ defined on $(\Omega, \mathcal{F}, P)$ such that for each $n \geq 1$, the distribution of X_n is given by ν_n: for all Borel set $B \in \mathcal{B}(\mathbb{R})$, $P(X_n \in B) = \nu_n(B)$.*

This will follow from a more general result, *Kolmogorov's Extension Theorem*, which allows to construct sequences $(X_n)_{n\geq 1}$ with particular dependencies, for example Markov chains (see Theorem 4.1.2).

The natural space, to construct sequences of real variables is the infinite product $\mathbb{R}^{\mathbb{N}}$, elements of which are sequences $\omega = (\omega_1, \omega_2, \ldots)$, $\omega_k \in \mathbb{R}$. Let $\mathcal{I}$ denote the set of intervals of the line of the type $(-\infty, a)$, $[a, b)$, or $[b, +\infty)$. A simple rectangle in $\mathbb{R}^n$ is a set of the form $I_1 \times \cdots \times I_n$, where each $I_k \in \mathcal{I}$. Let $\mathcal{R}^n$ denote the algebra generated by simple rectangles. As can be verified easily, elements of $\mathcal{R}^n$ are finite disjoint unions of simple rectangles. The product σ-algebra on $\mathbb{R}^n$ is $\mathcal{B}(\mathbb{R}^n) := \sigma(\mathcal{R}^n)$.

For $m \leq n$, define the canonical projection $\pi_m^n : \mathbb{R}^n \to \mathbb{R}^m$ by

$$\pi_m^n(\omega_1, \ldots, \omega_n) := (\omega_1, \ldots, \omega_m)\,. \tag{3.1.1}$$

When $n = \infty$, we write π_m for the projection $\pi_m : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^m$, defined by

$$\pi_m(\omega_1, \omega_2, \dots) := (\omega_1, \dots, \omega_m) . \tag{3.1.2}$$

We have

$$\pi^n_m \circ \pi_n = \pi_m , \qquad \text{and} \qquad (\pi^n_m)^{-1} = \pi_n \circ \pi_m^{-1} . \tag{3.1.3}$$

On $\mathbb{R}^{\mathbb{N}}$, the algebra of cylinders of size n is defined by $\mathcal{C}_n := \pi_n^{-1}(\mathcal{R}^n)$. Let $\mathcal{C} := \bigcup_{n\geq 1} \mathcal{C}_n$. Since $\mathcal{C}_n \subset \mathcal{C}_p$ when $n \leq p$, $\mathcal{C}$ is an algebra on $\mathbb{R}^{\mathbb{N}}$, called the algebra of cylinders. Finally, $\mathcal{B}(\mathbb{R}^{\mathbb{N}}) := \sigma(\mathcal{C})$.

Theorem 3.1.2 (Kolmogorov's Extension Theorem). *Let $(\mu_n)_{n\geq 1}$, where μ_n is a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, satisfying the following compatibility condition: for all $n \geq m$,*

$$\mu_n \circ (\pi^n_m)^{-1} = \mu_m . \tag{3.1.4}$$

Then there exists a unique probability measure P on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ such that

$$P \circ \pi_n^{-1} = \mu_n , \ \forall n \geq 1 . \tag{3.1.5}$$

Proof. The compatibility condition (3.1.4) will allow to define a probability P on the algebra $\mathcal{C}$, which we then extend to $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ using Carathéodory's Extension Theorem. Let $B \in \mathcal{C}$, i.e. $B = \pi_n^{-1}(R)$ for some $n \geq 1$, and $R \in \mathcal{R}^n$. Define

$$P(B) := \mu_n(R) .$$

Since $\mathcal{C}_n \subset \mathcal{C}_p$ when $n \leq p$, the representation of B is not unique, and we must verify that each such representation leads to the same number $P(B)$. So assume $B = \pi_n^{-1}(R) = \pi_p^{-1}(R')$, where $R' \in \mathcal{C}_p$ with $p \geq n$. This implies, by (3.1.4) and (3.1.3),

$$\mu_n(R) = \mu_p((\pi^p_n)^{-1}(R)) = \mu_p((\pi_p(\pi_n^{-1}(R))) = \mu_p(R') .$$

This shows that P is well defined on $\mathcal{C}$. We verify that P is additive: let $A, B \in \mathcal{C}$, $A \cap B = \emptyset$. There exist m, n such that $A = \pi_n^{-1}(R)$, $B = \pi_m^{-1}(R')$, where $R \in \mathcal{R}^n$, $R' \in \mathcal{R}^m$. Letting $p := \max\{n, m\}$, we get $A \cup B = \pi_p^{-1}((\pi^p_n)^{-1}(R) \cup (\pi^p_m)^{-1}(R'))$, and since $(\pi^p_n)^{-1}(R) \cap (\pi^p_m)^{-1}(R') = \emptyset$,

$$\begin{aligned} P(A \cup B) &= \mu_p((\pi^p_n)^{-1}(R) \cup (\pi^p_m)^{-1}(R')) \\ &= \mu_p((\pi^p_n)^{-1}(R)) + \mu_p(\pi^p_m)^{-1}(R')) \\ &= \mu_n(R) + \mu_m(R') \\ &= P(A) + P(B) . \end{aligned}$$

To verify σ-additivity, let $B_n \in \mathcal{C}$, $B_n \searrow \emptyset$. We will show that $P(B_n) \searrow 0$ *ad absurdum*: assuming that $P(B_n) \searrow 2\lambda > 0$, we will show, using a compactness argument, that $\bigcap_n B_n \neq \emptyset$, a contradiction.

We can assume, without loss of generality, that $B_n \in \mathcal{C}_n$, i.e. $B_n = \pi_n^{-1}(R_n)$ and $P(B_n) = \mu_n(R_n)$ for some $R_n \in \mathcal{R}^n$. Since μ_n is a probability measure, there exists a compact $\widehat{R}_n \subset R_n$ such that

$$\mu_n(R_n \backslash \widehat{R}_n) \leq \frac{\lambda}{2^n}\,. \tag{3.1.6}$$

Then, there exists $R_n' \in \mathcal{R}^n$ and a compact $\widetilde{R}_n$ such that $\widehat{R}_n \subset R_n' \subset \widetilde{R}_n \subset R_n$, and such that, with $B_n' := \pi_n^{-1}(R_n')$,

$$P(B_n \backslash B_n') = P(\pi_n^{-1}(R_n \backslash R_n')) = \mu_n(R_n \backslash R_n') \leq \mu_n(R_n \backslash \widehat{R}_n) \leq \frac{\lambda}{2^n}\,.$$

Set $\widetilde{B}_n := \pi_n^{-1}(\widetilde{R}_n)$, and $\widetilde{D}_n := \bigcap_{k=1}^n \widetilde{B}_k$, which is decreasing. We will show that $\bigcap_n \widetilde{D}_n \neq \emptyset$, which yields a contradiction since $\bigcap_n \widetilde{D}_n \subset \bigcap_n B_n = \emptyset$. Let also $D_n' := \bigcap_{k=1}^n B_k'$.

$$P(D_n') \geq P(D_n' \cap B_n) = P(B_n) - P(B_n \backslash D_n') \geq 2\lambda - \sum_{k=1}^n P(B_k \backslash B_k') \geq \lambda\,.$$

In particular, $D_n' \neq \emptyset$, and so $\widetilde{D}_n \neq \emptyset$. For each n, take $\omega^n = (\omega_k^n)_{k\geq 1} \in \widetilde{D}_n$. Consider the sequence $(\omega_1^n)_{n\geq 1}$. Since for all $n \geq 1$ $\omega_1^n \in \pi_1(\widetilde{D}_n) \subset \pi_1(\widetilde{D}_1)$, which is compact, there exists a subsequence of $1, 2, \dots$, denoted $(n(1,j))_{j\geq 1}$, such that

$$\omega_1^* := \lim_{j\to\infty} \omega_1^{n(1,j)} \quad \text{exists.}$$

Then consider the subsequence $(\omega_2^{n(1,j)})_{j\geq 1}$. Since $\omega_2^{n(1,j)} \in \pi_2(\widetilde{D}_1)$, which is compact, there exists a further subsequence of $(n(1,j))_{j\geq 1}$, denoted $(n(2,j))_{j\geq 1}$, such that

$$\omega_2^* := \lim_{j\to\infty} \omega_2^{n(2,j)} \quad \text{exists.}$$

This procedures goes on until having constructed, for all k, some subsequence of $(n(k-1,j))_{j\geq 1}$ denoted $(n(k,j))_{j\geq 1}$, such that

$$\omega_k^* := \lim_{j\to\infty} \omega_k^{n(k,j)} \quad \text{exists.}$$

Consider the diagonal sequence $(n(j,j))_{j\geq 1}$, which is a subsequence of all the previous ones. Therefore,

$$\lim_{j\to\infty}(\omega_1^{n(j,j)},\dots,\omega_k^{n(j,j)}) = (\omega_1^*,\dots,\omega_k^*)$$

for all k. Since $(\omega_1^{n(j,j)},\dots,\omega_k^{n(j,j)}) \in \pi_k(\widetilde{D}_{n(j,j)})$ for large enough j, and since $\pi_k(\widetilde{D}_{n(j,j)})$ is closed, $\pi_k(\widetilde{D}_{n(j,j)}) \subset \pi_k(\widetilde{D}_k)$, we have $(\omega_1^*,\dots,\omega_k^*) \in \pi_k(\widetilde{D}_k)$. This implies that the full sequence $\omega^* := (\omega_1^*,\omega_2^*,\dots) \in \widetilde{D}_k$ for all k, which proves that $\bigcap_k \widetilde{D}_k \neq \emptyset$. □

Remark 3.1.1. Usually, the cylinders are defined to be sets of the form $\pi_n^{-1}(B)$, where $B \in \mathcal{B}(\mathbb{R}^n)$ (rather than $B \in \mathcal{R}^n$). This definition can be shown to lead to the same σ-field $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$. The only difference, in the proof of the Extension Theorem, is that to guarantee the existence of the compact set $\widehat{R}_n$ in (3.1.6) requires a classical theorem from measure theory that says that any measure on a complete separable metric space is *tight*, i.e. the measure of any Borel set can be approximated from below by compact sets. We will come back to these properties in Chapter 7.2.

Proof of Theorem 3.1.1: Consider $\Omega := \mathbb{R}^{\mathbb{N}}$, with the σ-algebra $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ defined above. For each $n \geq 1$, consider the product measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ defined by $\mu_n := \nu_1 \otimes \cdots \otimes \nu_n$. The sequence $(\mu_n)_{n\geq 1}$ clearly satisfies the compatibility condition (3.1.4). The measure P is thus the one given by Kolmogorov's Extension Theorem. Moreover, defining $X_k : \Omega \to \mathbb{R}$ by $X_k(\omega) := \omega_k$, we have that

$$P(X_{i_1} \in B_1,\dots,X_{i_k} \in B_k) = \nu_{i_1}(B_1)\dots\nu_{i_k}(B_k) = P(X_{i_1} \in B_1)\dots P(X_{i_k} \in B_k)\,,$$

which shows that the X_k are independent. □

Kolmogorov's Extension Theorem shows that one can always consider a family of random variables $(X_n)_{n\geq 1}$ as constructed on the product space $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$. Namely, assume $(X_n)_{n\geq 1}$ lives on a probability space $(\Omega, \mathcal{F}, Q)$. Define the marginals

$$\mu_n(B_1 \times \cdots \times B_n) := Q(X_1 \in B_1,\dots,X_n \in B_n)\,,$$

and construct the measure P on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ using the Kolmogorov Extension Theorem. Then, define for all $\omega = (\omega_1,\omega_2,\dots)$,

$$\widetilde{X}_n(\omega) := \omega_n\,.$$

Then $(\widetilde{X}_n)_{n\geq 1}$ is distributed exactly as $(X_n)_{n\geq 1}$. The advantage of working on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ makes it easier to define *shifting* and *permuting* operations, as will be seen later.

3.2. The 0-1 Law of Kolmogorov

Consider a sequence of random variables $(X_n)_{n\geq 1}$ defined on a same probability space $(\Omega, \mathcal{F}, P)$, for example on the product space of the previous section. Events which are relevant in the study of $(X_n)_{n\geq 1}$ are events which are sensitive only to *asymptotic* properties of this sequence. For example, the Strong Law of Large Numbers asserts that if the sequence is i.i.d., then

$$\frac{S_n}{n} \to E[X_1] \quad \text{a.s.}$$

One could ask, for example: why is the preceding limit not random? It happens that the event $\{\frac{S_n}{n} \text{ has a limit}\}$ does not depend on any finite number of variables X_k, and such events have the particularity of not being random, in that their probability is either 0 or 1, as we shall see below.

Let $\mathcal{F}_n^\infty := \sigma(X_{n+1}, X_{n+2}, \dots)$ denote the smallest σ-algebra for which each X_k, $k > n$, is measurable. The sequence of σ-algebras $(\mathcal{F}_n^\infty)_{n\geq 1}$ is decreasing: $\mathcal{F}_n^\infty \supset \mathcal{F}_{n+1}^\infty$. The **tail σ-field** is defined by

$$\mathcal{T}_\infty := \bigcap_{n\geq 1} \mathcal{F}_n^\infty \,. \tag{3.2.1}$$

The events in $\mathcal{T}_\infty$ are called **tail events**. Observe for example that $\{\lim_n X_n = c\}$ is a tail event, since

$$\{\lim_{n\to\infty} X_n = c\} = \underbrace{\bigcap_{m\geq 1} \underbrace{\bigcup_{n\geq m} \underbrace{\bigcap_{j\geq n} \underbrace{\{|X_j - c| \leq \frac{1}{m}\}}_{\in \mathcal{F}_j^\infty \subset \mathcal{F}_n^\infty}}_{\in \mathcal{F}_n^\infty \subset \mathcal{F}_m^\infty}}_{\in \mathcal{F}_m^\infty}} \,.$$

On the other hand, $\{\sum_{n\geq 1} X_n = c\} \notin \mathcal{T}_\infty$.

THEOREM 3.2.1 (0-1 Law of Kolmogorov). *Assume the variables $(X_n)_{n\geq 1}$ are independent. Then $P(A) \in \{0, 1\}$ for all $A \in \mathcal{T}_\infty$.*

The proof relies on two lemmas. The first, although intuitively obvious, requires a proof of a rather abstract nature; see Corollary B.0.3 of Appendix B.

LEMMA 3.2.1. *Assume the variables* $(X_n)_{n\geq 1}$ *are independent. Then for all* $k \geq 1$, $\sigma(X_1, \ldots, X_k)$ *and* $\sigma(X_{k+1}, \ldots)$ *are independent.*

LEMMA 3.2.2 (Approximation Lemma). *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, where* $\mathcal{F}$ *is generated by an algebra* $\mathcal{A}$. *Then for all* $E \in \mathcal{F}$ *there exists a sequence* $A_n \in \mathcal{A}$ *such that* $P(E \triangle A_n) \to 0$.

PROOF. Let $\mathcal{G}$ denote the family of sets $E \in \mathcal{F}$ for which the property holds. We show that $\mathcal{G}$ is a σ-algebra; since it contains $\mathcal{A}$, this will show the lemma. Obviously, $\emptyset, \Omega \in \mathcal{G}$, and since $P(E^c \triangle A_n^c) = P(E \triangle A_n)$, $\mathcal{G}$ is stable under complementation. Then, let $E_n \in \mathcal{G}$. We verify that $E = \bigcup_n E_n \in \mathcal{G}$. Take $\epsilon > 0$. Consider, for each $n \geq 1$, some $A_n \in \mathcal{A}$ such that $P(E_n \triangle A_n) \leq \epsilon 2^{-(n+1)}$. If $D := \bigcup_{n\geq 1} A_n$, we have $P(E \triangle D) = \lim_{n\to\infty} P(E \triangle D_N)$, where $D_N := \bigcup_{n=1}^N A_n$. But

$$\begin{aligned} P(E \triangle D) &= P(E \cap D^c) + P(D \cap E^c) \\ &\leq \sum_{n\geq 1} P(E_n \cap D^c) + \sum_{m\geq 1} P(A_m \cap E^c) \\ &\leq \sum_{n\geq 1} P(E_n \cap A_n^c) + \sum_{m\geq 1} P(A_m \cap E_m^c) = \sum_{n\geq 1} P(E_n \triangle A_n) \leq \frac{\epsilon}{2}. \end{aligned}$$

Therefore, if N is sufficiently large, $P(E \triangle D_N) \leq \epsilon$. Since $D_N \in \mathcal{A}$, this shows that $E \in \mathcal{G}$. □

PROOF OF THEOREM 3.2.1: Let $A \in \mathcal{T}_\infty \subset \sigma(X_1, X_2, \ldots)$. Since $\sigma(X_1, X_2, \ldots) = \sigma(\mathcal{A})$, where $\mathcal{A} = \bigcup_{n\geq 1} \sigma(X_1, \ldots, X_n)$ (see Exercise 3.4), we can use the Approximation Lemma: there exists a sequence $A_n \in \mathcal{A}$ such that $\lim_n P(A \triangle A_n) = 0$. Therefore, $P(A) = P(A \cap A) = \lim_n P(A \cap A_n)$. Since $A_n \in \sigma(X_1, \ldots, X_m)$ for some sufficiently large m, and since $A \in \sigma(X_{m+1}, \ldots)$, Lemma 3.2.1 implies that A and A_n are independent: $P(A \cap A_n) = P(A)P(A_n)$. Therefore, $P(A) = \lim_n P(A)P(A_n) = P(A)^2$ and so $P(A)$ equals 0 or 1. □

An immediate consequence of the $0-1$-Law of Kolmogorov is

THEOREM 3.2.2. *Assume the variables* $(X_n)_{n\geq 1}$ *are independent. Then any* $\mathcal{T}_\infty$*-measurable random variable* Z *is almost surely constant: there exists* $-\infty \leq c \leq +\infty$ *such that* $P(Z = c) = 1$.

Proof. By hypothesis, $\{Z \leq x\} \in \mathcal{T}_\infty$ and therefore $P(Z \leq x) \in \{0,1\}$ for all x. Since $x \mapsto P(Z \leq x)$ is non-decreasing and right-continuous, one can define $c := \inf\{x : P(Z \leq x) = 1\}$, which gives $P(Z = c) = 1$. □

For example, consider the simple random walk on $\mathbb{Z}$, denoted $(S_n)_{n\geq 1}$. Define

$$Z := \limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} .$$

Clearly, Z is $\mathcal{T}_\infty$-measurable, and by the previous theorem there exists $c \in \mathbb{R} \cup \{\pm\infty\}$ such that $P(Z = c) = 1$. As we saw in Section 2, $c = +\infty$ if $p > \frac{1}{2}$, $c = -\infty$ if $p < \frac{1}{2}$, and by the Law of the Iterated Logarithm, $c = 1$ if $p = \frac{1}{2}$.

Let us see further consequences of the $0-1$ Law, in the general case where the variables X_k are independent. Of central interest in the asymptotic properties of $(X_n)_{n\geq 1}$ are the partial sums $S_n = \sum_{k=1}^n X_k$. These lead to two particular types of random variables. First, if one considers the random series $\lim_n S_n = \sum_{n\geq 1} X_n$, then the 0-1 Law says that

$$P\Big(\sum_n X_n \text{ converges}\Big) \in \{0,1\} .$$

To guarantee that the above number is 1 can require some work. We will see later how a simple condition on the variances of the X_ns allows to show that this probability is indeed 1. See Theorem 5.4.3 in the chapter on martingales.
Second, let $(a_n)_{n\geq 1}$ with $a_n \to \infty$. Define the average

$$A_n := \frac{S_n}{a_n} .$$

Laws of Large of Numbers study conditions under which these averages have limits. Since one can always write, for all $m \leq n$,

$$A_n = \frac{1}{a_n}\sum_{k=1}^{m-1} X_k + \frac{1}{a_n}\sum_{k=m}^{n} X_k ,$$

we see that when $n \to \infty$, the first term always vanishes. This shows that $\{\lim_n A_n \text{ exists}\}$ is invariant under a change of finitely many random variables X_k. Therefore, $P(\lim_n A_n \text{ exists}) \in \{0,1\}$. Considering in particular random variables of the type $X_n - E[X_n]$ with $a_n = n$,

$$P\Big(\frac{1}{n}\sum_{k=1}^{n}(X_k - E[X_k]) \text{ converges}\Big) \in \{0,1\}\,,$$

which shows that the Strong Law of Large numbers reduces to giving conditions on the variables X_k under which this probability is 1 (and not 0!).

More generally, a tail-σ-algebra can be created starting from any decreasing countable collection of σ-algebras: let $(\mathcal{F}_n)_{n\geq 1}$ be such that $\mathcal{F}_{n+1} \subset \mathcal{F}_n$. Then One can define as before

$$\mathcal{T}_\infty := \bigcap_{n\geq 1} \mathcal{F}_n\,. \tag{3.2.2}$$

Above we had $\mathcal{F}_n = \sigma(X_n, X_{n+1}, \dots)$. Another example can be constructed as follows: let $E_1, E_2, \dots$ be any sequence of events, and take $\mathcal{F}_n := \sigma(E_n, E_{n+1}, \dots)$. Then the events $\limsup_n E_n$ and $\liminf_n E_n$ clearly belong to $\mathcal{T}_\infty$. Moreover, if the events E_n are independent, we conclude that both $P(\limsup_n E_n)$ and $P(\liminf_n E_n)$ equal either 0 or 1. The Borel-Cantelli then gives a criterium to decide whether it is 0 or 1 by considering the convergence of the series $\sum_n P(E_n)$.

3.3. The 0-1 Law of Hewitt-Savage

In the case of the simple random walk on $\mathbb{Z}$, it can be seen that $\{S_n = 0 \text{ i.o.}\}$ is *not* a tail event. Therefore, one can not deduce directly from the 0-1 Law of Kolmogorov that $P(S_n = 0 \text{ i.o.}) \in \{0,1\}$, although we know this to be true by the recurrence properties of the random walk. Observe that although $\{S_n = 0 \text{ i.o.}\}$ may be sensitive to a change of a finite number of the variables X_k, it nevertheless remains invariant under the *permutation* of two variables X_n, X_m: the Hewitt-Savage gives a 0-1 Law for events that are invariant under finite permutations of variables.

The notion of “permutation” of two variables needs to be made precise. Observe that since the events under consideration are defined only in terms of the variables X_k (that is, contained in the σ-algebra $\sigma(X_1, \dots)$) we might as well use for the underlying probability space the infinite product

$$(\Omega, \mathcal{F}) := (\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$$

constructed in Section 3.1, and take $X_k(\omega) := \omega_k$. Notice that in this case, $\sigma(X_1, \dots, X_n) \equiv \mathcal{B}(\mathbb{R}^n)$ and $\sigma(X_1, X_2, \dots) \equiv \mathcal{B}(\mathbb{R}^{\mathbb{N}})$.

Let $\varphi : \mathbb{N} \to \mathbb{N}$ be a finite permutation, i.e. a bijection such that $\{k \in \mathbb{N} : \varphi(k) \neq k\}$ is finite. Let Π denote the set of all finite permutations. Define $\varphi : \Omega \to \Omega$ by

$$\varphi(\omega_1, \omega_2, \dots) := (\omega_{\varphi(1)}, \omega_{\varphi(2)}, \dots)\,.$$

LEMMA 3.3.1. *For each $\varphi \in \Pi$, $\varphi : \Omega \to \Omega$ is measurable: $\forall A \in \mathcal{F}$, $\varphi^{-1}(A) \in \mathcal{F}$.*

PROOF. Let $\mathcal{D}$ denote the class of events $A \in \mathcal{F}$ for which $\varphi^{-1}(A) \in \mathcal{F}$. Then clearly $\mathcal{D} \supset \mathcal{C}$, and is easy to verify that $\mathcal{D}$ is a σ-algebra. Therefore, $\mathcal{D} = \mathcal{F}$. □

LEMMA 3.3.2. *Assume the variables $(X_n)_{n\geq 1}$ are independent and identically distributed. Then $P \circ \varphi^{-1} = P$ for all $\varphi \in \Pi$.*

PROOF. Take $\varphi \in \Pi$ and define $Q := P \circ \varphi^{-1}$. Then, as can be easily seen, $Q(A) = P(A)$ for all cylinder $A \in \mathcal{C}$, and therefore Q and P coincide everywhere. □

An event $A \in \mathcal{F}$ is called exchangeable if $\varphi^{-1}(A) = A$ $\forall \varphi \in \Pi$. The class of exchangeable events forms a σ-algebra denoted $\mathcal{E}$. Observe that all tail events are exchangeable: $\mathcal{T}_\infty \subset \mathcal{E}$. Nevertheless, $\{S_n = 0 \text{ i.o.}\} \in \mathcal{E} \backslash \mathcal{T}_\infty$. The following is thus a generalization of the 0-1 Law of Kolmogorov.

THEOREM 3.3.1 (0-1 Law of Hewitt-Savage). *Assume the variables $(X_n)_{n\geq 1}$ are independent and identically distributed. Then $P(A) \in \{0, 1\}$ for all $A \in \mathcal{E}$.*

PROOF OF THEOREM 3.3.1: Let $E \in \mathcal{E}$. We proceed as in the proof of the 0-1 Law of Kolmogorov. Since $\mathcal{E} \subset \sigma(X_1, X_2, \dots)$, and since $\sigma(X_1, X_2, \dots)$ is generated by the algebra $\mathcal{A} = \bigcup_{n\geq 1} \sigma(X_1, \dots, X_n)$, we can use the Approximation Lemma: there exists a sequence $A_n \in \mathcal{A}$ such that $\lim_n P(E \triangle A_n) = 0$. We can assume without loss of generality that $A_n \in \sigma(X_1, \dots, X_n)$. For each $n \geq 1$, consider the finite permutation $\varphi_n : \mathbb{N} \to \mathbb{N}$ defined by

$$\varphi_n(k) := \begin{cases} k+n & \text{if } k = 1, \dots, n\,, \\ k-n & \text{if } k = n+1, \dots, 2n\,, \\ k & \text{if } k > 2n. \end{cases}$$

Observe that φ_n satisfies $\varphi_n^2 = \text{id}$, i.e. $\varphi_n^{-1} = \varphi_n$, and that $\varphi_n^{-1}(A_n) \in \sigma(X_{n+1}, \dots)$. Therefore, since the variables X_k are independent, Lemma 3.2.1 gives $P(A_n \cap \varphi_n^{-1}(A_n)) = P(A_n)P(\varphi_n^{-1}(A_n))$. By Lemma 3.3.2, $P(\varphi_n^{-1}(A_n)) = P(A_n)$. Therefore, $P(A_n \cap \varphi_n^{-1}(A_n)) \to P(E)^2$ when $n \to \infty$, and we can write

$$\begin{aligned} |P(E) - P(E)^2| &= \lim_{n\to\infty} |P(E) - P(A_n \cap \varphi_n^{-1}(A_n))| \\ &\le \limsup_{n\to\infty} [P(E \triangle A_n) + P(E \triangle \varphi_n^{-1}(A_n))] \\ &= \limsup_{n\to\infty} P(E \triangle \varphi_n^{-1}(A_n)) \\ &= \limsup_{n\to\infty} P(E \triangle \varphi_n^{-1}(A_n))\,. \end{aligned} \tag{3.3.1}$$

We used the inequality $|P(E) - P(C_1 \cap C_2)| \le P(E\triangle C_1) + P(E\triangle C_2)$, which can easily be verified [1]. Now since $E \in \mathcal{E}$, we have $E = \varphi_n^{-1}(E)$. Therefore,

$$P(E\triangle\varphi_n^{-1}(A_n)) = P(\varphi_n^{-1}(E)\triangle\varphi_n^{-1}(A_n)) = P(\varphi_n^{-1}(E\triangle A_n)) = P(E\triangle A_n)\,,$$

where we used the fact that φ_n is a bijection in the second equality, and Lemma 3.3.2 in the last. But since $P(E\triangle A_n) \to 0$ when $n \to \infty$, (3.3.1) shows that $P(E) = P(E)^2$, and finishes the proof. ☐

The 0-1 Law of Hewitt-Savage gives the following weak characterization of recurrence for the simple random walk on $\mathbb{Z}$, denoted S_n.

THEOREM 3.3.2. *Let $(S_n)_{n\ge1}$ denote the simple random walk on $\mathbb{Z}$. Then, with probability one, exactly one of the following events occurs:*

(1) $\{\lim_{n\to\infty} S_n = +\infty\}$,
(2) $\{\lim_{n\to\infty} S_n = -\infty\}$,
(3) $\{\liminf_{n\to\infty} S_n = -\infty\} \cap \{\limsup_{n\to\infty} S_n = +\infty\}$.

PROOF. Define $Z_+ := \limsup_{n\to\infty} S_n$, $Z_- := \liminf_{n\to\infty} S_n$. The 0-1 Law of Hewitt-Savage implies that these random variables are constant almost surely: there exists $c_+, c_- \in \mathbb{R} \cup \{\pm\infty\}$ such that $P(Z_\pm = c_\pm) = 1$. It suffices to show that the constants $c_\pm$ can't be

[1] First verify that $|P(E) - P(B)| \le P(E\triangle B)$, and then take $B = C_1 \cap C_2$.

finite. Namely, assume that c_+ is finite. Then, since $S'_n := S_{n+1} - X_1$ has the same distribution as S_n, we must have $c_+ = c_+ - X_1$, i.e. $X_1 = 0$ almost surely, which is absurd. □

3.4. Exercises

EXERCISE 3.1. Show that the σ-algebra on $\mathbb{R}^n$ generated by rectangles, i.e. set of the form $B_1 \times \dots B_n$, where $B_k \in \mathcal{B}(\mathbb{R})$, equals $\mathcal{B}(\mathbb{R}^n)$.

EXERCISE 3.2. Prove Theorem 3.1.1.

EXERCISE 3.3. Show that the Kolmogorov's Extension Theorem generalizes easily to families $(X_t)_{t\in I}$, where I is an arbitrary set of indices, for example $I = [0, 1]$. (voir les deux gus)

EXERCISE 3.4. Seja $X_1, X_2, \dots$ uma seqüência qualquer de variáveis aleatórias. Mostre que

$$\sigma(X_1, X_2 \dots) = \sigma\Big(\bigcup_{n\geq 1} \sigma(X_1, \dots, X_n)\Big).$$

EXERCISE 3.5. Show that a random variable which is $\mathcal{T}_\infty$-measurable is almost-surely constant.

EXERCISE 3.6. [**Chu01**] p. 270. Let $(X_n)_{n\geq 1}$ be independent, such that $P(X_n = 4^{-n}) = P(X_n = -4^{-n}) = \frac{1}{2}$. Is the tail field $\mathcal{T}_\infty = \sigma(S_n, n \geq 1)$ is trivial?

EXERCISE 3.7. Let $\mathcal{T}_\infty := \bigcap_{n\geq 1} \sigma(X_n, X_{n+1}, \dots)$. Set $S_n = \sum_{k=1}^n X_k$. Determine which of the following events are in $\mathcal{T}_\infty$.

$$\{X_n \in I_n \text{ i.o.}\}, \quad \big\{\lim_n S_n \text{ exists}\big\}, \quad \big\{\lim_n S_n \text{ exists and is } \leq c\big\},$$

$$\big\{\limsup_n X_n < \infty\big\}, \quad \big\{\limsup_n S_n = \infty\big\}, \quad \big\{\limsup_n S_n > 0\big\},$$

What must the sequence $(c_n)_{n\geq 1}$ satisfy in order to guarantee that the event $\{\limsup_n S_n/c_n > x\} \in \mathcal{T}_\infty$?

EXERCISE 3.8. Consider independent site percolation on $\mathbb{Z}^d$, $d \geq 1$. Characterize the tail-σ-field. Which of the events

$$\{|C_0| = \infty\}, \quad \{\text{there exists an infinite connected cluster}\}$$

are trivial?

EXERCISE 3.9. Let (S_n) denote the simple symmetric random walk. Show that $P(\limsup_n S_n = \infty) = 1$, without using the Law of the Iterated Logarithm.

EXERCISE 3.10. Consider the simple random walk on $\mathbb{Z}$. Is it that $\{S_n = 0 \text{ i.o.}\} \in \mathcal{T}_\infty$? Show that

- if $p \neq \frac{1}{2}$, then $P(S_n = 0 \text{ i.o.}) = 0$ (in particular, (S_n) is transient).
- if $p = \frac{1}{2}$, then $P(S_n = 0 \text{ i.o.}) = 1$ (in particular, (S_n) is recurrent).

Hint: the first follows by a direct application of Borel-Cantelli. For the second, observe that it suffices to show that $P(A^+) = P(A^-) = 1$, where $A^\pm$ are the tail events

$$A^+ = \left\{ \limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = \infty \right\}, \quad A^- = \left\{ \liminf_{n\to\infty} \frac{S_n}{\sqrt{n}} = -\infty \right\}.$$

Observe that $A^\pm = \bigcap_{k\geq 1} A_k^\pm$, where $A_k^\pm$ are the tail events

$$A_k^+ = \left\{ \limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} \geq k \right\}, \quad A_k^- = \left\{ \liminf_{n\to\infty} \frac{S_n}{\sqrt{n}} \leq -k \right\}.$$

Use the Central Limit Theorem and the inclusion (prove!) $\{\limsup_n Z_n \geq c\} \subset \limsup_n \{Z_n \geq c\}$ to show that $P(A_k^\pm) = 1$ for all k.

EXERCISE 3.11. Show that for the simple random walk on $\mathbb{Z}$, $P(T_0 < \infty) \in \{0, 1\}$.

EXERCISE 3.12. [**Wil91**] p. 229. Let $(S_n)_{n\geq 0}$ denote the SRRW. Define $\mathcal{A} := \sigma(X_1, \ldots)$, $\mathcal{T}_n := \sigma(S_{n+1}, \ldots)$. Let

$$\mathcal{L} := \bigcap_{n\geq 1} \sigma(\mathcal{A}, \mathcal{T}_n), \quad \mathcal{M} := \sigma\Big(\mathcal{A}, \bigcap_{n\geq 1} \mathcal{T}_n\Big).$$

Show that $\mathcal{L} \neq \mathcal{M}$. *Hint:* show that X_1 is $\mathcal{L}$-measurable and independent of $\mathcal{M}$.

EXERCISE 3.13. Show that for the simple random walk on $\mathbb{Z}$, $\limsup_n S_n$ is constant almost surely.

CHAPTER 4

Markov Chains

This chapter is inspired partly by [**Nev70**] and [**R.88**].

4.1. Definitions and Basic Properties

Let S be a finite or countable set, which we call state space, endowed with the σ-algebra $\mathcal{P}(S) = \{A : A \subset S\}$.

DEFINITION 4.1.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A sequence of S-valued random variables $(X_n)_{n\geq 0}$ is a Markov chain (with state space S) if for all $n \geq 0$, $X_n : \Omega \to S$ is measurable and*

$$P(X_{n+1} = x_{n+1}|X_n = x_n, \dots, X_0 = x_0) = P(X_{n+1} = x_{n+1}|X_n = x_n) \tag{4.1.1}$$

for all $x_0, \dots, x_{n+1} \in S$.

We avoided to mention that (4.1.1) holds P-almost surely, and will usually continue doing so in the sequel. We don't yet worry about the structure of the underlying probability space $(\Omega, \mathcal{F}, P)$, although a canonical choice will be made in Section 4.1.1.

We will mostly consider the case where the probability $P(X_{n+1} = x_{n+1}|X_n = x_n)$ does not depend on n, that is where

$$P(X_{n+1} = y|X_n = x) = P(X_1 = y|X_0 = x)$$

for all $n \geq 1$. In such case, the chain is called homogeneous, and the dependence among the random variables is determined by the numbers $P(X_1 = y|X_0 = x)$, called transition probabilities. Observe that these satisfy $\sum_{y\in S} P(X_1 = y|X_0 = x) = 1$ (P-a.s.) for all $x \in S$. We are interested in the study of Markov chains for which the transition probabilities are specified a priori.

DEFINITION 4.1.2. *A collection $Q(x, y)$, $x, y \in S$, is called a transition probability matrix if $Q(x, y) \in [0, 1]$ and if $\sum_{y\in S} Q(x, y) = 1$ for all $x \in S$. A homogeneous Markov chain $(X_n)_{n\geq 0}$ has transition probability matrix Q if*

$$P(X_{n+1} = y|X_n = x) = Q(x, y) \quad P\text{-a.s.}$$

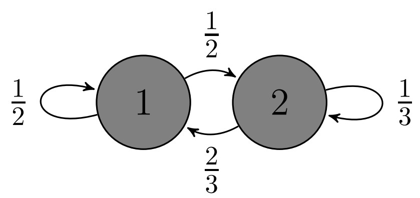
for all $n \geq 0$, $x, y \in S$.

Although it might seem trivial at this point, observe that for all $y \in S$, $x \mapsto Q(x, y)$ is measurable. The existence of a Markov chain associated to a transition probability matrix will be shown in Section 4.1.1. Before going further we give a serie of examples.

EXAMPLE 4.1.1. Independent variables furnish a trivial example of Markov chain. Let $(X_n)_{n\geq 0}$ be a sequence of i.i.d random variables with distribution μ over $(S, \mathcal{P}(S))$. If we define $Q(x, y) := \mu(y)$, then by independence,

$$P(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_0 = x_0) = \mu(x_{n+1}) \equiv Q(x_n, x_{n+1}) \,.$$

EXAMPLE 4.1.2. The two state Markov chain is defined for $S = \{1, 2\}$. An example of a transition matrix is given in the following graphical representation:



EXAMPLE 4.1.3. The random walk on $S = \mathbb{Z}^d$. We considered the simplest case of random walk in Section 2. Consider a sequence $(Y_n)_{n\geq 1}$ of $\mathbb{Z}^d$-valued independent identically distributed random variables, and denote their common distribution by p. Define $S_0 := 0$, and for all $n \geq 1$, $S_n := \sum_{k=1}^n Y_k$. The sequence $(S_n)_{n\geq 0}$ is called a random walk on $\mathbb{Z}^d$. Observe that, since Y_{n+1} is independent of $S_1, \ldots, S_n$, we have

$$\begin{aligned} P(S_{n+1} &= x_{n+1} | S_n = x_n, \ldots, S_0 = x_0) \\ &= P(Y_{n+1} = x_{n+1} - x_n | S_n = x_n, \ldots, S_0 = x_0) \\ &= P(Y_{n+1} = x_{n+1} - x_n) \\ &= P(Y_{n+1} = x_{n+1} - x_n | S_n = x_n) \\ &= P(S_{n+1} = x_{n+1} | S_n = x_n) \,. \end{aligned}$$

Therefore, since $P(Y_{n+1} = x_{n+1} - x_n) = p(x_{n+1} - x_n)$, $(S_n)_{n\geq 0}$ is a Markov chain with state space $S = \mathbb{Z}^d$ and transition matrix $Q(x, y) = p(y - x)$. When

$$p(x) = \begin{cases} \frac{1}{2d} & \text{if } \|x\|_1 = 1 \,. \\ 0 & \text{otherwise,} \end{cases} \tag{4.1.2}$$

that is when $p(\pm e_i) = \frac{1}{2d}$ where the $e_1, \dots, e_d$ are the canonical unit vectors of $\mathbb{R}^d$, the random walk is called simple, symmetric. More will be said on random walks in Section 4.3.2.

EXAMPLE 4.1.4. Uniform Random Walk on a Graph. Let $G = (V, E)$ be a simple graph without loops. For each $x \in V$, we assume that $A_x := \{y \in V : \{x, y\} \in E\}$ is finite: $|A_x| < \infty$. Setting $S \equiv V$, one can define a transition matrix by

$$Q(x,y) := \begin{cases} \frac{1}{|A_x|} & \text{if } \{x,y\} \in E\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{4.1.3}$$

The simple random walk of the previous example is a particular case.

EXAMPLE 4.1.5. The Ehrenfest chain. Consider two urns with a total of r balls. Each urn can be considered as a box with a certain number of molecules, the total number of molecules being r. At each time step, a ball is chosen at random (in either box) and its position switched to the other box. Let X_n be the number of balls in the first box at time n. Then $(X_n)_{n\geq 0}$ is a Markov Chain with state space $S = \{0, 1, 2, \dots, r\}$ and transition matrix Q given by

$$Q(k, k+1) = \frac{r-k}{r}\,, \quad Q(k, k-1) = \frac{k}{r}\,,$$

and zero otherwise.

EXAMPLE 4.1.6. Birth and death chains. Consider $S = \{0, 1, 2, \dots\}$, in which $X_n = x$ means that population at time n is x, and $Q(x, y) > 0$ only if $|x - y| \leq 1$. Therefore, the chain is determined by the transition probabilities $r_x = Q(x, x)$, $q_x = Q(x, x-1)$ (clearly, $q_0 = 0$), $p_x = Q(x, x+1)$. See Figure 1.

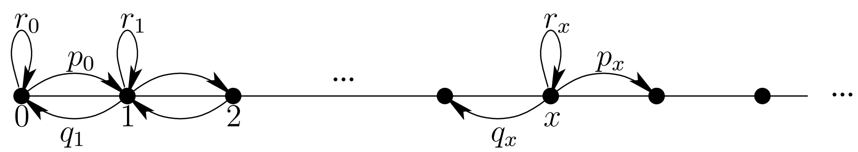


FIGURE 1. The birth and death chain.

EXAMPLE 4.1.7. Renewal chains. Consider $S = \{0, 1, 2, \dots\}$ and a sequence $(p_k)_{k\geq 1}$ with $\sum_k p_k = 1$. Then $Q(0, k) = p_k$, and $Q(k, k-1) = 1$ for all $k \geq 2$.

EXAMPLE 4.1.8. The Branching Process was introduced by Galton and Watson to understand extinction or survival of family names. Let $(Y_k^{(n)})_{n\geq 0,k\geq 1}$ be an array of i.i.d. $\mathbb{N}$-valued random variables, with distribution ρ: $P(Y_k^{(n)} = j) = \rho(j)$ for all $j \geq 1$. $Y_k^{(n)}$ is the number of children of the kth individual of the nth generation. Let $X_0 := 1$, and define the total number of individuals of the $n+1$th generation:

$$X_{n+1} := \sum_{k=1}^{X_n} Y_k^{(n)} . \tag{4.1.4}$$

Let us show that $(X_n)_{n\geq 0}$ is a Markov chain with state space $S = \{0, 1, 2, \dots\}$.

$$\begin{aligned} P(X_{n+1} = y | X_n = x_n, \dots, X_0 = x_0) &= P\Big(\sum_{k=1}^{X_n} Y_k^{(n)} = y \Big| X_n = x_n, \dots, X_0 = x_0\Big) \\ &= P\Big(\sum_{k=1}^{x_n} Y_k^{(n)} = y\Big) . \end{aligned}$$

Since the variables $Y_k^{(n)}$ are independent, the distribution of the sum $\sum_{k=1}^{x_n} Y_k^{(n)}$ is given by the convolution $\rho * \rho * \cdots * \rho$ (x_n times), which we denote by ρ^{*x_n}. This shows that $(X_n)_{n\geq 0}$ is a Markov chain with transition probability matrix given by

$$Q(x, y) = \rho^{*x}(y) \quad \forall x, y \in S \tag{4.1.5}$$

It is well known that in the subcritical case, i.e. when $\lambda := E[Y_1^{(0)}] \leq 1$, the population dies out P-almost surely. In the supercritical case, i.e. for $\lambda > 1$, then the population explodes with positive probability.

We define the iterates of a transition matrix as follows: $Q^{(1)} := Q$, and for $n \geq 2$,

$$Q^{(n)}(x, z) := \sum_{y\in S} Q^{(n-1)}(x, y)Q(y, z) . \tag{4.1.6}$$

Clearly, each $Q^{(n)}$ is well defined and is again a transition matrix. Let us give an important equivalent characterization of Markov chains.

LEMMA 4.1.1. *Let Q be a transition probability matrix. A sequence $(X_n)_{n\geq 0}$ is a Markov chain with transition matrix Q if and only if for* *all* $n \geq 1$ *and all* $x_0, \ldots, x_n \in S$,

$$P(X_0 = x_0, \ldots, X_n = x_n) = P(X_0 = x_0)Q(x_0, x_1) \ldots Q(x_{n-1}, x_n) . \tag{4.1.7}$$

In particular, if $P(X_0 = x_0) > 0$*, then*

$$P(X_n = y | X_0 = x_0) = Q^{(n)}(x_0, y) . \tag{4.1.8}$$

PROOF. Assume $(X_n)_{n\geq 0}$ is a Markov chain with transition matrix Q. If $n = 1$, (4.1.7) is trivial. Indeed, if $P(X_0 = x_0) = 0$ then $P(X_0 = x_0, X_1 = x_1) = 0$ and so $P(X_0 = x_0, X_1 = x_1) = P(X_0 = x_0)Q(x_0, x_1)$. If $P(X_0 = x_0) = 0 > 0$ the same holds. So assume that (4.1.7) holds for n. Again, if $P(X_0 = x_0, \ldots, X_n = x_n) = 0$ then $P(X_0 = x_0, \ldots, X_{n+1} = x_{n+1}) = 0$ and the result follows. If $P(X_0 = x_0, \ldots, X_n = x_n) > 0$ then

$$\begin{aligned}
&P(X_0 = x_0, \ldots, X_{n+1} = x_{n+1}) \\
&\quad = P(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_0 = x_0) P(X_0 = x_0, \ldots, X_n = x_n) \\
&\quad = P(X_{n+1} = x_{n+1} | X_n = x_n) P(X_0 = x_0) Q(x_0, x_1) \ldots Q(x_{n-1}, x_n) \\
&\quad = P(X_0 = x_0) Q(x_0, x_1) \ldots Q(x_{n-1}, x_n) Q(x_n, x_{n+1}) ,
\end{aligned}$$

which shows the validity of (4.1.7) for $n + 1$. For (4.1.8), use (4.1.7) as follows:

$$\begin{aligned}
P(X_n = x_n, X_0 = x_0) &= \sum_{x_1, \ldots, x_{n-1}} P(X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) \\
&= P(X_0 = x_0) \sum_{x_1, \ldots, x_{n-1}} Q(x_0, x_1) \ldots Q(x_{n-1}, x_n) \\
&\equiv P(X_0 = x_0) Q^{(n)}(x_0, x_n) , \qquad (4.1.9)
\end{aligned}$$

which gives (4.1.8) if $P(X_0 = x_0) > 0$. □

Observe that by (4.1.7), the transition matrix Q completely specifies the evolution of the chain, once the distribution of X_0 is known. Let therefore μ be a distribution on $(S, \mathcal{P}(S))$. When the distribution of X_0 is given by μ, we will denote the law of $(X_n)_{n\geq 0}$ by P_μ. That is, by (4.1.7),

$$P_\mu(X_0 = x_0, \ldots, X_n = x_n) = \mu(x_0) Q(x_0, x_1) \ldots Q(x_{n-1}, x_n) . \tag{4.1.10}$$

When μ is a Dirac mass, i.e. $\mu(x) = 1$ for some $x \in S$, we will write P_x rather than P_μ, and interpret x as being a deterministic initial condition. For example, (4.1.8) gives

$$P_x(X_n = y) = Q^{(n)}(x, y) . \tag{4.1.11}$$

As can be easily verified, the measure P_μ can be reconstructed by convex combination of the measures $\{P_x\}_{x\in S}$:

$$P_\mu = \sum_{x\in S} \mu(x) P_x \,.$$

We denote expectations with respect to P_x by E_x. We have

$$\begin{aligned} E_x(f(X_n)) &= \sum_{y\in S} f(y) P_x(X_n = y) \\ &= \sum_{y\in S} f(y) Q^{(n)}(x,y) \\ &\equiv Q^{(n)} f(x)\,, \end{aligned} \tag{4.1.12}$$

where for each $n \geq 1$, the function $Q^{(n)} f : S \to \mathbb{R}$ is defined by

$$Q^{(n)} f(x) := \sum_{y\in S} Q^{(n)}(x,y) f(y)\,. \tag{4.1.13}$$

(4.1.8) says that the distribution of X_n, conditionned on X_0, is given by the nth iterate of Q. This distribution can be written as $P(X_n = y|X_0 = x_0) = E[1_{\{X_n=y\}}|X_0 = x_0]$. Since we will also be interested in functions depending on the process, of the form $f : S \to \mathbb{R}$, we might therefore be interested in studying more general conditional expectations of the form $E[f(X_n)|X_0 = x_0]$.

LEMMA 4.1.2. *Let $(X_n)_{n\geq 1}$ be a Markov chain with transition matrix Q. If $f : S \to \mathbb{R}$, then for all $n \geq 0$,*

$$E[f(X_{n+1})|X_n = x_n, \ldots, X_0 = x_0] = Qf(x_n)\,. \tag{4.1.14}$$

More generally, for any set $\{i_1, \ldots, i_k\} \subset \{1, 2, \ldots, n-1\}$,

$$E[f(X_{n+1})|X_n = x_n, X_{i_k} = x_{i_k} \ldots, X_{i_1} = x_{i_1}] = Qf(x_n)\,. \tag{4.1.15}$$

4.1.1. The Canonical Chain. Up to now the underlying probability space on which the chain is defined hasn't had an important role, but one should of course verify that at least one such space exists.

THEOREM 4.1.1. *Let μ be a probability distribution on $(S, \mathcal{P}(S))$ and Q a transition probability matrix. Then there exists a probability space $(\Omega', \mathcal{F}', P'_\mu)$ and a sequence of S-valued random variables $(X_n)_{n\geq 0}$ on $(\Omega', \mathcal{F}', P'_\mu)$ which form a Markov Chain with transition probability matrix Q:*

$$P'_\mu(X_0 = x_0, \ldots, X_n = x_n) = \mu(x_0) Q(x_0, x_1) \ldots Q(x_{n-1}, x_n)\,. \tag{4.1.16}$$

Proof. By Theorem 3.1.1, one can construct simultaneously a family of i.i.d. random variables $(Y_n)_{n\geq 1}$ on the product space $\Omega' = [0,1]^{\mathbb{N}}$, with uniform distribution on $[0,1]$ with respect to the Lebesgue measure. Ω' is endowed with the product σ-algebra $\mathcal{F}'$ and P' is the product of Lebesgue measures. Let us enumerate S in an arbitrary way: $S = \{y_1, y_2, \dots\}$. Fix some initial condition $x \in S$. We define a process $(X_n^x)_{n\geq 0}$ on $(\Omega', \mathcal{F}', P')$ as follows. First, $X_0^x := x$. Then, we need to define X_1^x in such a way that $P'(X_1^x = y_k | X_0^x = x) = Q(x, y_k)$ for all $k \geq 1$. Define, for all $z \in S$,

$$\alpha_k(z) := \sum_{1\leq i\leq k} Q(z, y_i) \,.$$

Observe that $0 \leq \alpha_1(z) \leq \alpha_2(z) \leq \cdots \leq 1$, and $\alpha_k(z) \to 1$ when $k \to \infty$. Then, set

$$X_1^x = y_k \quad \text{if and only if} \quad \alpha_{k-1}(x) < Y_1 \leq \alpha_k(x) \,.$$

Clearly, $P'(X_1^x = y_k | X_0^x = x) = P'(\alpha_{k-1}(x) < Y_1 \leq \alpha_k(x)) \equiv Q(x, y_k)$. For $n \geq 2$, X_n^x is defined by

$$X_n^x = y_k \quad \text{if and only if} \quad \alpha_{k-1}(X_{n-1}^x) < Y_n \leq \alpha_k(X_{n-1}^x) \,.$$

One then gets, by the independence and uniformity of the Y_ns,

$$\begin{aligned} P'(X_{n+1}^x &= y_k | X_n^x = x_n, \dots, X_0^x = x_0) = \\ &= P'(\alpha_{k-1}(X_n^x) < Y_n \leq \alpha_k(X_n^x) | X_n^x = x_n, \dots, X_0^x = x_0) \\ &= P'(\alpha_{k-1}(x_n) < Y_n \leq \alpha_k(x_n)) \\ &= Q(x_n, y_k) \,, \end{aligned} \tag{4.1.17}$$

which shows that $(X_n^x)_{n\geq 0}$ is a Markov chain with transition probability matrix Q and initial condition x. One can obtain a chain with initial distribution μ by taking convex combinations. Write the process constructed above $(X_n)_{n\geq 0}$, and denote its law by P'_x, in order to have $P'_x(X_0 = x) = 1$. Now define

$$P'_\mu := \sum_{x\in S} \mu(x) P'_x \,.$$

Then, using Lemma 4.1.1,

$$\begin{aligned} P'_\mu(X_0 = x_0, \dots, X_n = x_n) &= \sum_{x \in S} \mu(x) P'_x(X_0 = x_0, \dots, X_n = x_n) \\ &= \sum_{x \in S} \mu(x) 1_{\{x = x_0\}} Q(x_0, x_1) \dots Q(x_{n-1}, x_n) \\ &= \mu(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n)\,, \end{aligned}$$

which is (4.1.16). □

As will become clearer in the sequel, the study of homogeneous Markov chains is greatly facilitated by the introduction of a certain time translation operator on the process and of its random version, which will lead to the proofs of all recurrence results of Section 4.3. In the present section we construct a canonical space on which this operator will be naturally defined.

Each realization $\omega \in \Omega$ yields a sequence $X_1(\omega), X_2(\omega), \dots$, which we call a trajectory of the chain. A natural candidate for the simplest probability space describing an S-valued Markov Chain $(X_n)_{n\geq 0}$ is therefore the space in which each element ω *is itself* a trajectory, that is, the elements of which are the sequences $\omega = (\omega_0, \omega_1, \dots)$ where each $\omega_k \in S$:

$$\Omega := S^{\{0,1,2,\dots\}}\,. \tag{4.1.18}$$

For each $k \geq 0$, consider the coordinate map $X_k : \Omega \to \mathbb{R}$ defined by $X_k(\omega) := \omega_k$. The σ-algebra $\mathcal{F}$ is defined as the smallest collection of subsets of Ω for which each X_k is measurable, that is $\mathcal{F} := \sigma(X_k,\, k \geq 0)$. The σ-algebra $\mathcal{F}$ can also be obtained by considering the σ-algebra generated by thin cylinders, i.e. subsets of Ω of the form

$$[x_0, x_1, \dots, x_n] = \{\omega \in \Omega : \omega_0 = x_0, \omega_1 = x_1, \dots, \omega_n = x_n\}\,, \tag{4.1.19}$$

where $x_0, \dots, x_n \in S$. The intersection of two thin cylinders is either empty or is again a thin cylinder. The algebra of cylinders is obtained by taking finite unions of thin cylinders, and is denoted $\mathcal{C}$. Then clearly, $\mathcal{F} = \sigma(\mathcal{C})$.

Theorem 4.1.2. *Let μ be a probability distribution on $(S, \mathcal{P}(S))$ and Q be a transition probability matrix. Then there exists a unique probability measure P_μ on $(\Omega, \mathcal{F})$ such that on $(\Omega, \mathcal{F}, P_\mu)$, the coordinate maps $(X_n)_{n\geq 1}$ form a Markov Chain with state space S, transition* *probability matrix Q, and initial distribution μ:*

$$P_\mu(X_0 = x_0, \dots, X_n = x_n) = \mu(x_0)Q(x_0, x_1) \dots Q(x_{n-1}, x_n) . \quad (4.1.20)$$

PROOF. Consider the probability space $(\Omega', \mathcal{F}', P'_\mu)$ constructed in Theorem 4.1.1, together with the process constructed therein, which we temporarily denote by $(X'_n)_{n\geq 0}$ in order to distinguish it from the coordinate maps on Ω. Consider the map $\varphi : \Omega' \to \Omega$ defined by $\varphi(\omega')_n := X'_n(\omega')$ for all $n \geq 0$.

LEMMA 4.1.3. *φ is measurable: $\varphi^{-1}(A) \in \mathcal{F}'$ for all $A \in \mathcal{F}$.*

PROOF. Let $\mathcal{A} := \{A \in \mathcal{F} : \varphi^{-1}(A) \in \mathcal{F}'\}$. Then $\mathcal{A}$ is a σ-algebra. Moreover, it contains all sets of the form $X_n^{-1}(\{x\})$, $x \in S$, $n \geq 0$, since $\varphi^{-1}(X_n^{-1}(\{x\})) = (X_n \circ \varphi)^{-1}(\{x\}) = {X'}_n^{-1}(\{x\}) \in \mathcal{F}'$ by definition (the ${X'}_n$s are random variables). Therefore, $\mathcal{A} \equiv \mathcal{F}$. □

Since φ is measurable, we can define the image measure $P_\mu := P'_\mu \circ \varphi^{-1}$. We have

$$\begin{aligned} P_\mu(X_0 = x_0, \dots, X_n = x_n) &= P'_\mu(X'_0 = x_0, \dots, X'_n = x_n) \\ &= \mu(x_0)Q(x_0, x_1) \dots Q(x_{n-1}, x_n) , \end{aligned}$$

which shows that $(X_n)_{n\geq 0}$ has the wanted properties. Regarding uniqueness, assume $\tilde{P}_\mu$ is another measure also satisfying (4.1.20). But (4.1.20) implies that P_μ and $\tilde{P}_\mu$ coincide on thin cylinders, and since these generate $\mathcal{F}$, they are equal. □

ALTERNATE PROOF OF THEOREM 4.1.2: Let $[x_0, x_1, \dots, x_n]$ be any thin cylinder and define

$$P([x_0, x_1, \dots, x_n]) := \mu(x_0)Q(x_0, x_1) \dots Q(x_{n-1}, x_n) . \quad (4.1.21)$$

We need to show that P extends uniquely to a probability on $(\Omega, \mathcal{F})$ and that under P. But this follows immediately from Kolmogorov's Extension Theorem 3.1.2. By (4.1.21), the coordinate maps $(X_n)_{n\geq 0}$ clearly define a Markov chain with transition probability matrix Q and initial distribution μ. □

The following proposition shows that the canonical representation is sufficient for the study of Markov chains, in the sense that one cannot distinguish the distribution of the canonical chain from any other.

PROPOSITION 4.1.1. *Let $(Y_n)_{n\geq 0}$ be a Markov chain with initial distribution μ and transition matrix Q, constructed on some probability* *space* $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$. *Let* $(X_n)_{n\geq 0}$ *be the Canonical Markov chain with initial distribution* μ *and transition matrix* Q, *constructed on the product space* $(\Omega, \mathcal{F}, P)$ *as above. Then* P *is the image of* $\widetilde{P}$ *under the measurable map* $\varphi : \widetilde{\Omega} \to \Omega$ *defined by* $\varphi(\tilde{\omega}) := (Y_n(\tilde{\omega}))_{n\geq 0}$. *That is,* $P = \widetilde{P} \circ \varphi^{-1}$.

Proof. We already saw in Lemma 4.1.3 that φ is measurable. Now for any thin cylinder $[x_0, \dots, x_n]$,

$$\begin{aligned} \widetilde{P} \circ \varphi^{-1}([x_0, \dots, x_n]) &= \widetilde{P}(Y_0 = x_0, \dots, Y_n = x_n) \\ &= \mu(x_0)Q(x_0, x_1) \cdots Q(x_{n-1}, x_n) \\ &\equiv P(X_0 = x_0, \dots, X_n = x_n) \,. \end{aligned} \tag{4.1.22}$$

Since thin cylinders generate $\mathcal{F}$, this proves the proposition. □

4.2. The Markov Property

The basic relation defining a Markov chain, (4.1.1), says that conditionnally on a given past up to time n, $X_0, X_1, \dots, X_n$, the distribution of X_{n+1} depends only on X_n. Since this and time homogeneity suggest a certain translation in time, the canonical space constructed in the previous section appears well adapted to the precise formulation of a more general version of this property: conditionnally on a given past up to time n, $X_0, X_1, \dots, X_n$, the distribution of the *entire future* $X_{n+1}, X_{n+2}, \dots$ depends only on X_n. So from now on, the Markov chain under consideration will always be considered as built on the canonical product space Ω defined in (4.1.18). Define the transformation $\theta : \Omega \to \Omega$, called the shift, by

$$\theta(\omega)_n := \omega_{n+1} \quad \forall n \geq 0 \,.$$

Since $\theta^{-1}(X_n^{-1}(\{x\})) = X_{n+1}^{-1}(\{x\}) \in \mathcal{F}$, θ is measurable. One can of course iterate the shift: $\theta_1 := \theta$, and $\theta_{n+1} := \theta_n \circ \theta$.

To use the language of conditional expectation, we encode the information contained in the past of n, $X_0, X_1, \dots, X_n$, in the σ-algebra $\mathcal{F}_n := \sigma(X_0, X_1, \dots, X_n)$.

Theorem 4.2.1 (Simple Markov Property). *Let* $x \in S$, *and* $n \geq 1$. *Let* $\varphi : \Omega \to \mathbb{R}$ *be bounded, positive and* $\mathcal{F}_n$*-measurable. Then for all bounded, positive, measurable* $\psi : \Omega \to \mathbb{R}$,

$$E_x[\varphi \cdot \psi \circ \theta_n] = E_x[\varphi \cdot E_{X_n}(\psi)] \,. \tag{4.2.1}$$

In the right-hand side of (4.2.1) appears the random variable $E_{X_n}(\psi)$, which is just $E_x(\psi)$ evaluated at X_n [1]. Observe that by taking $\varphi = 1_A$ for each $A \in \mathcal{F}_n$, (4.2.1) is equivalent to the P_x-almost sure statement

$$E_x[\psi \circ \theta_n | \mathcal{F}_n] = E_{X_n}[\psi] . \tag{4.2.2}$$

Proof. We first consider the case where φ and ψ are indicators of thin cylinders: $\varphi = 1_C$, with $C = [x_0, \dots, x_n]$, $\psi = 1_D$ with $D = [y_0, \dots, y_p]$. We have

$$\begin{aligned} E_{X_n}[\psi] &= \sum_{x'_0,\dots,x'_p} \psi(x'_0, \dots, x'_p) P_{X_n}(X_0 = x'_0, \dots, X_p = x'_p) \\ &= 1_{\{X_n = y_0\}} Q(y_0, y_1) \dots Q(y_{p-1}, y_p) , \end{aligned}$$

which leads to

$$E_x[\varphi \cdot E_{X_n}(\psi)] = 1_{\{x_0 = x\}} Q(x_0, x_1) \dots Q(x_{n-1}, x_n) 1_{\{x_n = y_0\}} Q(y_0, y_1) \dots Q(y_{p-1}, y_p) .$$

On the other hand,

$$\begin{aligned} E_x[\varphi \cdot \psi \circ \theta_n] &= E_x[1_{\{X_0 = x_0\}} \dots 1_{\{X_n = x_n\}} 1_{\{X_n = y_0\}} 1_{\{X_{n+1} = y_1\}} \dots 1_{\{X_{n+p} = y_p\}}] \\ &= P_x(X_0 = 0, \dots X_n = x_n, X_n = y_0, X_{n+1} = y_1, \dots, X_{n+p} = y_p) \\ &= 1_{\{x_0 = x\}} Q(x_0, x_1) \dots Q(x_{n-1}, x_n) 1_{\{x_n = y_0\}} Q(y_0, y_1) \dots Q(y_{p-1}, y_p) , \end{aligned}$$

which shows (4.2.1) in the particular case. We then show that for the same φ, (4.2.1) holds also in the case where $\psi = 1_A$, where $A \in \mathcal{F}$. Consider the class $\mathcal{A} = \{A \in \mathcal{F} : E_x[\varphi \cdot 1_A \circ \theta_n] = E_x[\varphi \cdot E_{X_n}(1_A)]\}$. We know that $\mathcal{A}$ contains all thin cylinders, and therefore all cylinders by summation. It is then easy to verify that $\mathcal{A}$ is a Dynkin system, and so $\mathcal{A} = \mathcal{F}$ by Theorem B.0.1. Now, the extension to arbitrary bounded positive functions follows by uniform approximation by simple functions. □

A simple application of the Markov Property is the following identity, known as the Chapman-Kolmogorov Equation:

$$P_x(X_{m+n} = y) = \sum_{z \in S} P_x(X_m = z) P_z(X_n = y) . \tag{4.2.3}$$

[1] Observe here that $x \mapsto E_x(\psi)$ is $\mathcal{P}(S)$-measurable, and that $\omega \mapsto E_{X_n(\omega)}(\psi)$ is a random variable.

Namely, one can write $P_x(X_{m+n} = y) = E_x[E_x(1_{X_{m+n}=y}|\mathcal{F}_m)]$, and then

$$\begin{aligned} E_x[E_x(1_{X_{m+n}=y}|\mathcal{F}_m)] &= E_x[E_x(1_{X_n=y} \circ \theta_m|\mathcal{F}_m)] \\ &= E_x[E_{X_m}(1_{X_n} = y)] \\ &= E_x[P_{X_m}(X_n = y)] \\ &= \sum_{z \in S} P_x(X_m = z) P_z(X_n = y) \,. \end{aligned} \tag{4.2.4}$$

The reader should convince himself that any other proof of (4.2.3) will necessary end up requiring one or another form of the Simple Markov Property.

The Markov Property deserves an extension to the case there the time n is replaced by a random time T. The reason is the following. Suppose we are interested in the following question: if the chain returns back to its starting point with probability one, is it true that it will do so an infinite number of times? This seems clear since at the time of first return, the chain is back at its original position and therefore by the Markov Property the probability of coming back a second time is again one, and so on. Nevertheless, the times at which the chain returns to its starting point are *random*, and the simple Markov Property can't be used in its actual form.

Random times are usually called *stopping times*. We will define them here in the framework of Markov chains; in Section 5 these will be used extensively in the chapter on martingales. A stopping time satisfies a list of properties which we first illustrate on a simple example. Let $(X_n)_{n\geq 0}$ be the random walk on the integers with initial condition $X_0 = 0$. Considering n as a parameter describing time, an example of a random time is the first return of the walk to the origin, which we already encountered in Section 2:

$$T_0 := \inf\{n \geq 1 : X_n = 0\} \,. \tag{4.2.5}$$

If the walk never returns to the origin, i.e. $\{n \geq 1 : X_n = 0\} = \emptyset$, we set $T_0 = \infty$. So T_0 is a random variable taking values in $\{1, 2, \dots\} \cup \{\infty\}$. Moreover, the event $\{T_0 = n\}$ is insensitive to the change of any of the variables X_k for $k > n$. This is made clear by noting that $\{T_0 = n\} = \{X_1 \neq 0, \dots, X_{n-1} \neq 0, X_n = 0\}$. In other words, $\{\tau_0 = n\}$ is $\mathcal{F}_n$-measurable, where $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$. We call the sequence $(\mathcal{F}_n)_{n\geq 0}$ the **natural filtration** associated to the chain $(X_n)_{n\geq 0}$.

Clearly, $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for all $n \geq 0$. The natural filtration can be defined for any random process.

DEFINITION 4.2.1. *Consider the natural filtration $(\mathcal{F}_n)_{n\geq 0}$ associated to a Markov chain $(X_n)_{n\geq 0}$. A* **stopping time** *is a $\{1, 2, \dots\} \cup \{\infty\}$-valued random variable T such that for all $n \geq 0$, $\{T = n\} \in \mathcal{F}_n$.*

In the Simple Markov Property, we considered a Markov chain at a fixed time n, conditionned with respect to $\mathcal{F}_n$. We now want to consider the same chain at a random time T, and condition with respect to the σ-algebra which contains events that depend only on what happened *before* T. Since it doesn't make sense to write "$\sigma(X_0, X_1, \dots, X_T)$", we say that $A \in \mathcal{F}_T$ if each time $T \leq n$ then $A \in \mathcal{F}_n$. So define the **stopped σ-algebra generated by T**:

$$\mathcal{F}_T := \{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n \, \forall n \geq 0\} . \tag{4.2.6}$$

Fot example, $\{T < \infty\} \in \mathcal{F}_T$. It can be easily verified that $\mathcal{F}_T$ is a σ-algebra (see Exercise 4.12). The position of a Markov chain at time T is naturally defined by the random variable

$$X_T(\omega) := \begin{cases} X_n(\omega) & \text{if } T(\omega) = n \,, \\ \text{“0”} & \text{if } T(\omega) = \infty \,, \end{cases}$$

where "0" is any fixed point of S. We also define $\theta_\infty := \mathrm{id}$. We are now ready to move on to the study of the Markov Property when the conditionning is done with respect to the past of a random stopping time.

THEOREM 4.2.2 (Strong Markov Property). *Let $x \in S$. Let T be a stopping time adapted to the natural filtration $(\mathcal{F}_n)_{n\geq 0}$. Let $\varphi : \Omega \to \mathbb{R}$ be bounded, positive and $\mathcal{F}_T$-measurable. Then for all bounded, positive, measurable $\psi : \Omega \to \mathbb{R}$,*

$$E_x[1_{\{T<\infty\}} \cdot \varphi \cdot \psi \circ \theta_T] = E_x[1_{\{T<\infty\}} \cdot \varphi \cdot E_{X_T}(\psi)] . \tag{4.2.7}$$

In particular, if $P_x(T < \infty) = 1$, then

$$E_x[\varphi \cdot \psi \circ \theta_T] = E_x[\varphi \cdot E_{X_T}(\psi)] . \tag{4.2.8}$$

The analogue of (4.2.2) for random times reads

$$E_x[1_{\{T<\infty\}} \cdot \psi \circ \theta_T | \mathcal{F}_T] = 1_{\{T<\infty\}} E_{X_T}[\psi] \quad P_x\text{-a.s.} \tag{4.2.9}$$

PROOF. The proof follows by writing $\{T < \infty\} = \bigcup_{n\geq 0}\{T = n\}$. Then, since $1_{\{T=n\}} \cdot \varphi$ is $\mathcal{F}_n$-measurable, by the Simple Markov Property,

$$\begin{aligned} E_x[1_{\{T=n\}} \cdot \varphi \cdot \psi \circ \theta_T] &= E_x[1_{\{T=n\}} \cdot \varphi \cdot \psi \circ \theta_n] \\ &= E_x[1_{\{T=n\}} \cdot \varphi \cdot E_{X_n}(\psi)] \\ &= E_x[1_{\{T=n\}} \cdot \varphi \cdot E_{X_T}(\psi)]\,. \end{aligned}$$

Summing over n gives (4.2.7). □

4.3. Recurrence and Classification

We now consider the *recurrence* problem mentionned before in the case of the random walk: when does a Markov chain come back to its starting point? As before, we will always consider the canonical chain constructed on the product space $\Omega = S^{\{0,1,2,\dots\}}$.

Two random variables are relevant in the study of recurrence. For each $x \in S$, the first visit at x is defined by

$$T_x := \inf\{n \geq 1 : X_n = x\}\,, \tag{4.3.1}$$

where we set $T_x := \infty$ if $\{n \geq 1 : X_n = x\} = \emptyset$. Observe that T_x is a stopping time since $\{T_x > n\} = \{X_1 \neq x, \dots, X_n \neq x\} \in \mathcal{F}_n$. On the other hand, the number of visits at site x is defined by

$$N_x := \sum_{n\geq 1} 1_{\{X_n = x\}}\,. \tag{4.3.2}$$

Clearly, $N_x \geq 1$ if and only if $T_x < \infty$, and so $P_x(N_x \geq 1) = P_x(T_x < \infty)$. A cornerstone in the study of recurrence for Markov chains is a generalization to the situation where $N_x \geq k$.

LEMMA 4.3.1. *Let $x, y \in S$, $k \geq 1$. Then*

$$P_x(N_y \geq k) = P_x(T_y < \infty)P_y(N_y \geq k-1)\,. \tag{4.3.3}$$

In particular, $P_x(N_x \geq k) = P_x(T_x < \infty)^{k-1}$.

PROOF. Observe that $N_y = N_y \circ \theta_{T_y} + 1$. Therefore, $N_y \geq k+1$ if and only if $T_y < \infty$ and $N_y \circ \theta_{T_y} \geq k$:

$$\begin{aligned} P_x(N_y \geq k+1) &= P_x(T_y < \infty, N_y \circ \theta_{T_y} \geq k) \\ &= E_x[1_{\{T_y<\infty\}} \cdot 1_{\{N_y \geq k\}} \circ \theta_{T_y}]\,. \end{aligned}$$

By the Strong Markov Property (with $\varphi = 1$) and since $X_{T_y} = y$,

$$\begin{aligned} E_x[1_{\{T_y<\infty\}} \cdot 1_{\{N_y \geq k\}} \circ \theta_{T_y}] &= E_x[1_{\{T_y<\infty\}} \cdot E_{X_{T_y}}[1_{\{N_y\geq k\}}]] \\ &= E_x[1_{\{T_y<\infty\}}]E_y[1_{\{N_y\geq k\}}] \\ &\equiv P_x(T_y < \infty)P_y(N_y \geq k)\,. \end{aligned} \tag{4.3.4}$$

The second affirmation follows from the first by induction. □

As an immediate corollary, we obtain the following formula:

$$E_x[N_x] = \sum_{k\geq 1} E_x[N_x \geq k] = \frac{1}{1 - P_x(T_x < \infty)} = \frac{1}{P_x(T_x = \infty)}\,. \tag{4.3.5}$$

This formula makes sense also when $P_x(T_x = \infty) = 0$, in which case $E_x[N_x] = \infty$.

Definition 4.3.1. *A point $x \in S$ is called*

- **recurrent** *if $P_x(T_x < \infty) = 1$,*
- **transient** *if $P_x(T_x < \infty) < 1$.*

Proposition 4.3.1. *Let $x \in S$. Then*

(1) *x is recurrent if and only if $P_x(N_x = \infty) = 1$,*
(2) *x is transient if and only if $P_x(N_x = \infty) = 0$.*

Proof. Assume x is recurrent. Then since $\{N_x \geq k\} \searrow \{N_x = +\infty\}$, and since $P_x(N_x \geq k) = 1$ by the previous lemma, we have $P_x(N_x = \infty) = 1$. Conversely, if $P_x(N_x = \infty) = 1$ then $P_x(N_x \geq k) = 1$ for all $k \geq 1$, which implies $P_x(T_x < \infty) = 1$ by the previous lemma: x is recurrent [2]. If x is transient then $P_x(T_x < \infty) < 1$, and by Lemma 4.3.1, $P_x(N_x = \infty) = \lim_{k\to\infty} P_x(N_x \geq k) = 0$. □

Observe that $\{N_x = \infty\}$ is a tail event, and we have proved that with respect to P_x, its probability is either 0 (when x is transient) or 1 (when x is recurrent). Nevertheless, we have not yet proved a 0-1 Law for Markov chains.

The goal of the rest of this section is to study the partition of S into recurrent and transient states. Comparison of recurrence properties of different points x, y, will be done by studying the expected number of visits at y when started from x:

$$u(x, y) := E_x[N_y]\,. \tag{4.3.6}$$

[2] Observação feita pela Luciana, abril 2008.

Lemma 4.3.2. *For any $x, y \in S$,*

(1) $u(x,y) = \sum_{n\geq 0} Q^{(n)}(x,y)$.
(2) *x is recurrent if and only if $u(x,x) = \infty$.*
(3) *If $x \neq y$, then*

$$u(x,y) = P_x(T_y < \infty)u(y,y)\,. \tag{4.3.7}$$

Each of these properties is intuitive: together, (1) and (2) give a computable criterium for verifying whether a point is recurrent, which will be used repeatedly in the sequel (in particular for random walks, in Section 4.3.2). The identity (4.3.7) gives a simple way of comparing recurrence properties of different points.

Proof of Lemma 4.3.2. (1) follows by the definition of N_y and (4.1.8), (2) was shown in (4.3.5). For (3), we use the Strong Markov Property. Since $N_y = 0$ on $\{T_y = \infty\}$,

$$\begin{aligned} E_x[N_y] = E_x[N_y, T_y < \infty] &= E_x[1_{\{T_y<\infty\}} \cdot N_y \circ \theta_{T_y}] \\ &= E_x[1_{\{T_y<\infty\}} \cdot E_y[N_y]] \\ &= P_x(T_y < \infty)E_y[N_y]\,, \end{aligned}$$

which is (4.3.7). □

"Recurrence is contagious", as seen hereafter.

Lemma 4.3.3. *Let x be recurrent, and $y \neq x$. If $u(x,y) > 0$, then $P_y(T_x < \infty) = 1$, $u(y,x) > 0$, y is recurrent and $P_x(T_y < \infty) = 1$. If y is transient, then $u(x,y) = 0$.*

Proof. Since x is recurrent, $P_x(N_x = \infty) = 1$ (Proposition 4.3.1), and so

$$\begin{aligned} 0 = P_x(N_x < \infty) &\geq P_x(T_y < \infty, T_x \circ \theta_{T_y} = \infty) \\ &= E_x[1_{\{T_y<\infty\}} \cdot 1_{\{T_x=\infty\}} \circ \theta_{T_y}] \\ &= E_x[1_{\{T_y<\infty\}} \cdot E_y[1_{\{T_x=\infty\}}]] \\ &= P_x(T_y < \infty)P_y(T_x = \infty)\,. \end{aligned} \tag{4.3.8}$$

Since $u(x,y) > 0$, there exists $n \geq 1$ such that $Q^{(n)}(x,y) > 0$, which implies $P_x(T_y < \infty) \geq Q^{(n)}(x,y) > 0$. (4.3.8) thus gives $P_y(T_x = \infty) = 0$, i.e. $P_y(T_x < \infty) = 1$. By (4.3.7), we obtain $u(y,x) = P_y(T_x < \infty)u(x,x) = \infty > 0$. Since $u(y,x) > 0$, there exists $m \geq 1$ such that

$Q^{(m)}(y,x) > 0$. Then, for all $p \geq 0$, by the Chapman-Kolmogorov Equation,

$$Q^{(n+m+p)}(y,y) \geq Q^{(m)}(y,x)Q^{(p)}(x,x)Q^{(n)}(x,y)\,,$$

and so

$$u(y,y) \geq \sum_{p\geq 0} Q^{(n+m+p)}(y,y) \geq Q^{(m)}(y,x)\Big[\sum_{p\geq 0} Q^{(p)}(x,x)\Big]Q^{(n)}(x,y) = \infty\,,$$

which implies that y is recurrent. Proceeding as above from y to x gives $P_x(T_y < \infty) = 1$. The last claim is then obvious. □

As an application, consider the simple random walk on $\mathbb{Z}$ with $0 < p < 1$. Let $x, y \in S$, $x < y$. Then $u(x,y) \geq Q^{(y-x)}(x,y) \geq p^{y-x} > 0$. Similarly, $u(y,x) \geq q^{y-x} > 0$. Therefore, all points are either recurrent, or transient. It is thus sufficient to consider the recurrence properties of the origin. By Theorem 2.1.1, we have that all points are recurrent if $p = \frac{1}{2}$, transient otherwise.

Going back to the general case, let us write S as a disjoint union $\mathcal{R}\cup\mathcal{T}$, where $\mathcal{R}$ are the recurrent points and $\mathcal{T}$ are the transient points. Define the following relation on $\mathcal{R}$: $x \sim y$ if and only if $u(x,y) > 0$. Then obviously $x \sim x$, and Lemma 4.3.3 shows that $\sim$ is reflexive: $x \sim y$ implies $y \sim x$. On the other hand, Then, if $x \sim y$ then there exists $n \geq 1$ with $Q^{(n)}(x,y) > 0$, if $y \sim z$ then there exists $m \geq 1$ with $Q^{(n)}(y,z) > 0$, and so $Q^{(n+m)}(x,z) > 0$, i.e. $x \sim z$. That is, $\sim$ is an equivalence relation, and we can consider the partition of $\mathcal{R}$ into equivalence classes. Since $\mathcal{R}$ is countable, this partition also is, and we denote it by $\mathcal{R} = \bigcup_{j\geq 1}\mathcal{R}_j$. Each $\mathcal{R}_j$ is called a recurrence class.

The Classification Theorem hereafter proves the following intuitive properties: the chain started at $x \in \mathcal{R}_j$ stays in $\mathcal{R}_j$ forever and visits any other $y \in \mathcal{R}_j$ an infinite number of times. The chain started at $x \in \mathcal{T}$ either never visits $\mathcal{R}$ and visits any transient point a finite number of times, or eventually enters a recurrence class $\mathcal{R}_j$ and stays there forever.

THEOREM 4.3.1. *The decomposition $S = \mathcal{T}\cup\bigcup_{j\geq 1}\mathcal{R}_j$ has the following properties:*

(1) *If $x \in \mathcal{R}_j$ then, P_x-almost surely, $N_y = \infty$ for all $y \in \mathcal{R}_j$ and $N_y = 0$ for all $y \in S\backslash\mathcal{R}_j$.*

(2) *If* $x \in \mathcal{T}$ *and* $T_{\mathcal{R}} := \inf\{n \geq 1 : X_n \in \mathcal{R}\}$ *then,* P_x*-almost surely,*
   (a) *either* $T_{\mathcal{R}} = \infty$ *and then* $N_y < \infty$ *for all* $y \in S$*,*
   (b) *or* $T_{\mathcal{R}} < \infty$ *and there exists a random* $j \geq 1$ *such that* $X_n \in \mathcal{R}_j$ *for all* $n \geq T_{\mathcal{R}}$*.*

PROOF. (1) Let $x \in \mathcal{R}_j$. Then $E_x(N_y) = u(x,y) = 0$ for all $y \in \mathcal{T}$ by Lemma 4.3.3, and for all $y \in \mathcal{R}_i$ ($i \neq j$) by definition. Therefore, $N_y = 0$ P_x-a.s. for all $y \in S \backslash \mathcal{R}_j$. If $y \in \mathcal{R}_j$, then by taking $k \to \infty$ in Lemma 4.3.1, we get

$$P_x(N_y = \infty) = P_x(T_y < \infty) P_y(N_y = \infty)\,. \tag{4.3.9}$$

But $P_x(T_y < \infty) = 1$ by Lemma 4.3.3, and $P_y(N_y = \infty) = 1$ by Proposition 4.3.1. Therefore, $N_y = \infty$ P_x-a.s.
(2) Let $x \in \mathcal{T}$. We first show (2a), which means

$$P_x(T_{\mathcal{R}} = \infty) = P_x(T_{\mathcal{R}} = \infty, N_y < \infty\, \forall y \in \mathcal{T})\,. \tag{4.3.10}$$

Since

$$P_x(T_{\mathcal{R}} = \infty, N_y < \infty\, \forall y \in \mathcal{T}) = P_x(T_{\mathcal{R}} = \infty) - P_x\Big(\{T_{\mathcal{R}} = \infty\} \cap \bigcup_{y \in \mathcal{T}} \{N_y = \infty\}\Big)\,,$$

it suffices to notice that for each $y \in \mathcal{T}$,

$$P_x(T_{\mathcal{R}} = \infty, N_y = \infty) \leq P_x(N_y = \infty)\,,$$

which is zero since y is transient (use (4.3.9) and Proposition 4.3.1). This proves (4.3.10). Then we show (2b), which means

$$P_x(T_{\mathcal{R}} < \infty) = P_x(T_{\mathcal{R}} < \infty, \exists j \geq 1 \text{ s.t. } X_n \in \mathcal{R}_j \forall n \geq T_{\mathcal{R}})\,. \tag{4.3.11}$$

Since the recurrence classes $\mathcal{R}_j$ are disjoint, we can compute

$$\begin{aligned} P_x(T_{\mathcal{R}} < \infty, X_n \in \mathcal{R}_j \forall n \geq T_{\mathcal{R}}) &= E_x[1_{\{T_{\mathcal{R}} < \infty\}} \cdot 1_{\{X_n \in \mathcal{R}_j \forall n \geq 0\}} \circ \theta_{T_{\mathcal{R}}}] \\ &= E_x[1_{\{T_{\mathcal{R}} < \infty\}} \cdot P_{X_{T_{\mathcal{R}}}}(X_n \in \mathcal{R}_j \forall n \geq 0)] \end{aligned}$$

But clearly, $P_{X_{T_{\mathcal{R}}}}(X_n \in \mathcal{R}_j \forall n \geq 0) = 1$ if $X_{T_{\mathcal{R}}} \in \mathcal{R}_j$, 0 if $X_{T_{\mathcal{R}}} \notin \mathcal{R}_j$. Therefore, the right hand side of (4.3.11) equals

$$\begin{aligned} \sum_{j \geq 1} E_x[1_{\{T_{\mathcal{R}} < \infty\}} \cdot P_{X_{T_{\mathcal{R}}}}(X_n \in \mathcal{R}_j \forall n \geq 0)] &= E_x\Big[1_{\{T_{\mathcal{R}} < \infty\}} \sum_{j \geq 1} 1_{\{X_{T_{\mathcal{R}}} \in \mathcal{R}_j\}}\Big] \\ &\equiv E_x[1_{\{T_{\mathcal{R}} < \infty\}}] \\ &= P_x(T_{\mathcal{R}} < \infty)\,. \end{aligned}$$

We have used the fact that $\sum_{j \geq 1} 1_{\{X_{T_{\mathcal{R}}} \in \mathcal{R}_j\}} = 1_{\{X_{T_{\mathcal{R}}} \in \mathcal{R}\}} = 1$ on $\{T_{\mathcal{R}} < \infty\}$. This finishes the proof of the theorem. $\square$

4.3.1. Irreducibility. The Classification Theorem shows that the long time evolution of a Markov chain depends on how the state space S splits into equivalence classes, via the use of the function u. It is natural to consider the case in which the chain has a single class.

DEFINITION 4.3.2. *A chain is called* irreducible *if* $u(x, y) > 0$ *for all* $x, y \in S$.

An equivalent definition of irreducibility is: for all $x, y \in S$, there exists an $n \geq 1$ such that $Q^{(n)}(x, y) > 0$. As seen hereafter, in an irreducible chain, all the points are of the same type.

THEOREM 4.3.2. *Let the chain be irreducible. Then*

(1) *either all the points are recurrent, there exists a single recurrence class* $S \equiv \mathcal{R}_1$, *and* $P_x(N_y = \infty \, \forall y \in S) = 1$ *for all* $x \in S$,
(2) *or all states are transient,* $S = \mathcal{T}$, *and* $P_x(N_y < \infty \, \forall y \in S) = 1$ *for all* $x \in S$.

When S is finite, only the first case can happen.

PROOF. (1) If there is a recurrent point, then by the irreducibility hypothesis and Lemma 4.3.3, all points are recurrent, and clearly there can exist only one recurrence class. The statement, as well as (2), follow from Theorem 4.3.1. For the last statement, assume $|S| < \infty$. If some $x \in S$ were transient, then by (2), we would have, P_x-a.s., $N_y < \infty$ for all $y \in S$. In particular, $\sum_{y \in S} N_y < \infty$. But this is absurd since

$$\sum_{y \in S} N_y = \sum_{y \in S} \sum_{n \geq 0} 1_{\{X_n = y\}} = \sum_{n \geq 0} \sum_{y \in S} 1_{\{X_n = y\}} = \infty \,.$$

(Indeed, for each $n \geq 0$, $\sum_{y \in S} 1_{\{X_n = y\}} = 1$.) □

Before going further and introduce invariant measures, we apply these results to the study of recurrence of random walks on $\mathbb{Z}^d$.

4.3.2. The Simple Symmetric Random Walk on $\mathbb{Z}^d$. The simple random walk on $\mathbb{Z}^d$ was introduced in Example 4.1.3: $S_n = \sum_{k=1}^n X_k$, where $S_0 = 0$ and the variables X_k are $\mathbb{Z}^d$-valued, i.i.d., with distribution p defined in (4.1.2). We denote the probability describing the walk by P (rather than P_0). Clearly, the chain is irreducible. By Theorem 4.3.2, the points are either all recurrent, or all transient. It is thus enough to consider the origin, whose time of first return is denoted T_0. The random walk is recurrent if $P(T_0 < \infty) = 1$, and transient otherwise (i.e. if $P(T_0 < \infty) < 1$). The main result for the simple random walk is the following.

THEOREM 4.3.3. *The simple symmetric random walk is recurrent for $d = 1, 2$, and transient for $d \geq 3$.*

Since $Q^{(n)}(0,0) = P(S_n = 0)$, which is zero when n is odd, Lemma 4.3.2 gives the following criterium for recurrence.

$$\text{The walk is recurrent } \Leftrightarrow \sum_{n \geq 1} P(S_{2n} = 0) = \infty\,. \tag{4.3.12}$$

Recurrence for $d = 1, 2$ will be obtained with the following property of symmetric random walks.

LEMMA 4.3.4. *If the walk is symmetric, then*

$$P(S_{2n} = 0) = \sup_{z \in \mathbb{Z}^d} P(S_{2n} = z)\,. \tag{4.3.13}$$

PROOF. We sum over the position at the nth step and use independence:

$$\begin{aligned} P(S_{2n} = z) &= \sum_{y \in \mathbb{Z}^d} P(S_n = y, S_{2n} = z) \\ &= \sum_{y \in \mathbb{Z}^d} P(S_n = y, S_{2n} - S_n = z - y) \\ &= \sum_{y \in \mathbb{Z}^d} P(S_n = y) P(S_n = z - y)\,. \end{aligned} \tag{4.3.14}$$

By the Cauchy-Schwartz Inequality and a change of variable,

$$\begin{aligned} \sum_{y \in \mathbb{Z}^d} P(S_n = y) P(S_n = z - y) &\leq \Big[\sum_{y \in \mathbb{Z}^d} P(S_n = y)^2\Big]^{\frac{1}{2}} \Big[\sum_{y \in \mathbb{Z}^d} P(S_n = z - y)^2\Big]^{\frac{1}{2}} \\ &= \sum_{y \in \mathbb{Z}^d} P(S_n = y)^2\,. \end{aligned}$$

Now if the walk is symmetric then $P(S_n = y) = P(S_n = -y)$, and so using again (4.3.14) with $z = 0$, we get

$$\sum_{y \in \mathbb{Z}^d} P(S_n = y)^2 = \sum_{y \in \mathbb{Z}^d} P(S_n = y) P(S_n = -y) = P(S_{2n} = 0)\,,$$

which proves the claim. □

Below, $\|\cdot\|$ denotes Euclidian distance in $\mathbb{Z}^d$.

Proof of Theorem 4.3.3: First consider $d = 1$: by Lemma 4.3.4,

$$1 = \sum_{y\in\mathbb{Z}:\|y\|\le 2n} P(S_{2n} = y) \le (4n+1)P(S_{2n} = 0)\,,$$

which gives $P(S_{2n} = 0) \ge (4n+1)^{-1}$. By (4.3.12), the walk is recurrent. For $d = 2$, we proceed in the same way. A straightforward computation using independence of the $X_k s$ yields $E[\|S_{2n}\|^2] = 2n$. By the Chebychev Inequality,

$$P(\|S_{2n}\| > 2\sqrt{n}) \le \frac{E[\|S_{2n}\|^2]}{4n} = \frac{1}{2}\,.$$

One can thus proceed as before and obtain

$$\frac{1}{2} \le P(\|S_{2n}\| \le 2\sqrt{n}) = \sum_{y\in\mathbb{Z}^2:\|y\|\le 2\sqrt{n}} P(S_{2n} = y) \le (8\sqrt{n}+1)^2 P(S_{2n} = 0)\,.$$

By (4.3.12), the walk is recurrent. For $d = 3$, we need an upper bound. Let $n_i \ge 0$, $i \in \{1, 2, 3\}$, be the number of positive steps done along the direction e_i. To be back at the origin after $2n$ steps, we must choose a triple (n_1, n_2, n_3) satisfying $n_1 + n_2 + n_3 = n$, and then choose a path which contains, for each $i = 1, 2, 3$, n_i steps along $+e_i$, and n_i steps along $-e_i$. There are

$$\binom{2n}{n_1\ n_1\ n_2\ n_2\ n_3\ n_3} = \frac{(2n)!}{(n_1!n_2!n_3!)^2}$$

ways of doing so. Since each path has probability $(\frac{1}{6})^{2n}$,

$$\begin{aligned} P(S_{2n} = 0) &= \sum_{\substack{(n_1,n_2,n_3):\\ n_1+n_2+n_3=n}} \frac{(2n)!}{(n_1!n_2!n_3!)^2}\frac{1}{6^{2n}} \\ &= \frac{1}{2^{2n}}\binom{2n}{n} \sum_{\substack{(n_1,n_2):\\ 0\le n_1+n_2\le n}} \left[\frac{n!}{n_1!n_2!(n-n_1-n_2)!}\frac{1}{3^n}\right]^2 \\ &\le \frac{1}{2^{2n}}\binom{2n}{n} \max_{\substack{(n_1,n_2):\\ 0\le n_1+n_2\le n}} \frac{n!}{n_1!n_2!(n-n_1-n_2)!}\frac{1}{3^n}\,. \end{aligned} \tag{4.3.15}$$

We have used the fact that the numbers in the brackets add up to one.

Lemma 4.3.5. *There exists $C > 0$ such that*

$$\max_{\substack{(n_1,n_2):\\ 0\le n_1+n_2\le n}} \frac{n!}{n_1!n_2!(n-n_1-n_2)!}\frac{1}{3^n} \le \frac{C}{n}\,. \tag{4.3.16}$$

Proof. As can be easily verified, the denominator in (4.3.16) decreases when the difference between the three numbers $n_1, n_2, n-n_1-n_2$ is reduced. One can therefore bound the maximum over all triples in which each term lies within distance at most one from $\frac{n}{3}$. This implies that for large n, the Stirling Formula can be used for each of the terms apearing in the ratio, which proves the lemma. □

Using the lemma and again the Stirling Formula for the first term in (4.3.15),

$$P(S_{2n}=0) \leq \frac{D}{n^{\frac{3}{2}}}\,.$$

With (4.3.12), we conclude that the simple random walk on $\mathbb{Z}^3$ is transient. The proof that the walk is transient in higher dimensions is left as an exercise. □

Observe that all the estimates we have obtained above for $P(S_{2n}=0)$ follow from a more general Local Limit Theorem, valid in all dimension (see Exercise 4.19):

$$P(S_{2n}=0) \sim \frac{1}{\sqrt{(2\pi n)^d}}\,.$$

4.4. Equilibrium: Stationary Distributions

Theorems (4.3.1) and 4.3.2 give a first general picture of what the asymptotic behaviour of a Markov chain looks like: starting from an arbitrary point x, it either falls into one of the recurrence classes $\mathcal{R}_j$, or remains transient forever. Our next objective is to take a closer look at what can happen in each of these cases. More precisely, we will look at things such at the average time spent by the chain at each point $x \in S$, leading to the natural notion of *invariant* measure. Before this we to introduce some notations for probability distributions on $(S, \mathcal{P}(S))$.

4.4.1. Invariant Measures. Let μ be a measure on $(S, \mathcal{P}(S))$, i.e. a collection of non-negative numbers $(\mu(x))_{x\in S}$. To avoid misleading it with E_μ, which acts on random variables living in another space, we denote the expectation, with respect to μ, of a positive bounded measurable function $f : S \to \mathbb{R}$ by either of the symbols

$$\int f d\mu = \mu(f) := \sum_{x\in S} \mu(x) f(x)\,.$$

It is sometimes useful to think of functions $f : S \to \mathbb{R}$ as *column vectors* and of measures μ on S as *row vectors*. The expectation $\mu(f)$ can then be naturally written as an inner product:

$$\langle f, \mu \rangle := \sum_{x \in S} \mu(x) f(x) \,.$$

If Q is a transition probability matrix, we define a new measure μQ by

$$\mu Q(x) := \sum_{y \in S} \mu(y) Q(y, x) \,. \tag{4.4.1}$$

Remembering (4.1.13):

$$Q^{(n)} f(x) := \sum_{y \in S} Q^{(n)}(x, y) f(y) \,, \tag{4.4.2}$$

we have the following identity:

$$\langle f, \mu Q \rangle = \langle Q f, \mu \rangle \,.$$

It does then make sense to say that Q act *from the left* on functions and *from the right* on measures. If μ is a probability (i.e. $\sum_x \mu(x) = 1$), then μQ is again a probability. Going back to Markov chains: if μ is the probability distribution of X_0 for the Markov chain $(X_n)_{n \geq 0}$ whose transition matrix is Q, i.e. $P_\mu(X_0 = x) = \mu(x)$, then μQ is the distribution of X_1. Indeed, by Lemma 4.1.1,

$$P_\mu(X_1 = x) = \sum_{y \in S} P_\mu(X_1 = x, X_0 = y) = \sum_{y \in S} \mu(y) Q(y, x) \equiv \mu Q(x) \,.$$

Similarly, the distribution of X_n is given by $\mu Q^{(n)}$:

$$P_\mu(X_n = x) = \mu Q^{(n)}(x) \,.$$

We see that understanding the large-n-behaviour of the chain goes through the study of the limits

$$\pi(x) := \lim_{n \to \infty} \mu Q^{(n)}(x) \,. \tag{4.4.3}$$

Giving a meaning to (4.4.3), conditions under which this limit exists, and its possible independence of μ, will be done in details later.

There is also a formula for the expectation of $f(X_n)$ with respect to E_μ:

$$E_\mu(f(X_n)) = \sum_{x \in S} P_\mu(X_n = x) f(x) = \mu Q^{(n)}(f) \,. \tag{4.4.4}$$

(4.4.4) says that the expectation of an observable made on the evolution can be obtained by an expectation of this observable over S with respect to the measure $\mu Q^{(n)}$. Observe that $\mu Q^{(n)}(f) = \langle f, \mu Q^{(n)} \rangle = \langle Q^{(n)} f, \mu \rangle = \mu(Q^{(n)} f)$. To motivate the following definition, assume for a while that the limit defining π in (4.4.3) exists for all $x \in S$. Then for all bounded f,

$$\langle f, \pi Q \rangle = \langle Qf, \pi \rangle = \lim_{n\to\infty} \langle Qf, \mu Q^{(n)} \rangle = \lim_{n\to\infty} \langle f, \mu Q^{(n+1)} \rangle = \langle f, \pi \rangle \,,$$

which implies that $\pi Q = \pi$. This motivates the following definition.

DEFINITION 4.4.1. *Let Q be a transition matrix, μ a measure on $(S, \mathcal{P}(S))$. If*

$$\mu Q = \mu \,, \tag{4.4.5}$$

then μ is called invariant with respect to Q.

(4.4.5) is sometimes called the balance relation. Consider the random walk of Example 4.1.3, with $Q(x, y) = p(y - x)$. Then the counting measure ($\mu(x) = 1$ for all x) is invariant:

$$\mu Q(x) = \sum_{y \in S} Q(y, x) = \sum_{y \in S} p(y - x) = 1 = \mu(x) \,.$$

By induction we see that if μ is invariant, then $\mu Q^{(n)} = \mu$ for all $n \geq 1$. Moreover, when the initial distribution μ of a Markov chain $(X_n)_{n\geq 0}$ with transition matrix Q is invariant under Q, then X_n has the same distribution as X_0. Namely, by (4.4.4),

$$E_\mu(f(X_n)) = \mu Q^{(n)}(f) = \mu(f) \equiv E_\mu(f(X_0)) \,.$$

In such a case, i.e. when the distribution of the chain is insensitive to the evolution under the transition matrix Q, we say that X_n is at equilibirum for all $n \geq 1$. Invariant measures will play an important role in the study of the asymptotics of the chain.

We will first be interested in the existence of invariant measures, then of invariant probability measures, and then we shall move on to the study of the existence of the limits (4.4.3).

4.4.2. Existence of Invariant Measures. Finding an invariant measure means, for the time being, solving a system of equations for $(\mu(x))_{x\in S}$:

$$\mu(x) = \sum_{y \in S} \mu(y) Q(y, x) \quad \forall x \in S \,.$$

DEFINITION 4.4.2. *A measure μ is* **reversible (with respect to Q)** *if*

$$\mu(x)Q(x,y) = \mu(y)Q(y,x) \quad \forall x, y \in S\,. \tag{4.4.6}$$

The set relations (4.4.6) are sometimes called the **relation of detailed balance**, since it is stronger than (4.4.5). Observe that if μ is reversible, then for all $x \in S$,

$$\mu Q(x) = \sum_{y \in S} \mu(y)Q(y,x) = \sum_{y \in S} \mu(x)Q(x,y) = \mu(x)\,.$$

We have thus shown

LEMMA 4.4.1. *If μ is reversible, then it is invariant.*

This result gives an easy way of finding invariant measures. For example, consider the uniform random walk on the graph, introduced in Example 4.1.4. Then the measure $\mu(x) := |A_x|$ is invariant. Namely, if $\{x,y\} \in E$,

$$\mu(x)Q(x,y) = |A_x|\frac{1}{|A_x|} = 1 = |A_y|\frac{1}{|A_y|} = \mu(y)Q(y,x)\,.$$

Another example is the simple random walk on $\mathbb{Z}$ with $Q(x,x+1) = p < 1$. It easy to verify, using the above criterium, that the measure

$$\mu(x) = \big(\frac{p}{1-p}\big)^x\,, \quad \forall x \in \mathbb{Z}$$

is invariant. Observe that $\mu(x)$ is bounded if and only if $p = \frac{1}{2}$. When $p > \frac{1}{2}$ (resp. $p < \frac{1}{2}$), then μ gives unbounded weight to points far to the right (resp. left), which reflects the transience of the walk. As an exercise, the reader can also compute an invariant measure for the Ehrenfest Model of Example 4.1.5 (Exercise 4.24).

The following result shows that the existence of at least one recurrent point x guarantees the existence of an invariant measure.

THEOREM 4.4.1. *Let $x \in S$ be recurrent. For all $y \in S$, define*

$$\nu_x(y) := E_x\Big[\sum_{k=0}^{T_x - 1} 1_{\{X_k = y\}}\Big] \equiv E_x[N_y, T_y < T_x]\,. \tag{4.4.7}$$

Then ν_x is an invariant measure [3]. *Moreover, $\nu_x(y) > 0$ if and only if y belongs to the recurrence class of x. Finally, $\nu_x(y) < \infty$ for all $y \in S$.*

PROOF. Observe that $\nu_x(x) = E_x(1) = 1$. We compute, for all $z \in S$,

$$\begin{aligned}\sum_{y\in S} \nu_x(y)Q(y,z) &= \sum_{y\in S}\sum_{k\geq 0} E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}]Q(y,z)\\ &= \sum_{y\in S}\sum_{k\geq 0} E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}1_{\{X_{k+1}=z\}}] \qquad (4.4.8)\\ &= \sum_{k\geq 0} E_x[1_{\{k<T_x\}}1_{\{X_{k+1}=z\}}]\,.\end{aligned}$$

This identity (4.4.8) is justified by observing that, since $1_{\{k<T_x\}}1_{\{X_k=y\}}$ is $\mathfrak{F}_k$-measurable, the Markov Property at time k gives

$$\begin{aligned}E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}1_{\{X_{k+1}=z\}}] &= E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}1_{\{X_1=z\}}\circ\theta_k]\\ &= E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}E_{X_k}[1_{\{X_1=z\}}]]\\ &= E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}E_y[1_{\{X_1=z\}}]]\\ &= E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}]Q(y,z)\,.\end{aligned}$$

Now, if $z \neq x$, then clearly $1_{\{k<T_x\}}1_{\{X_{k+1}=z\}} = 1_{\{k+1<T_x\}}1_{\{X_{k+1}=z\}}$, and so

$$\sum_{y\in S}\nu_x(y)Q(y,z) = \sum_{k\geq 0} E_x[1_{\{k+1<T_x\}}1_{\{X_{k+1}=z\}}] = E_x\Big[\sum_{k=0}^{T_x-2} 1_{\{X_{k+1}=z\}}\Big] = \nu_x(z)\,.$$

On the other hand, when $z = x$, then $E_x[1_{\{k<T_x\}}1_{\{X_{k+1}=x\}}] = P_x(T_x = k+1)$, and so, since x is recurrent,

$$\sum_{y\in S}\nu_x(y)Q(y,x) = \sum_{k\geq 0} P_x(T_x = k+1) = P_x(T_x<\infty) = 1 = \nu_x(x)\,.$$

This proves that ν_x is invariant. Then, if y belongs to the recurrence class of x, there exists some $m \geq 1$ such that $Q^{(m)}(x,y) > 0$, and so

$$\nu_x(y) = \sum_{z\in S}\nu_x(z)Q^{(m)}(z,y) \geq \nu_x(x)Q^{(m)}(x,y) > 0\,.$$

On the other hand, if y is not in the recurrence class of x, then $N_y = 0$ i.e. $1_{\{X_k=y\}} = 0$ for all $k \geq 0$ P_x-a.s. by Theorem 4.3.1, and so

[3]To see that ν_x is not completely trivial, i.e. that $\nu_x(y) < \infty$ for all $y \in S$, see [**R.88**] p. 301. of Neveu p. 50.

$\nu_x(y) = 0$. To show that $\nu_x(y)$ is finite, observe that invariance of ν_x implies that $\nu_x = \nu_x Q^{(n)}$ for all $n \geq 1$. In particular,

$$1 = \nu_x(x) = \nu_x Q^{(n)}(x) \geq \nu_x(y) Q^{(n)}(y,x) \quad \forall y \in S\,,$$

which implies $\nu_x(y) < \infty$ if $n \geq 1$ is such that $Q^{(n)}(y,x) > 0$. But this is true for at least one n when y belongs to the recurrence class of x. If y is not in the recurrence class of x, we have $\nu_x(y) = 0 < \infty$, as seen above. □

Observe that if there is more than one recurrence class, then the theorem above allows to construct invariant measures with disjoint supports.

THEOREM 4.4.2. *Let the chain be irreducible and all points be recurrent. Then the invariant measure (which exists by Theorem 4.4.1) is unique, up to a multiplicative constant.*

PROOF. Let $x \in S$ and consider the invariant measure ν_x of Theorem 4.4.1. We will show that for any other invariant measure μ,

$$\mu(y) \geq \mu(x)\nu_x(y) \quad \forall y \in S\,. \tag{4.4.9}$$

Assume for a while that this is true. We have, for all $n \geq 1$,

$$\mu(x) = \sum_{z\in S} \mu(z) Q^{(n)}(z,x) \geq \sum_{z \in S} \mu(x)\nu_x(z) Q^{(n)}(z,x) = \mu(x)\,,$$

which gives

$$\sum_{z\in S} [\mu(z) - \mu(x)\nu_x(z)] Q^{(n)}(z,x) = 0\,.$$

Therefore, $\mu(z) = \mu(x)\nu_x(z)$ each time $Q^{(n)}(z,x) > 0$ for some $n \geq 1$. But this is guaranteed by the irreducibility of the chain. Therefore, $\mu = c\nu_x$, with $c = \mu(x)$, proving the theorem. To obtain (4.4.9), we will show, by induction on $p \geq 0$, that ($a \wedge b := \min\{a,b\}$)

$$\mu(y) \geq \mu(x) E_x\Big[\sum_{k=0}^{p\wedge(T_x-1)} 1_{\{X_k = y\}}\Big]\,. \tag{4.4.10}$$

From this, (4.4.9) follows by taking $p \to \infty$. The inequality (4.4.10) is an equality when $y = x$, so we may always consider $y \neq x$. For $p = 0$, the inequality is trivial. Assuming (4.4.10) holds for p,

$$\begin{aligned}\mu(y) = \sum_{z\in S}\mu(z)Q(z,y) \geq \mu(x)\sum_{z\in S}E_x\Big[\sum_{k=0}^{p\wedge(T_x-1)} 1_{\{X_k=z\}}\Big]Q(z,y)\\
= \mu(x)\sum_{z\in S}\sum_{k=0}^{p}E_x[1_{\{k<T_x\}}1_{\{X_k=z\}}]Q(z,y)\,.\\
= \mu(x)\sum_{z\in S}\sum_{k=0}^{p}E_x[1_{\{k<T_x\}}1_{\{X_k=z\}}1_{\{X_{k+1}=y\}}]\\
\qquad (4.4.11)\\
= \mu(x)\sum_{k=0}^{p}E_x[1_{\{k<T_x\}}1_{\{X_{k+1}=y\}}]\end{aligned}$$

In (4.4.11) we used the Markov Property, as in the proof of Theorem 4.4.1. Now, since $y \neq x$,

$$\begin{aligned}\sum_{k=0}^{p}E_x[1_{\{k<T_x\}}1_{\{X_{k+1}=y\}}] = \sum_{k=0}^{p}E_x[1_{\{k+1<T_x\}}1_{\{X_{k+1}=y\}}]\\
= \sum_{l=1}^{p+1}E_x[1_{\{l<T_x\}}1_{\{X_l=y\}}] = E_x\Big[\sum_{l=0}^{(p+1)\wedge(T_x-1)} 1_{\{X_l=y\}}\Big]\,.\end{aligned}$$

This proves (4.4.10) for $p+1$. □

Now that the existence and uniqueness of invariant measures is settled, we turn to the problem of determining whether there exist *finite* measures, i.e. for which $\mu(S) < \infty$, or, which is equivalent, to finding probability distributions on S invariant under Q. This will require a further distinction among recurrent points. Before this, we give a simple result showing that invariant probability measures concentrate on recurrent points. From now on, invariant probability measures will be denoted by π.

Lemma 4.4.2. *Assume there exists an invariant probability π, then each point $x \in S$ with $\pi(x) > 0$ is recurrent.*

Proof. Since π is invariant we have $\pi Q^{(n)} = \pi$ for all $n \geq 1$. Assume $\pi(x) > 0$. Then, using Fubini's Theorem and recalling the definition (4.3.6),

$$\infty = \sum_{n\geq 1} \pi(x) = \sum_{n\geq 1}\sum_{y\in S} \pi(y)Q^{(n)}(y,x) \leq \sum_{y\in S} \pi(y)u(y,x) \leq u(x,x)\,.$$

We used (3) of Lemma 4.3.2 and the fact that π is a probability. By (2) of the same lemma, we conclude that x is recurrent. □

Proposition 4.4.1. *If the chain is irreducible and if there exists an invariant probability π, then it has the form*

$$\pi(x) = \frac{1}{E_x(T_x)} \quad \forall x \in S\,. \tag{4.4.12}$$

Proof. If there exists an invariant probability, then all points are recurrent. Indeed, if there existed a transient point then all points would be transient (since the chain is irreducible), and so $\pi(x) = 0$ for all x by Lemma 4.4.2, a contradiction. We choose any $x \in S$ and show that $\pi(x)$ has the form (4.4.12). By Theorem 4.4.1 there exists an invariant measure ν_x, given in (4.4.7). By Theorem 4.4.2 the invariant measure is unique up to a multiplicative constant. Therefore, if there exists an invariant probability π, then the total mass of ν_x must be finite, $\nu_x(S) < \infty$, and π have the form $\pi = \frac{\nu_x}{\nu_x(S)}$. But

$$\nu_x(S) = \sum_{y\in S} \nu_x(y) = E_x\Big[\sum_{k=0}^{T_x-1}\sum_{y\in S} 1_{\{X_k=y\}}\Big] \equiv E_x(T_x)\,.$$

In particular, $\pi(x) = \frac{\nu_x(x)}{E_x(T_x)} = \frac{1}{E_x(T_x)}$. This shows the theorem. □

The previous result shows that for an invariant measure to exist, one must have $E_x(T_x) < \infty$ for all recurrent point x. This leads to the following distinction among recurrent points.

Definition 4.4.3. *A recurrent point $x \in S$ is called*

- *positive-recurrent if $E_x(T_x) < \infty$,*
- *null-recurrent if $E_x(T_x) = \infty$.*

For example, the simple symmetric random walk on $\mathbb{Z}$ is recurrent, but null-recurrent, as we saw in Theorem 2.1.1. Positive recurrence is a class property: points belonging to the same recurrence class are either all positive-recurrent, or all null-recurrent.

Lemma 4.4.3. *Let the chain be irreducible. Then the following are equivalent.*

(1) *There exists one positive-recurrent point $x \in S$.*
(2) *There exists an invariant probability π.*
(3) *All points $x \in S$ are positive-recurrent.*

PROOF. (1) implies (2): Assume $x \in S$ is positive-recurrent. Consider the invariant measure ν_x. Then $\nu_x(S) = E_x[T_x] < \infty$, and so $\pi := \nu_x(S)^{-1}\nu_x$ is an invariant probability. (2) implies (3): As we saw in Lemma 4.4.2, the existence of an invariant probability implies that $\pi(x) > 0$ for all x. But $\pi(x) = E_x[T_x]^{-1}$ by Proposition 4.4.1, and so $E_x[T_x] < \infty$. (3) implies (1) trivially. □

We gather the results about invariant for irreducible chains in a theorem.

THEOREM 4.4.3. *Let the chain be irreducible and all points be recurrent. Then*

(1) *either each point is positive-recurrent, and there exists a unique invariant probability measure π, $\pi(S) = 1$, given by*

$$\pi(x) = \frac{1}{E_x(T_x)} \quad \forall x \in S \,, \tag{4.4.13}$$

(2) *or each point is null-recurrent, and any invariant measure μ has infinite mass ($\mu(S) = \infty$).*

4.5. Approach to Equilibrium

We now turn to the study of how equilibrium is approached along the time evolution of a Markov chain. Our main purpose is to show that the distribution of the chain converges, in the limit $n \to \infty$, to the invariant measure constructed in Theorem 4.4.3. We will therefore study the limits which appeared in (4.4.3). A detailed study of the convergence to equilibrium can be found in [**Str05**].

The convergence of distribution will be in the sense of the total variation norm, defined, for each $\rho : S \to \mathbb{R}$, by

$$\|\rho\|_{\mathrm{TV}} := \sum_{x \in S} |\rho(x)| \,. \tag{4.5.1}$$

We will say that a sequence of measures $(\mu_n)_{n\geq 1}$ on $(S, \mathcal{P}(S))$ converges to μ if $\|\mu_n - \mu\|_{\mathrm{TV}} \to 0$. As a short hand, we write $\mu_n \Rightarrow \mu$. Our aim is to find under which conditions can one obtain, for a recurrent chain with unique invariant probability π,

$$\mu Q^{(n)} \Rightarrow \pi \,.$$

We will present a standard proof based on a *coupling* argument.

Consider a Markov chain with state space S and transition matrix Q. A *coupling* consists in building two copies of this chain on the cartesian product $\mathbb{S} := S \times S$. We endow $\mathbb{S}$ with the σ-field $\mathcal{P}(\mathbb{S})$. If μ, ν are probability distributions on S, $\mu \otimes \nu$ denotes the probability distribution on $\mathbb{S}$ defined by $(\mu \otimes \nu)(x, y) := \mu(x)\nu(y)$. We then define the following transition matrix on $\mathbb{S}$:

$$\mathbb{Q}((x, y), (x'y')) := Q(x, x')Q(y, y') \,. \tag{4.5.2}$$

By Theorem 4.1.2, we can construct a canonical version of a Markov chain $(X_n, Y_n)_{n\geq 0}$ with state space $\mathbb{S}$, initial distribution $\mu \otimes \nu$ and transition matrix $\mathbb{Q}$. We denote the associated measure by $\mathbb{P}_{\mu\otimes\nu}$. It is clear that under $\mathbb{P}_{\mu\otimes\nu}$, the coupled chain $(X_n, Y_n)_{n\geq 0}$ describes two independent copies of the original markov chain. Its marginals are given by

$$\begin{aligned} &\mathbb{P}_{\mu\otimes\nu}(X_{n+1} = x'|X_n = x) = Q(x, x') \,, \quad \mathbb{P}_{\mu\otimes\nu}(X_n = x) = \mu Q^{(n)}(x) \,, \\ &\mathbb{P}_{\mu\otimes\nu}(Y_{n+1} = y'|Y_n = y) = Q(y, y') \,, \quad \mathbb{P}_{\mu\otimes\nu}(Y_n = y) = \nu Q^{(n)}(y) \,. \end{aligned}$$

A key idea is then to choose $\nu := \pi$, where π is the invariant measure of Q. This implies that under $\mathbb{P}_{\mu\otimes\nu}$, $(Y_n)_{n\geq 0}$ is at equilibrium for all $n \geq 0$:

$$\mathbb{P}_{\mu\otimes\pi}(Y_n = y) = \pi Q^{(n)}(y) = \pi(y) = \mathbb{P}_{\mu\otimes\pi}(Y_0 = y) \,.$$

Therefore,

$$\begin{aligned} P_\mu(X_n = y) - \pi(y) &= \mathbb{P}_{\mu\otimes\pi}(X_n = y) - \mathbb{P}_{\mu\otimes\pi}(Y_n = y) \\ &= \mathbb{E}_{\mu\otimes\pi}\big[1_{\{X_n = y\}} - 1_{\{Y_n = y\}}\big] \,. \end{aligned}$$

Let now $\mathbb{T}$ define the stopping time at which X_n and Y_n meet for the first time:

$$\mathbb{T} := \inf\{n \geq 1 : X_n = Y_n\} \,.$$

In other words, $\mathbb{T}$ is the first time the chain $(X_n, Y_n)_{n\geq 0}$ hits the diagonal $\{(x, x) : x \in S\}$. The point is that if the two chains meet at some time N, then the Markov Property implies that they become probabilistically undistinguishable for times $> N$. We therefore decompose the last expectation with respect to the stopping time $\mathbb{T}$ and to the position of the chain at time $\mathbb{T}$:

$$\begin{aligned}\mathbb{E}_{\mu\otimes\pi}\big[1_{\{X_n=y\}}-1_{\{Y_n=y\}}\big] &= \mathbb{E}_{\mu\otimes\pi}\big[1_{\{\mathbb{T}>n\}}(1_{\{X_n=y\}}-1_{\{Y_n=y\}})\big] \\ &\quad+\sum_{k=1}^{n}\sum_{z\in S}\mathbb{E}_{\mu\otimes\pi}\big[1_{\{\mathbb{T}=k,X_k=Y_k=z\}}(1_{\{X_n=y\}}-1_{\{Y_n=y\}})\big]\,.\end{aligned}$$

This last sum is zero. Indeed, using twice the Markov Property at time k,

$$\begin{aligned}\mathbb{E}_{\mu\otimes\pi}[1_{\{\mathbb{T}=k,X_k=Y_k=z\}}1_{\{X_n=y\}}] &= \mathbb{E}_{\mu\otimes\pi}[1_{\{\mathbb{T}=k,X_k=Y_k=z\}}1_{\{X_{n-k}=y\}}\circ\theta^k] \\ &= \mathbb{E}_{\mu\otimes\pi}[1_{\{\mathbb{T}=k,X_k=Y_k=z\}}]Q^{(n-k)}(z,y) \\ &= \mathbb{E}_{\mu\otimes\pi}[1_{\{\mathbb{T}=k,X_k=Y_k=z\}}1_{\{Y_{n-k}=y\}}\circ\theta^k] \\ &= \mathbb{E}_{\mu\otimes\pi}[1_{\{\mathbb{T}=k,X_k=Y_k=z\}}1_{\{Y_n=y\}}]\,.\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{y\in S}|P_\mu(X_n=y)-\pi(y)| &= \sum_{y\in S}\big|\mathbb{E}_{\mu\otimes\pi}\big[1_{\{\mathbb{T}>n\}}(1_{\{X_n=y\}}-1_{\{Y_n=y\}})\big]\big| \\ &\le 2\sum_{y\in S}\mathbb{E}_{\mu\otimes\pi}[1_{\{\mathbb{T}>n\}}1_{\{Y_n=y\}}] = 2\mathbb{P}_{\mu\otimes\pi}(\mathbb{T}>n)\,,\end{aligned}$$

We are left with

$$\|\mu Q^{(n)}-\pi\|_{\mathrm{TV}} \le 2\mathbb{P}_{\mu\otimes\pi}(\mathbb{T}>n)\,, \tag{4.5.3}$$

which is the standard coupling inequality. We will thus obtain $\mu Q^{(n)}\Rightarrow\pi$ if we can show that the chain $(X_n,Y_n)_{n\ge0}$ is recurrent. The most general way of obtaining this recurrence is under a condition on the chain S called *aperiodicity*, to which we shall turn in a while. Before this we consider a more restrictive condition, but which gives a rate of convergence for the speed at which $\|\mu Q^{(n)}-\pi\|_{\mathrm{TV}}\to 0$.

LEMMA 4.5.1. *Assume the chain S satisfies the following condition: there exists $\ell\ge1$ such that*

$$\inf_{x,y\in S} Q^{(\ell)}(x,y)\ge\delta>0\,. \tag{4.5.4}$$

Then, for all probability distributions μ, ν, we have

$$\mathbb{P}_{\mu\otimes\nu}(\mathbb{T}>k\ell)\le(1-\delta)^k\,,\quad \forall k\ge1\,. \tag{4.5.5}$$

PROOF. We will prove the lemma for Dirac masses $\mu = \delta_x$, $\nu = \delta_y$, in which case the measure is denoted $\mathbb{P}_{(x,y)}$. That is, we will show that for all $k \geq 1$,

$$\mathbb{P}_{(x,y)}(\mathbb{T} > k\ell) \leq (1-\delta)^k \,, \quad \forall (x,y) \in \mathbb{S} \,. \tag{4.5.6}$$

The general case (4.5.5) then follows by summation over $(x, y) \in \mathbb{S}$ [4]. We show (4.5.6) by induction on k. Consider first the case $k = 1$. For any pair (x, y), we have, by (4.5.4),

$$\begin{aligned}\mathbb{P}_{(x,y)}(\mathbb{T} \leq \ell) \geq \mathbb{P}_{(x,y)}(X_\ell = Y_\ell) &= \sum_{z\in S} \mathbb{P}_{(x,y)}(X_\ell = Y_\ell = z) \\ &= \sum_{z\in S} Q^{(\ell)}(x,z)Q^{(\ell)}(y,z) \geq \delta \sum_{z\in S} Q^{(\ell)}(y,z) = \delta \,,\end{aligned}$$

which shows (4.5.6) for $k = 1$. Assume then that (4.5.6) holds for k and for all pair (x, y). Then,

$$\mathbb{P}_{(x,y)}(\mathbb{T} > (k+1)\ell) = \sum_{(s,t)\in\mathbb{S}} \mathbb{P}_{(x,y)}(\mathbb{T} > (k+1)\ell, X_{k\ell} = s, Y_{k\ell} = t) \,.$$

Using the Markov Property at time $k\ell$,

$$\mathbb{P}_{(x,y)}(\mathbb{T} > (k+1)\ell, X_{k\ell} = s, Y_{k\ell} = t) = \mathbb{P}_{(x,y)}(\mathbb{T} > k\ell, X_{k\ell} = s, Y_{k\ell} = t)\mathbb{P}_{(s,t)}(\mathbb{T} > \ell)$$

Using $\mathbb{P}_{(s,t)}(\mathbb{T} > \ell) \leq 1 - \delta$, resumming over $(s, t) \in \mathbb{S}$ and using the induction hypothesis yields (4.5.6) for $k + 1$. □

A direct corollary is then

THEOREM 4.5.1. *Assume the chain S is irreducible and positive recurrent and satisfies* (4.5.4) *for some $\delta > 0$, $\ell \geq 1$. Let π denote the unique invariant probability measure. Then*

$$\|\mu Q^{(n)} - \pi\|_{\mathrm{TV}} \leq 2(1-\delta)^{\lfloor \frac{n}{\ell} \rfloor} \tag{4.5.7}$$

uniformly in all initial distribution μ. In particular, $\mu Q^{(n)} \Rightarrow \pi$.

To emphasize the fact that a chain as above *forgets about its initial condition*, consider two distinct initial distributions μ, μ'. By the triangle inequality,

$$\|\mu Q^{(n)} - \mu' Q^{(n)}\|_{\mathrm{TV}} \leq \|\mu Q^{(n)} - \pi\|_{\mathrm{TV}} + \|\pi - \mu Q^{(n)}\|_{\mathrm{TV}} \to 0 \,,$$

and so the distribution of X_n with initial distribution μ becomes, asymptotically, indistinguishable from the one started with μ'.

[4] A mettre en exercice!

Assumption (4.5.4) is a strong mixing condition. It forces trajectories to meet, in the sense that *any* pair of points (x, y) can be joined during a time interval of length ℓ with positive probability. This implies that two trajectories meet at *some* of the times $\ell, 2\ell, 3\ell, \dots$, and so the coupled chain $\mathbb{S}$ is recurrent. Clearly, (4.5.4) is not realistic when S is infinite, and one can cook up simple examples in which it is not satisfied even in case where S is finite. Consider for example the case where S are the vertices of a square, where the particle can jump to either of its two nearest neighbours with probability $\frac{1}{2}$.

DEFINITION 4.5.1. *Let $x \in S$ be recurrent. Let $I(x) := \{n \geq 1 : Q^{(n)}(x, x) > 0\}$ be the set of times at which a return to x is possible when starting from x. The greatest common divisor of $I(x)$, denoted $d(x)$, is called the* **period** *of x.*

Since x is recurrent, $u(x, x) = \infty > 0$, and so $Q^{(n)}(x, x) > 0$ for infinitely many ns. Therefore, $I(x)$ contains an infinite number of numbers. Moreover, observe that $I(x)$ is stable under addition: if $n, m \in I(x)$ then by the Chapman-Kolmogorov Equation (4.2.3),

$$Q^{(n+m)}(x, x) \geq Q^{(n)}(x, x)Q^{(m)}(x, x) > 0 \,,$$

and so $n + m \in I(x)$.

LEMMA 4.5.2. *If $x, y \in S$ belong to the same recurrence class, then $d(x) = d(y)$.*

PROOF. Since x, y are in the same class, there exists $K \geq 1$ such that $Q^{(K)}(x, y) > 0$ and $L \geq 1$ such that $Q^{(L)}(y, x) > 0$. Therefore,

$$Q^{(K+L)}(y, y) \geq Q^{(L)}(y, x)Q^{(K)}(x, y) > 0 \,,$$

which means that $K + L \in I(y)$, and therefore, $d(y)$ divides $K + L$. Then, consider any $n \in I(x)$. We have

$$Q^{(K+n+L)}(y, y) \geq Q^{(L)}(y, x)Q^{(n)}(x, x)Q^{(K)}(x, y) > 0 \,,$$

which means that $K+n+L \in I(y)$ and therefore, $d(y)$ divides $K+n+L$. Therefore, $d(y)$ divides n. Since this holds for all $n \in I(x)$, $d(y)$ is a divisor of $I(x)$. As a consequence [5], $d(y)$ divides $d(x)$. Changing the roles of y and x shows that $d(x)$ divides $d(y)$, and so $d(x) = d(y)$. □

LEMMA 4.5.3. *If $d(x) = 1$, then there exists m_0 such that $Q^{(n)}(x, x) > 0$ for all $n \geq m_0$.*

[5] Ça je l'ai pas encore bien compris.

PROOF. We first show that when $d(x) = 1$, $I(x)$ must contain two consecutive integers. So let $n_0, n_0 + k \in I(x)$. If $k = 1$ then there is nothing to do. If $k > 1$ then (since $k > d(x)$) there must exist some $l \in I(x)$ which k does not divide. Write $l = km + r$, $0 < r < k$. Since $I(x)$ is stable under addition, the two numbers $(m+1)(n_0+k)$, $(m+1)n_0 + n_1$, are both in $I(x)$. But these numbers differ by less than k:

$$(m+1)(n_0+k) - (m+1)n_0 + n_1 = (m+1)k - (km+r) = k - r < k\,.$$

Proceeding by induction we finally obtain a number N such that $\{N, N+1\} \in I(x)$. Let $m_0 := N^2$. Then each $n \geq N^2$ can be written as $n = N^2 + kN + r$ for some $k \geq 0$, $0 \leq r < N$. We can therefore write n as $n = (N+1)r + N(N - r + k)$, which shows that $n \in I(x)$. □

This lemma says that any point with $d(x) = 1$ can come back to its original position in an arbitray number n of steps, as long as n is sufficiently large. This clearly means that if two independent walks are started at points x, x' with $d(x) = d(x') = 1$, they can meet at any $y \in S$ at time n with positive probability, as soon as n is taken sufficiently large. Of course, depending on x, y, n might have to be taken larger. This shows that imposing $d(x) = 1$ for all $x \in S$ leads to the same recurrence property as (4.5.4), without uniformity in x, y.

DEFINITION 4.5.2. *If $d(x) = 1$ for all $x \in S$, the chain is called aperiodic.*

Aperiodicity is an algebraic property that turns all initial conditions equivalent; it does entail that two trajectories started at two different points have a positive probability of meeting along the evolution, but only just (with no uniformity on the time or points). This is enough to guarantee convergence to equilibrium.

THEOREM 4.5.2. *Let the chain be irreducible and aperiodic. Assume π is an invariant probability. Then for all initial distribution μ,*

$$\|\mu Q^{(n)} - \pi\|_{\mathrm{TV}} \to 0\,.$$

PROOF. We first show that $\mathbb{S}$ is irreducible. So let $(x, y), (x', y')$ be points in $\mathbb{S}$. Since the original chain is irreducible there exist $K \geq 1$ such that $Q^{(K)}(x, x') > 0$ and $L \geq 1$ such that $Q^{(L)}(y, y') > 0$. By Lemma 4.5.3 there exists $n_0 \geq 1$ such that $Q^{(n)}(x, x) > 0$ for all

$n \geq n_0$, and $m_0 \geq 1$ such that $Q^{(m)}(y,y) > 0$ for all $m \geq m_0$. For all $n \geq \max\{n_0, m_0\}$, we have

$$\begin{aligned}\mathbb{Q}^{(K+L+n)}((x,y),(x',y')) &= Q^{(K+L+n)}(x,x')Q^{(K+L+n)}(y,y') \\ &\geq Q^{(L+n)}(x,x)Q^{(K)}(x,x')Q^{(K+n)}(y,y)Q^{(L)}(y,y') > 0\,.\end{aligned}$$

Then, since $\pi \otimes \pi$ is an invariant probability for $\mathbb{Q}$, the chain $\mathbb{S}$ is recurrent (Lemma 4.4.2). By the coupling inequality (4.5.3), this shows that $\|\mu Q^{(n)} - \pi\|_{\mathrm{TV}} \to 0$. □

Observe that there exists a chain which is irreducible, aperiodic, recurrent, but in which two copies don't necessarily meet (see [**R.88**] p. 313).

4.6. The Ergodic Theorem

The notion of *invariant measure*, together with the convergence properties described in Theorems 4.5.1 and 4.5.2, gives a fairly satisfactory description of the asymptotic behaviour of an irreducible Markov chain. What still needs to be done is to see how the *empirical quantities* relate to this asymptotic behaviour. For example: what is, up to time n, the time spent by a chain at a site $y \in S$?

THEOREM 4.6.1 (Ergodic Theorem). *Assume the chain is irreducible and positive recurrent. Let π denote the unique invariant probability measure, and consider a non-negative function $f : S \to \mathbb{R}$, integrable with respect to π: $\int |f| d\pi < \infty$. Then for all $x \in S$,*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \longrightarrow \int f d\pi\,, \quad P_x\text{-a.s.} \tag{4.6.1}$$

This results answers the previous question (in the case of an irreducible, positive recurrent chain). Namely, take $f = \delta_y$. Then $\int f d\pi = \pi(y)$ and by (4.6.1), the fraction of time spent by the chain at y is

$$\frac{1}{n}\sharp\{0 \leq k \leq n-1 : X_k = y\} = \frac{1}{n}\sum_{k=0}^{n-1} 1_{\{X_k=y\}} \longrightarrow \pi(y)\,, \quad P_x\text{-a.s.} \tag{4.6.2}$$

This is very different from the convergence obtained in the previous section. Namely, in the aperiodic case for example, we had obtained $\|\mu Q^{(n)} - \pi\|_{\mathrm{TV}} \to 0$, which implies $P_\mu(X_n = y) \to \pi(y)$ for all $y \in S$, which is a probability of what happens *at time n*. On the other hand, (4.6.2) gives an *almost sure* convergence of the time the *trajectory* spends at y *up to* time n.

PROOF OF THEOREM 4.6.1: Let $x \in S$. We consider the partition of the trajectory into the successive returns of the chain at x:

$$0 =: T_x^0 < T_x^1 < T_x^2 < \dots ,$$

where $T_x^1 := T_x$, and for $k \geq 2$,

$$T_x^k := \inf\{n > T_x^{k-1} : X_n = x\} .$$

Since the chain is irreducible and since there exists an invariant probability, the chain is recurrent (Lemma 4.4.2), and each T_x^k is P_x-almost surely finite. The result will follow from the fact that the events happening during the time intervals $[T_x^k, T_x^{k+1})$ are independent, and from the Law of Large Numbers. Fix $f : S \to \mathbb{R}$ and define, for all $k \geq 0$,

$$Z_k := \sum_{j=T_x^{(k)}}^{T_x^{(k+1)}-1} f(X_j) .$$

Clearly, $Z_k = Z_0 \circ \theta_{T_x^{(k)}}$.

LEMMA 4.6.1. *The sequence* $(Z_n)_{n\geq 0}$ *is i.i.d.*

PROOF. First observe that for all positive measurable bounded $g : \mathbb{R} \to \mathbb{R}$, the Markov Property at time $T_x^{(k)}$ gives

$$E_x[g(Z_k)] = E_x[(g \circ Z_0) \circ \theta_{T_x^{(k)}}] = E_x[g(Z_0)] ,$$

and so the Z_ks are identically distributed. For the independence, it is sufficient to show that for all $k \geq 0$,

$$E_x[g_0(Z_0) \dots g_k(Z_k)] = E_x[g_0(Z_0)] \dots E_x[g_k(Z_0)] , \tag{4.6.3}$$

where $g_j : \mathbb{R} \to \mathbb{R}$, $j = 0, 1, \dots, k$ are arbitrary bounded functions. This is trivially true when $k = 0$, so assume (4.6.3) holds for $k - 1$. Using again the Markov property at time $T_x^{(k)}$ and the induction hypothesis,

$$\begin{aligned} E_x[g_0(Z_0) \dots g_k(Z_k)] &= E_x[g_0(Z_0) \dots g_{k-1}(Z_{k-1})] E_x[g_k(Z_0)] \\ &= E_x[g_0(Z_0)] \dots E_x[g_k(Z_0)] . \end{aligned}$$

This shows (4.6.3) for k. □

Now, observe that since π is invariant it must have the form $\pi = \pi(x)\nu_x$ for all x, where ν_x is the invariant measure of Theorem 4.4.1. This implies that

$$\begin{aligned} E_x[|Z_0|] = E_x[Z_0] = E_x\Big[\sum_{j=0}^{T_x-1} f(X_j)\Big] \\ &= \sum_{y\in S} f(y)E_x\Big[\sum_{j=0}^{T_x-1} 1_{\{X_j=y\}}\Big] \\ &= \int f d\nu_x = \frac{1}{\pi(x)}\int f d\pi < \infty\,. \end{aligned}$$

Therefore, by the Strong Law of Large Numbers,

$$\frac{1}{n}\sum_{k=0}^{n-1} Z_k \longrightarrow \frac{1}{\pi(x)}\int f d\pi \quad P_x\text{-a.s.} \tag{4.6.4}$$

Let $N_x(n)$ be the number of visits of the chain at x up to time n. Then $T_x^{N_x(n)} \leq n < T_x^{N_x(n)+1}$, and since f is non-negative,

$$\frac{1}{N_x(n)}\sum_{k=0}^{T_x^{N_x(n)}-1} f(X_k) \leq \frac{1}{N_x(n)}\sum_{k=0}^{n} f(X_k) \leq \frac{1}{N_x(n)}\sum_{k=0}^{T_x^{N_x(n)}+1} f(X_k)\,,$$

which is the same as

$$\frac{1}{N_x(n)}\sum_{j=0}^{N_x(n)-1} Z_j \leq \frac{1}{N_x(n)}\sum_{k=0}^{n} f(X_k) \leq \frac{1}{N_x(n)}\sum_{j=0}^{N_x(n)} Z_j\,,$$

By (4.6.4) and since $N_x(n)\to\infty$ P_x-a.s. when $n\to\infty$ (Proposition 4.3.1),

$$\frac{1}{N_x(n)}\sum_{k=0}^{n} f(X_k) \longrightarrow \int f d\nu_x\,.$$

The same expression with $f = 1$ gives $\frac{n}{N_x(n)} \to \nu_x(S) = \frac{1}{\pi(x)}$. This finishes the proof. □

4.7. Exercises

Generalities.

EXERCISE 4.1. [**GS05**] p. 219. A Die is rolled repeatedly. Which of the following are Markov chains? For those that are, supply the transition matrix.

- The largest number X_n shown up to time n.
- The number N_n of sixes in n rolls.
- At time r, the time C_r since the most recent six.
- At time r, the time B_r until the next six.

EXERCISE 4.2. [**GS05**] p. 219. Let $(X_n)_{n\geq 0}$ be the simple random walk starting at the origin. Are $(|X_n|)_{n\geq 0}$ and $(M_n)_{n\geq 0}$ Markov chains? (We defined $M_n := \max\{X_k : 0 \leq k \leq n\}$.) When this is the case, compute the transition matrix. Show that $Y_n := M_n - X_n$ defines a Markov chain. What happens if $X_0 \neq 0$.

EXERCISE 4.3. [**GS05**] p. 220. Let X_n, Y_n be Markov chains on $S = \mathbb{Z}$. is $X_n + Y_n$ necessarily a Markov chain?

EXERCISE 4.4. [**GS05**] p. 220. Let X_n be a Markov chain. Show that for all $1 < r < n$,

$$\begin{aligned} &P(X_r = x | X_i = x_i,\ i = 1, 2, \ldots, r-1, r+1, \ldots, n) \\ &\qquad = P(X_r = x | X_{r-1} = x_{r-1}, X_{r+1} = x_{r+1})\,. \end{aligned}$$

EXERCISE 4.5. Consider "Markov's Other chain" ([**GS05**] p. 218): let $Y_1, Y_3, Y_5, \ldots$ be a sequence of independent identically distributed random variables such that

$$P(Y_{2k+1} = -1) = P(Y_{2k+1} = +1) = \frac{1}{2}.$$

Define then $Y_{2k} := Y_{2k-1}Y_{2k+1}$. Check that $Y_2, Y_4, Y_6, \ldots$ are identically distributed, with the same distribution as above. Is $(Y_k)_{k\geq 1}$ a Markov chain? Enlarge the state space to $\{\pm 1\}^2$ and define $Z_n := (Y_n, Y_{n+1})$.

EXERCISE 4.6. [**R.88**] p.281. Two state Markov chain. Let $S = \{0, 1\}$ with transition matrix

$$Q = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$$

For example, if X_n represents the weather on day n, with say 0 being bad weather, 1 being nice weather, a reasonable optimistic choice is $\alpha = 0.6$, $\beta = 0.2$.
Let μ be a probability distribution on S. Show by induction that

$$P_\mu(X_n = 0) = \frac{\beta}{\beta+\alpha} + (1-\alpha-\beta)^n\Big(\mu(0) - \frac{\beta}{\beta+\alpha}\Big)\,.$$

Compute $n \to \infty$. Set $p_n := P_x(X_n = 1)$, and show that when $0 < \alpha + \beta < 2$, (voir les notes de Durrett)

$$|p_n - \frac{\beta}{\beta + \alpha}| \leq |1 - \alpha - \beta|^n$$

In which cases does p_n *not* converge? Hint: set $p_n := P_x(X_n = 1)$. Use the Markov Property, write $p_n = (1-\alpha)p_{n-1} + (1-p_{n-1})\beta$, and show that

$$p_n - \frac{\beta}{\beta+\alpha} = \Big(p_{n-1} - \frac{\beta}{\beta+\alpha}\Big)(1-\alpha-\beta)$$

EXERCISE 4.7. Let $(Y_n)_{n\geq 0}$ be i.i.d., $P(Y_k = 0) = 1 - P(Y_k = 1) = \frac{1}{2}$. Show that $X_n := (Y_n, Y_{n+1})$ is a Markov chain and compute its transition probability matrix Q. What is $Q^{(2)}$?

EXERCISE 4.8. Prove Lemma 4.1.1: $(X_n)_{n\geq 0}$ is a Markov chain with transition matrix Q if and only if for all $n \geq 1$ and all $x_0, \dots, x_n \in S$,

$$P(X_0 = x_0, \dots, X_n = x_n) = P(X_0 = x_0)Q(x_0, x_1)\dots Q(x_{n-1}, x_n)\,.$$

EXERCISE 4.9. The Canonical Markov Chain.

(1) Show the equivalence between the σ-algebra generated by the coordinate maps $\sigma(X_k)_{n\geq 0}$ and the one generated by cylinders.
(2) Show that the shift θ is measurable.
(3) Show the equivalence between (4.2.1) and (4.2.2).
(4) Let $\psi : \Omega \to \mathbb{R}$ be measurable, positive, bounded. Show that $x \mapsto E_x(\psi)$ is measurable.

The Markov Property.

EXERCISE 4.10. Fill in the details at the end of the proof of Theorem 4.2.1.

EXERCISE 4.11. (Durrett p. 283) Using the Markov Property, show that if $A \in \sigma(X_0, \dots, X_n)$ and $B \in \sigma(X_n, X_{n+1}, \dots)$, then

$$P_\mu(A \cap B | X_n) = P_\mu(A|X_n)P_\mu(B|X_n)\,.$$

Hint: Write the left-hand side as $E_\mu(E_\mu(1_A 1_B|\mathcal{F}_n)|X_n)$.

EXERCISE 4.12. Stopping times.

(1) Show that $\mathcal{F}_T$ is a σ-algebra.
(2) Show that X_T is $\mathcal{F}_T$-measurable.

EXERCISE 4.13. Let $T_y := \inf\{n \geq 1 : X_n = y\}$. Use the Strong Markov Property to show that

$$P_x(X_n = y) = \sum_{m=1}^{n} P_x(T_y = m) P_y(X_{n-m} = y) .$$

Compare with the Chapman-Kolmogorov Equation, whose proof requires only the Simple Markov Property.

Recurrence and Classification.

EXERCISE 4.14. Dacunha-Castelle p.185. Classify the states of the Markov chains on $S = \{1, 2, 3, 4\}$ whose transition matrices are given respectively by

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}, \quad \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

EXERCISE 4.15. Classify the states of the Markov chain of Exercise 4.6 in function of α and β.

Random Walks.

EXERCISE 4.16. [**Str05**] p. 16. Prove the Cauchy-Schwartz Inequality: for any pair of sequences $(a_n)_{n\in\mathbb{Z}}$, $(b_n)_{n\in\mathbb{Z}}$,

$$\sum_{n\in\mathbb{Z}} |a_n b_n| \leq \Big(\sum_{n\in\mathbb{Z}} a_n^2\Big)^{\frac{1}{2}} \Big(\sum_{n\in\mathbb{Z}} b_n^2\Big)^{\frac{1}{2}}$$

(1) Show that it is sufficient to consider the case in which $a_n = b_n = 0$ for all but a finite number of ns.
(2) Given $f(x) = Ax^2 + 2Bx + C$, show that $f \geq 0$ if and only if $C \geq 0$ and $B^2 \leq AC$.
(3) In the case where $a_n = b_n = 0$ for all but a finite number of ns, set $g(x) = \sum_n (a_n x + b_n)^2$ and apply the previous step.

EXERCISE 4.17. Consider a random walk on $\mathbb{Z}^d$ given by a transition matrix

$$Q(x, y) = \frac{1}{2^d} \prod_{i=1}^{d} 1_{|x_i - y_i| = 1} .$$

Study the recurrence of this chain (in function of the dimension) by computing $Q^{(n)}(0, 0)$.

Exercise 4.18. Dacunha-Castelle p. 193. Consider the simple symmetric random walk on $\mathbb{Z}^d$ starting at the origin. Let T_n be the number of visits of the walk at the origin up to time n. Study the asymptotics of $E[T_n]$ for large n, for the cases $d = 1, 2, 3$. For example, show that in $d = 1$,

$$E[T_n] \sim \frac{2}{\sqrt{\pi}}\sqrt{n}\,.$$

Hint: First, observe that $\{T_n = k\} = \{T_0^{(k)} \leq n, T_0^{(k+1)} > n\}$, and so $P(T_n = k) = P(T_0^{(k)} \leq n) - P(T_0^{(k+1)} \leq n)$, which gives

$$E[T_n] = \sum_{k\geq 1} P(T_0^{(k)} \leq n) = \sum_{j=1}^{2n} P(X_j = 0)\,.$$

Exercise 4.19. Consider a random walk on $\mathbb{Z}^d$, $S_0 = 0$, $S_n = X_1 + \cdots + X_n$, whose increments X_k have the distribution $P(X_k = x) = p(x)$. The purpose of this exercise to show the following Recurrence Criterium (compare with (4.3.12)):

$$\text{The walk is recurrent} \quad \Leftrightarrow \quad \int_{[-\pi,\pi]^d} \frac{1}{1-\varphi(\xi)} d\xi = \infty\,. \qquad (4.7.1)$$

Here, $\varphi(\xi) = E[e^{i\xi\cdot X_1}]$, $\xi \in \mathbb{R}^d$, is the characteristic function of X_1.

(1) For any $\mathbb{Z}^d$-valued random variable Z, define $p_Z(x) := P(Z = x)$, and denote the characteristic function of Z by φ_Z. Prove the inversion formula:

$$p_Z(x) = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} e^{-i\xi\cdot x}\varphi_Z(\xi) d\xi$$

(2) Compute $P(S_n = 0)$ and $\sum_n P(S_n = 0)$, prove the recurrence criterium (Hint: compute first $\sum_n \theta^n P(S_n = 0)$ and then take $\theta \to 1^-$).
(3) Apply the criterium to show that the simple symmetric random walk is recurrent for $d = 1, 2$, and transient for all $d \geq 3$. Hint: once you have computed the characteristic function of X_1, show that $1 - \cos x \geq \frac{x^2}{4}$ when $|x| \leq \frac{\pi}{3}$.
(4) Apply the criterium to the random walk of Exercise 4.17.

Invariant Distributions.

Exercise 4.20. Vares-Olivieri Chap. 4.3. Show that if a distribution is reversible with respect to a birth and death chain, then it is invariant. Find an example of a distribution which is invariant but not reversible.

Exercise 4.21. Total Variation Norm. Let S be countable and $\rho : S \to \mathbb{R}$. Define

$$\|\rho\|_{\mathrm{TV}} := \sum_{x \in S} |\rho(x)| .$$

(1) Show that $\| \cdot \|_{\mathrm{TV}}$ is a norm.
(2) Consider the normed vector space $\mathcal{M}^1 := \{\rho : S \to \mathbb{R} : \|\rho\|_{\mathrm{TV}} < \infty\}$. Let Q be a transition probability matrix on S. Show that for all $\rho \in \mathcal{M}^1$, $\rho Q : S \to \mathbb{R}$ is well defined: $\rho Q(x) := \sum_{y \in S} \rho(y) Q(y,x)$, and that $\|\rho Q\|_{\mathrm{TV}} \leq \|\rho\|_{\mathrm{TV}}$.
(3) A sequence $(\rho_n)_{n \geq 1}$ in $\mathcal{M}^1$ is Cauchy if for all $\epsilon > 0$ there exists N_0 such that $\|\rho_n - \rho_m\|_{\mathrm{TV}} \leq \epsilon$ for all $n, m \geq N_0$. Show that $\mathcal{M}^1$ is complete: $(\rho_n)_{n \geq 1}$ is Cauchy if and only if there exists ρ such that $\|\rho_n - \rho\|_{\mathrm{TV}} \to 0$.

Exercise 4.22. ([**GS06**] p. 77) Find an invariant probability measure for the Markov chain of Exercise 4.6. Suppose that $0 < \alpha\beta < 1$. Find $Q^{(n)}$. Fix some initial distribution μ and study $\|\mu Q^{(n)} - \pi\|_{\mathrm{TV}}$ for large n. For what values of α, β is the chain reversible in equilibrium?

Exercise 4.23. [**R.88**] p. 305 Show that the random walk on a graph (see Example 4.1.4) is irreducible if and only if the graph is connected. Show that the walk is positive-recurrent if and only if the graph is finite. In this case, show that the invariant probability measure is given by $\mu(x) = \frac{d(x)}{N}$, where $d(x)$ is the degree of the vertex x and N is the number of vertices of the graph ([**GS05**] p. 236, ex. 6).

Exercise 4.24. Ehrenfest Urn Model.

(1) Compute the invariant measure of the Ehrenfest model. (Voir p.187 de Dacunha-Castelle ou Le Gall p.19.)
(2) Does $P_\mu(X_n = x)$ converge? Why? Find two distributions μ, ν for which

$$\liminf_n \|\mu Q^{(n)} - \nu Q^{(n)}\|_{\mathrm{TV}} > 0 .$$

Exercise 4.25. Give an example of a Markov chain and of a distribution which is invariant but not reversible.

Exercise 4.26. [**GS06**] p. 69 Let $(S_n)_{n\geq 0}$, $(S'_n)_{n\geq 0}$ be two copies of the simple symmetric random walk. Let $Z_n := (S_n, S'_n)$ be constrained to lie in the region $S_n \geq 0$, $S'_n \geq 0$, $S_n + S'_n \leq a$, for some integer $a \geq 1$. Find the stationary distribution of Z_n. What happens when $a \to \infty$?

Exercise 4.27. ([**GS06**] p. 77) Let $S = \{0, 1, 2, \dots\}$ and $Q(x, x+1) = p_x$, $Q(x, 0) = 1 - p_x$, where $0 < p_x < 1$. Let $b_x = p_0 p_1 \dots p_{x-1}$. Show that the chain is

(1) recurrent if and only if $b_x \to 0$ when $x \to \infty$,
(2) non-null recurrent if and only if $\sum_x b_x < \infty$,

and write down the stationary distribution if the latter condition holds. Then, let $a, \beta > 0$ and assume $p_x = 1 - ax^{-\beta}$. Show that the chain is

(1) transient if $\beta > 1$,
(2) non–null recurrent if $\beta < 1$,

and that if $\beta = 1$,

(1) non-null recurrent if $a > 1$,
(2) null recurrent if $a \geq 1$.

VOIR AUSSI NEVEU p. 78., Varadhan p. 145

Exercise 4.28. Birth and Death chains. Voir Durrett p. 297, Le Gall p. 23. ET NEVEU p. 81., VARADHAN p. 145 Let $S = \{0, 1, 2, \dots\}$, with the transition matrix $Q(x, x+1) = p_x$, $Q(x, x-1) = q_x$, $Q(x, x) = r_x$, $q_0 = 0$, $p_x + q_x + r_x = 1$ for all x.

(1) Show that $\mu(0) := p_1$,
$$\mu(x) = \frac{p_0 p_1 \cdots p_{x-1}}{q_1 q_2 \cdots q_x} \quad \forall x \geq 1$$
is the only invariant measure (up to a multiplicative constant).
(2) Consider the particular case $r_x = 0$, $p_x = 1 - q_x = p$. For which ps is the measure μ finite? For which ps is the chain null-recurrent? recurrent? transient? (Hint: use Lemma 4.4.2)

Exercise 4.29. [**GS05**] ex. 7, p. 236. Show that the random walk on the infinite binary tree is transient. Hint: use either Exercise 4.28.

Exercise 4.30. [**R.88**] p. 313. The Bernoulli-Laplace Model of Diffusion. Consider two boxes having each N particles, b of which are black (we assume $b \leq N$), $2N - b$ of which are white. At each time n, we pick a ball from each box and interchange them. Let X_n denote the number of black particles in the left box. Compute the transition matrix of the Markov chain $(X_n)_{n\geq 1}$. Study convergence to equilibrium. See *www.math.uah.edu/stat/markov/index.xhtml*.

Exercise 4.31. [**R.88**] p. 302. Consider the simple symmetric random walk on $\mathbb{Z}$ starting at the origin. Show that the expected number of visits to a site $x \neq 0$ before the time of first return is 1. *Hint:* use Theorems 4.4.1 and 4.4.2. Il suffit de prendre la fomule en page 47 de Neveu: $E_x(N_y^x) = \frac{\mu(y)}{\mu(x)}$, donc ici puisque la mesure de comptage est invariante ca montre le resultat.

Exercise 4.32. [**R.88**] p. 302.

(1) Show that $\nu_x(y)\nu_y(z) = \nu_x(z)$.
(2) Let $w_{xy} = P_x(T_y < T_x)$. Show that

$$\nu_x(y) = \frac{w_{xy}}{w_{yx}} \quad \forall y \in S\,.$$

Use this to solve Exercise 4.31.

Exercise 4.33. [**R.88**] p. 305. Compute the expected number of moves it takes a knight to return to its initial position if it starts in a corner of the chessboard, assuming there are no other pieces on the board, and each time it chooses a move at random from its legal moves. *Hint:* a chessboard has $S = \{0, 1, \ldots, 7\}^2$. A knight's move is L-shaped: two steps in one direction followed by one step in a perpendicular direction. Same with a king.

Exercise 4.34. [**GS06**] p. 68 A particle performs a random walk on the bow tie ABCDE drawn on Figure 2, where C is the knot. From any vertex the next step is equally likely to be any neighbouring vertex. The particle starts at A. Find the expected value of

(1) the time of first return to A
(2) the number of visits to D before returning to A
(3) the number of visits to C before returning to A
(4) the time of first return to A, given no prior visit by the particle to E
(5) the number of visits to D before returning to A, given no prior visit by the particle to E.

Exercise 4.35. [**GS06**] p. 68. A particle performs a symmetric random walk on the square graph of Figure 2, starting from A. Find the expected number of visits to B before it returns to A.

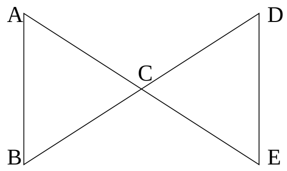


FIGURE 2

EXERCISE 4.36. [**R.88**] p. 279, 304. Show that an irreducible renewal chain is positive recurrent if and only if $\sum_k k f_k < \infty$.

EXERCISE 4.37. (Dacunha-Castelle, p. 191) Consider the successive return times at a given site x: $T_x^{(1)} < T_x^{(2)} < \dots$. Define

$$g(s) = \sum_{n\geq 0} Q^{(n)}(x,x)s^n\,, \qquad h(s) = \sum_{n\geq 1} P_x(T_x = n)s^n\,.$$

Express g as a function of h. The following formula might be useful:

$$Q^n(x,x) = \sum_{k=1}^{n} P_x(T = k)Q^{(n-k)}(x,x)$$

What should g satisfy in order for T to be almost surely finite? Show that for all sequence $n_1 < n_2 < \dots$,

$$\begin{aligned} P_x(T_x^{(1)} = n_1, T_x^{(2)} - T_x^{(1)} = n_2 \dots, T_x^{(k)} - T_x^{(k-1)} = n_k) \\ = \prod_{j=1}^{k} P_x(T_x^{(j)} - T_x^{(j-1)} = n_j, T_x^{(j)} < \infty) \\ = \prod_{j=1}^{k} P_x(T_x^{(1)} = n_j)\,. \qquad (4.7.2) \end{aligned}$$

Assume x is transient. Show that the chain visits x almost surely a finite number of times (hint: use (4.7.2) and Borel-Cantelli for the sequence of events $\{T_x^{(k)} < \infty\}$). From (4.7.2), deduce that the sequence $(T_x^{(k)} - T_x^{(k-1)})_{k\geq 1}$ is i.i.d.

EXERCISE 4.38. Stroock [**Str05**] p. 28., Lindvall [**Lin92**] p. 54. Doeblin's Coupling. Let Q be a transition probability matrix on S. The aim of this exercise is to show the following result, due to Doeblin ([**Doe38**], 1938): *Assume there exists $y_0 \in S$, $\epsilon > 0$ such that*

$$\inf_{x\in S} Q(x, y_0) \geq \epsilon\,. \qquad (4.7.3)$$

Then Q has a unique invariant distribution π, and for any initial distribution μ,

$$\|\mu Q^{(n)} - \pi\|_{\text{TV}} \leq 2(1-\epsilon)^n \,.$$

We use the notations and results of Exercise 4.21.

(1) Show first that if (4.7.3) holds, and if $\rho : S \to \mathbb{R}$ is such that $\sum_{x\in S} \rho(x) = 0$, then

$$\|\rho Q\|_{\text{TV}} \leq (1-\epsilon)$$

Hint: Write $\sum_{x\in S} \rho(x) Q(x,y) = \sum_{x\in S} \rho(x)(Q(x,y) - \epsilon\delta_{y,y_0})$.

(2) By induction, show that for all $n \geq 1$,

$$\|\rho Q^{(n)}\|_{\text{TV}} \leq (1-\epsilon)^n \|\rho\|_{\text{TV}} \,.$$

(3) Let μ be an initial distribution, set $\mu_n := \mu Q^{(n)}$. Show that $\{\mu_n\}_{n\geq 1}$ is a Cauchy sequence to conclude that $\pi := \lim_n \mu Q^{(n)}$ exists. Hint: for $n \geq m$, write $\mu_n - \mu_m = (\mu Q^{(n-m)} - \mu) Q^{(m)}$.
(4) Show that π is stationary. Conclude.

Show the following generalization of the previous theorem: *Assume there exists $y_0 \in S$, $M \geq 1$ and $\epsilon > 0$ such that*

$$\inf_{x\in S} Q^{(M)}(x, y_0) \geq \epsilon \,. \tag{4.7.4}$$

Then Q has a unique invariant distribution π, and for any initial distribution μ,

$$\|\mu Q^{(n)} - \pi\|_{\text{TV}} \leq 2(1-\epsilon)^{\lfloor \frac{n}{M} \rfloor} \,.$$

(1) Set $\widetilde{Q} := Q^{(M)}$, apply the preceding theorem.
(2) Write any n as $n = mM + r$, $0 \leq r < M$, $\mu Q^{(n)} - \pi = (\mu Q^{(r)} - \pi)\widetilde{Q}^{(m)}$, and use the first part of the proof of the first theorem to conclude.
(3) Compute the rate of convergence for some of the finite state Markov chains encountered above.

Exercise 4.39. Consider the Ehrenfest model of Exercise 4.24, together with its invariant measure π. Although we saw in Exercise 4.24 that $P(X_n = k)$ has no limit when $n \to \infty$, show that the average time spent by the chain at k has a limit when $n \to \infty$. Compute this limit.

Exercise 4.40. [**Str05**] p. 124. Gibbs Distributions. Voir aussi Olivieri-Vares, p. 250 pour le champ moyen.

EXERCISE 4.41. $http://faculty.uml.edu/jpropp/584/$, bouquin de Charles M. Grinstead and J. Laurie Snell p. 418., The Fundamental Matrix for absorbing chains.

CHAPTER 5

Martingales

Many probabilists specialize in limit theorems, and much of applied probability is devoted to finding such results. The accumulated literature is vast and the techniques multifarious. One of the most usefull skills for establishing such results is that of martingale divination, because the convergence of martingales is guaranteed.

G. Grimmet and D. Stirzaker, [**GS05**].

Consider the simple random walk on $\mathbb{Z}$: $S_n = Y_1 + \cdots + Y_n$, where $P(Y_k = +1) = 1 - P(Y_k = -1) = p$, with $p \in (0,1)$. Observe that $|S_n| \leq n$ and so

$$S_n \in L^1 \quad \forall n \geq 1 \,. \tag{5.0.5}$$

Moreover, if we define $(\mathcal{F}_n)_{n\geq 1}$ by $\mathcal{F}_n := \sigma(Y_1, \dots, Y_n)$, then

$$E[S_{n+1}|\mathcal{F}_n] = E[S_n + Y_{n+1}|\mathcal{F}_n] = S_n + E[Y_{n+1}|\mathcal{F}_n] = S_n + 2p - 1 \,. \tag{5.0.6}$$

In particular, if $p = \frac{1}{2}$, then

$$E[S_{n+1}|\mathcal{F}_n] = S_n \quad \forall n \geq 1 \,. \tag{5.0.7}$$

(5.0.5) and (5.0.7) are the two properties defining a *martingale*.

5.1. Definition and Examples

Martingales describe sequences of integrable random variables $(X_n)_{n\geq 1}$ which respect a condition of the type (5.0.7): interpreting n as discrete time, the expectation of X_{n+1}, conditionned on the information encoded in the variables $X_1, \dots, X_n$, is equal to X_n. We will describe a slightly more general situation, where the sequence $(\mathcal{F}_n)_{n\geq 1}$ is defined *a priori*, without necessary reference to a sequence of variables. Throughout this section, $(\Omega, \mathcal{F}, P)$ is an arbitrary probability space.

Definition 5.1.1. *A* ***filtration*** *is an increasing sequence* $(\mathcal{F}_n)_{n\geq 1}$ *of sub-*σ*-algebras* $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}$. *A sequence of random variables* $(X_n)_{n\geq 1}$ *on* $(\Omega, \mathcal{F})$ *is* ***adapted to*** $(\mathcal{F}_n)_{n\geq 1}$ *if* X_n *is* $\mathcal{F}_n$*-measurable for all*

$n \geq 1$. A double sequence $(X_n, \mathcal{F}_n)_{n\geq 1}$, where $(\mathcal{F}_n)_{n\geq 1}$ is a filtration and $(X_n)_{n\geq 1}$ is adapted to $(\mathcal{F}_n)_{n\geq 1}$, is called a stochastic sequence [1].

When, as above, a filtration is associated to a sequence of random variables as $\mathcal{F}_n := \sigma(Y_1, \ldots, Y_n)$, we call it the natural (or canonical) filtration associated to $(Y_n)_{n\geq 1}$.

DEFINITION 5.1.2. *A stochastic sequence $(X_n, \mathcal{F}_n)_{n\geq 1}$ in which $X_n \in L^1$ is called*

(1) *a martingale if for all $n \geq 1$:*
$$E[X_{n+1}|\mathcal{F}_n] = X_n \,, \tag{5.1.1}$$
(2) *a submartingale if for all $n \geq 1$:*
$$E[X_{n+1}|\mathcal{F}_n] \geq X_n \,, \tag{5.1.2}$$
(3) *a supermartingale if for all $n \geq 1$:*
$$E[X_{n+1}|\mathcal{F}_n] \leq X_n \,, \tag{5.1.3}$$

When the filtration under consideration is clear in the context, we will write $(X_n)_{n\geq 1}$ rather than $(X_n, \mathcal{F}_n)_{n\geq 1}$. We have omitted to mention that each of the expressions (5.1.1)-(5.1.3) holds P-almost everywhere. We will continue to do so in the sequel, unless when the specification of the measure P will be necessary.

There is a condition equivalent to (5.1.1) called the martingale property, which characterizes a martingale. Namely: for all $n \geq 1$,
$$\int_A X_n \, dP = \int_A X_m \, dP \,, \quad \forall A \in \mathcal{F}_n \,, \forall m \geq n \,. \tag{5.1.4}$$
In particular, using this identity with $A = \Omega$, we see that
$$E[X_n] = E[X_1] \,, \quad \forall n \geq 1 \,. \tag{5.1.5}$$
Considering submartingales (resp. supermartingales), the same holds with $=$ in (5.1.4)-(5.1.5) replaced by $\geq$ (resp. $\leq$).

Observe that sequences of independent random variables don't, in general, form martingales. The simplest example of a submartingale is provided by a non-decreasing sequence $(a_n)_{n\geq 1}$: if $X_n = a_n$ for all n, then $(X_n)_{n\geq 1}$ is a submartingale with respect to any filtration. Let us

[1] This terminology is taken from Shiryayev [**Shi84**]

give more interesting examples to which we will come back often in the sequel.

Example 5.1.1. Consider the random walk $(S_n)_{n\geq 1}$ described above, with its natural filtration. Then (5.0.6) shows that $(S_n)_{n\geq 1}$ is a martingale when $p = \frac{1}{2}$, a submartingale when $p > \frac{1}{2}$, and a supermartingale when $p < \frac{1}{2}$.

Example 5.1.2. A similar construction can be done with products: let $M_n = Y_1 Y_2 \cdots Y_n$, where the Y_k are independent, non-negative, and $E[Y_k] = 1$. Then $E[M_n] = E[Y_1]^n = 1 < \infty$. If $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$, then $(M_n, \mathcal{F}_n)_{n\geq 1}$ is a martingale. Namely, since M_n is $\mathcal{F}_n$-measurable [2],

$$E[M_{n+1}|\mathcal{F}_n] = E[M_n Y_{n+1}|\mathcal{F}_n] = M_n E[Y_{n+1}|\mathcal{F}_n] = M_n E[Y_{n+1}] = M_n\,.$$

Example 5.1.3. Let $(\mathcal{F}_n)_{n\geq 1}$ be any filtration, $X \in L^1$. Set $X_n := E[X|\mathcal{F}_n]$. Then $(X_n, \mathcal{F}_n)_{n\geq 1}$ is a martingale. Namely, X_n is $\mathcal{F}_n$-measurable and

$$E[|X_n|] = E[|E[X|\mathcal{F}_n]|] \leq E[E[|X||\mathcal{F}_n]] = E[|X|] < \infty\,,$$

which implies $X_n \in L^1$. Then, since $\mathcal{F}_n \subset \mathcal{F}_{n+1}$,

$$E[X_{n+1}|\mathcal{F}_n] = E[E[X|\mathcal{F}_{n+1}]|\mathcal{F}_n] = E[X|\mathcal{F}_n] = X_n\,.$$

Martingales of this type are called closed. X_n can be interpreted as the best approximation of X given the partial information contained in the occurence or non-occurence of the events in $\mathcal{F}_n$. Intuitively, X_n should converge to X, which will be confirmed in Lévy's Upward Theorem.

The following example shows that martingales will also be useful in the study of purely measure-theoretical problems.

Example 5.1.4. Consider the probability space $([0,1), \mathcal{B}([0,1)), \lambda)$, where $\mathcal{B}([0,1))$ denotes the Borel σ-algebra, and λ the Lebesgue measure. Consider the dyadic intervals $I_k^n = [\frac{k-1}{2^n}, \frac{k}{2^n})$ and the associated dyadic filtration $(\mathcal{F}_n)_{n\geq 1}$, defined by

$$\mathcal{F}_n := \sigma\big(I_k^n : i = 1, 2, \dots, 2^n\big)\,.$$

If μ is any finite measure on $([0,1), \mathcal{B}([0,1)))$, we define the random variables

$$X_n := \sum_{i=1}^{2^n} \frac{\mu(I_k^n)}{\lambda(I_k^n)} 1_{I_k^n}\,. \tag{5.1.6}$$

[2] We refer the reader to Chapter 1 for the general properties of conditional expectation that are used throughout the present chapter.

For large n, these random variables give accurate local comparisons of the measures μ and λ. Observe that $X_n \in L^1(\lambda)$ and that for all $A \in \mathcal{F}_n$,

$$\int_A X_{n+1} d\lambda = \mu(A) \equiv \int_A X_n d\lambda \,, \tag{5.1.7}$$

which is (5.1.4). This shows that $(X_n)_{n\geq 1}$ is a martingale with respect to the dyadic filtration. By writing

$$\mu(A) = \int_A X_n d\lambda \quad \forall A \in \mathcal{F}_n \,, \tag{5.1.8}$$

we see that X_n is a good candidate for the construction of a density of μ with respect to λ. For (5.1.8) to hold for all $A \in \mathcal{B}([0,1))$, we see that a limiting procedure $n \to \infty$ is necessary.

Example 5.1.5. Consider the Branching Process of Example 4.1.8:

$$X_{n+1} = \sum_{k=1}^{X_n} Y_k^{(n)} \,,$$

where the $Y_k^{(n)}$ are i.i.d. with distribution $P(Y_k^{(n)} = j) = \rho(j)$, $j \geq 0$. Assume $\lambda := E[Y_k^{(n)}] = \sum_{j\geq 0} j\rho(j) < \infty$. Consider the natural filtration

$$\mathcal{F}_n := \sigma(Y_i^{(k)}, i \geq 1, k \leq n) \,.$$

We have $X_n \in L^1$ since

$$E[X_{n+1}] = \sum_{k\geq 1} E[1_{\{k\leq X_n\}} Y_k^{(n)}] = \sum_{k\geq 1} E[1_{\{k\leq X_n\}}] E[Y_k^{(n)}] = \lambda E[X_n] = \lambda^n < \infty \,.$$

Similarly,

$$E[X_{n+1}|\mathcal{F}_n] = \sum_{k\geq 1} 1_{\{k\leq X_n\}} E[Y_k^{(n)}|\mathcal{F}_n] = \lambda \sum_{k\geq 1} 1_{\{k\leq X_n\}} = \lambda X_n \,,$$

we see that $Z_n := \lambda^{-n} X_n$ is a non-negative martingale.

5.2. Martingale Transforms

If X_n is a martingale, what about $|X_n|$? More generally, what about $\phi(X_n)$? We call $(\phi(X_n))$ a transformation of (X_n). For example, convex functions transform martingales into submartingales:

Lemma 5.2.1. *Let $(X_n)_{n\geq 1}$ be a martingale (resp. submartingale), and let $\phi = \phi(x)$ be a convex (resp. convex and non-decreasing) function* *such that $\phi \circ X_n \in L^1$ for all $n \geq 1$. Then $(\phi(X_n))_{n\geq 1}$ is a submartingale. In particular, $(X_n^+)_{n\geq 1}$, $(X_n^-)_{n\geq 1}$ and $(|X_n|)_{n\geq 1}$ are submartingales, and if $X_n \in L^2$, then $(X_n^2)_{n\geq 1}$ is a submartingale.*

PROOF. This follows at once from the conditional version of Jensen's Inequality: if $(X_n)_{n\geq 1}$ is a martingale and ϕ is convex, then

$$\phi(X_n) = \phi(E[X_{n+1}|\mathfrak{F}_n]) \leq E[\phi(X_{n+1})|\mathfrak{F}_n] \,.$$

If ϕ is non-decreasing and $(X_n)_{n\geq 1}$ is a submartingale, then

$$\phi(X_n) \leq \phi(E[X_{n+1}|\mathfrak{F}_n]) \leq E[\phi(X_{n+1})|\mathfrak{F}_n] \,.$$

The final statement is a consequence of the first. □

Transformations that will play a fundamental role later are those which measure in a certain sense the increments of a martingale, i.e. $X_n - X_{n-1}$. See (5.2.1) hereafter.

DEFINITION 5.2.1. *Let $(\mathfrak{F}_n)_{n\geq 0}$ (observe that $n \geq 0$) be a filtration. A sequence of random variables $(C_n)_{n\geq 1}$ is called* predictable *if C_n is $\mathfrak{F}_{n-1}$-measurable for all $n \geq 1$.*

A typical example of predictable sequence is given in gambling terms: C_n is the amount of money the gambler bets on the game at time n, which is of course based on the knowledge of all the games up to time $n-1$. If $(X_n)_{n\geq 0}$ is adapted and $(C_n)_{n\geq 1}$ is predictable, we define the transformation of $(X_n)_{n\geq 0}$ by $(C_n)_{n\geq 1}$ as the sequence $((C\cdot X)_n)_{n\geq 1}$, where

$$(C \cdot X)_n := \sum_{k=1}^{n} C_k(X_k - X_{k-1}) \,. \tag{5.2.1}$$

LEMMA 5.2.2. *Let $(C_n)_{n\geq 1}$ be predictable and bounded.*

(1) *If $(X_n)_{n\geq 0}$ is a martingale, then $(C \cdot X)_n$ is a martingale.*
(2) *If $(X_n)_{n\geq 0}$ is a sub/super-martingale and if $C_n \geq 0$, then $(C \cdot X)_n$ is a sub/super-martingale.*

PROOF. Let $Z_n = (C \cdot X)_n$. Clearly, Z_n is adapted to $\mathfrak{F}_n$, and since C_n is bounded, $Z_n \in L^1$. Then, since (C_n) is predictable,

$$\begin{aligned} E[Z_{n+1}|\mathfrak{F}_n] &= Z_n + E[C_{n+1}(X_{n+1} - X_n)|\mathfrak{F}_n] \\ &= Z_n + C_{n+1}E[(X_{n+1} - X_n)|\mathfrak{F}_n] \,. \end{aligned}$$ □

Transformations by predictable sequences will play a crucial role in the following section.

5.3. Doob's Optional Stopping Theorem

We have seen in (5.1.5) that the expectation of a martingale X_n, $E[X_n]$, does not depend on time: $E[X_n] = E[X_1]$ for all $n \geq 1$. A deep property of martingales is that this remains true when the time n becomes a random variable. This property is the key to all convergence results that will be derived in further sections, and turns martingales a powerful tool in many applications.

Modeling a random time is done using the notion of *stopping time*, which we already encountered in the study of Markov chains. Here we define them with respect to an arbitrary filtration $(\mathcal{F}_n)_{n\geq 1}$, which we fix throughout the section.

Definition 5.3.1. *Let $(\mathcal{F}_n)_{n\geq 1}$ be a filtration. A $\{1, 2, \dots\} \cup \{\infty\}$-valued random variable T is a* **stopping time with respect to** *$(\mathcal{F}_n)_{n\geq 1}$ if*

$$\{T \leq n\} \in \mathcal{F}_n\,, \quad \forall n \geq 1\,. \tag{5.3.1}$$

Observe that the defining condition (5.3.1) can be replaced by

$$\{T = n\} \in \mathcal{F}_n\,, \quad \forall n \geq 1\,. \tag{5.3.2}$$

We also have $\{T < \infty\} = \bigcup_n \{T = n\} \in \mathcal{F}$ and so $\{T = \infty\} = \{T < \infty\}^c \in \mathcal{F}$.

Consider the random walk of Example 5.1.1. Let I be any subset of $\mathbb{Z}$. Then $T_I := \inf\{n \geq 1 : X_n \in I\}$, the **first visit at** I, is a stopping time with respect to the natural filtration. Observe that the **last visit at** I, $\sup\{n \geq 1 : X_n \in I\}$ is not a stopping time. The proof of the following lemma is left as an exercise.

Lemma 5.3.1. *If T is a stopping time and if $k \geq 1$ is any integer then $T \wedge k$ is a stopping time. If T_1, T_2 are two stopping times, then $T_1 + T_2$, $T_1 \wedge T_2$, $T_1 \vee T_2$ are stopping times.*

For each stopping time T, define the collection

$$\mathcal{F}_T := \big\{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n\,, \forall\, n \geq 1\big\}\,. \tag{5.3.3}$$

Lemma 5.3.2. *Let T be a stopping time.*

(1) *$\mathcal{F}_T \subset \mathcal{F}$ is a σ-algebra called the* **stopped σ-field** *generated by T,*
(2) *T is $\mathcal{F}_T$-measurable,*
(3) *If $T_1 \leq T_2$, then $\mathcal{F}_{T_1} \subset \mathcal{F}_{T_2}$,*

Proof. Clearly, $\emptyset$ and Ω belong to $\mathcal{F}_T$, and if $A \in \mathcal{F}_T$, then $A^c \cap \{T \leq n\} = \{T \leq n\} \cap (A \cap \{T \leq n\})^c \in \mathcal{F}_n$, i.e. $A^c \in \mathcal{F}_T$. The stability under countable unions is immediate. This shows that $\mathcal{F}_T$ is a σ-algebra. It is trivial to verify that T is $\mathcal{F}_T$-measurable. Finally, assume $T_1 \leq T_2$ and take $A \in \mathcal{F}_{T_1}$. We have, for all $n \geq 1$,

$$A \cap \{T_2 \leq n\} = \bigcup_{k=0}^{n} A \cap \{T_2 = k\} = \bigcup_{k=0}^{n} \bigcup_{j=1}^{k} \underbrace{A \cap \{T_1 \leq j\}}_{\in \mathcal{F}_j} \cap \underbrace{\{T_2 = k\}}_{\in \mathcal{F}_k} \in \mathcal{F}_n\,,$$

since $\mathcal{F}_j, \mathcal{F}_k \subset \mathcal{F}_n$. Therefore $A \in \mathcal{F}_{T_2}$. □

When considering the observation of a random sequence $X_1, X_2, \dots$ at some random time T, the output is X_T, where $X_T : \Omega \to \mathbb{R}$ is

$$X_T(\omega) := \begin{cases} X_n(\omega) & \text{if } T(\omega) = n\,, \\ 0 & \text{if } T(\omega) = \infty\,, \end{cases} \tag{5.3.4}$$

which can also be written

$$X_T(\omega) = \sum_{n \geq 1} X_n 1_{\{T=n\}}(\omega)\,.$$

In this form, it is clear that X_T is a random variable. Then, for each $x \in \mathbb{R}$,

$$\{X_T \leq x\} \cap \{T \leq n\} = \bigcup_{j=1}^{n} \{X_j \leq x\} \cap \{T = j\} \in \mathcal{F}_n\,,$$

since $\{X_j \leq x\} \in \mathcal{F}_j$ and $\{T = j\} \in \mathcal{F}_j$. Since this holds for all n, this shows that $\{X_T \leq x\} \in \mathcal{F}_T$: X_T is $\mathcal{F}_T$-measurable.

Here is a first concrete use of stopping times, which doesn't yet involve the concept of martingale. If $X_1, X_2, \dots$ are i.i.d. and integrable, $S_n := X_1 + \cdots + X_n$, then, by linearity, $E[S_n] = nE[X_1]$. The following gives a generalization of this fact to the situation where n is changed into a stopping time T.

Theorem 5.3.1 (Wald's Identity). *Let $X_1, X_2, \dots$ be i.i.d. with $X_1 \in L^1$, $S_n := X_1 + \cdots + X_n$. Let T be a stopping time with respect to the natural filtration $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, integrable: $T \in L^1$. Then $E[S_T] = E[T]E[X_1]$.*

Proof. Since $T \in L^1$ it is a.s. finite. We first consider the case $X_k \geq 0$:

$$E[S_T] = \sum_{k\geq 1} E[S_k 1_{\{T=k\}}] = \sum_{k\geq 1}\sum_{j=1}^{k} E[X_j 1_{\{T=k\}}] = \sum_{j\geq 1} E[X_j 1_{\{T\geq j\}}] \, .$$

Since $\{T \geq j\} = \{T \leq j-1\}^c \in \mathcal{F}_{j-1}$ and since X_j is independent of $\mathcal{F}_{j-1}$, we have $E[X_j 1_{\{T\geq j\}}] = E[X_j]E[1_{\{T\geq j\}}]$. The result follows since $E[X_j] = E[X_1]$ and $\sum_{j\geq 1} E[1_{\{T\geq j\}}] = E[T]$. In the general case, we must justify the interchange of the sums over j and k. Observe that

$$\infty > E[|X_1|]E[T] = \sum_{j\geq 1} E[|X_j|]P(T \geq j) = \sum_{j\geq 1}\sum_{k\geq j} E[|X_j|]P(T = k) \, .$$

By Fubini's Theorem, the change of order of summation is therefore justified. □

Wald's Identity has the following interesting consequence. Let (S_n) denote the simple symmetric random walk starting at the origin. Let $T := \inf\{n \geq 1 : S_n = 1\}$. By recurrence of the walk, $P(T < \infty) = 1$. Since $S_T = 1$ on $\{T < \infty\}$, we have $P(S_T = 1) = 1$, i.e. $E[S_T] = 1$. Since the walk is symmetric, $E[X_1] = 0$ for all n. If $T \in L^1$, the previous theorem would imply that $E[S_T] = 0$, a contradiction. Therefore, $E[T_x] = \infty$.

Let us move to the major tool of Martingale Theory, which says that the property (5.1.5) is preserved when the martingale is observed at random times. The central ingredient is the following lemma.

Lemma 5.3.3. *Let $(X_n)_{n\geq 0}$ be a martingale (resp. a supermartingale), T a stopping time. Then $(X_{n\wedge T})_{n\geq 0}$ is again a martingale (resp. supermartingale).*

Proof. For all $n \geq 1$, define $C_n := 1_{\{T\geq n\}}$. Since $\{T \geq n\} = \{T < n\}^c \in \mathcal{F}_{n-1}$, $(C_n)_{n\geq 1}$ is predictable. By explicit computation, $(C \cdot X)_n = X_{n\wedge T} - X_0$. Therefore,

$$X_{n\wedge T} = (C \cdot X)_n + X_0 \quad \forall n \geq 0 \, . \tag{5.3.5}$$

Since C_n is bounded, the result follows from Lemma 5.2.2. □

Theorem 5.3.2 (Doob's Optional Stopping Theorem). *Let $(X_n)_{n\geq 0}$ be an adapted sequence, T a stopping time. Consider the following conditions:*

(1) *T is bounded,*
(2) *T is a.s. finite and (X_n) is bounded,*
(3) *$T \in L^1$ and the increments $(X_n - X_{n-1})$ are bounded,*
(4) *T is a.s. finite and $X_n \geq 0$.*

If (X_n) is a supermartingale, and if either of the conditions (1)-(4) is satisfied, then $X_T \in L^1$ and $E[X_T] \leq E[X_0]$.
If (X_n) is a martingale, and if either of the conditions (1)-(3) is satisfied, then $X_T \in L^1$ and $E[X_T] = E[X_0]$.

Proof. We first consider the case of supermartingales. Under (1), there exists K such that $T \leq K$. Therefore, $X_T = X_{K\wedge T}$ and so, by Lemma 5.3.3, $X_T \in L^1$ and $E[X_T] = E[X_{K\wedge T}] \leq E[X_0]$. Assume then that (2) holds, i.e. $|X_n| \leq M$. Then

$$E[|X_T|] = \int_{T<\infty} |X_T| dP \leq M ,$$

and so $X_T \in L^1$. By Dominated Convergence and the supermartingale property ($\{T \leq n\} \in \mathcal{F}_n$),

$$E[X_T] = \int_{T<\infty} X_T dP = \lim_{n\to\infty} \int_{T\leq n} X_{n\wedge T} dP \leq \limsup_{n\to\infty} \int_{T\leq n} X_0 dP = E[X_0] .$$

Assume (3) holds. Then T is a.s. finite and $|X_n - X_{n-1}| \leq M$. Therefore,

$$|X_{n\wedge T}| \leq |X_0| + |X_{n\wedge T} - X_0| \leq |X_0| + \sum_{k=1}^{n\wedge T} |X_k - X_{k-1}| \leq |X_0| + MT .$$

This implies $X_{n\wedge T} \in L^1$. Since $\lim_n X_{n\wedge T} = X_T$ a.s., by the Lemma of Fatou,

$$\int |X_T| dP \leq \liminf_{n\to\infty} \int |X_{n\wedge T}| dP \leq E[|X_0|] + ME[T] < \infty$$

and so $X_T \in L^1$. Then, by Dominated Convergence,

$$E[X_T] = \lim_{n\to\infty} \int_{T\leq n} X_{n\wedge T} dP \leq \limsup_{n\to\infty} \int_{T\leq n} X_0 dP = E[X_0] .$$

Under (4), we use again the Lemma of Fatou:

$$E[|X_T|] = E[X_T] \leq \liminf_{n\to\infty} \int X_{n\wedge T} dP \leq \int_{T<\infty} X_0 dP = E[X_0] .$$

Now if (X_n) is a martingale then both (X_n) and $(-X_n)$ are supermartingales, and therefore, under either of the conditions (1)-(3), $E[X_T] \leq E[X_0]$ and $E[-X_T] \leq E[-X_0]$. This implies $E[X_T] = E[X_0]$. □

The Optional Stopping Theorem applies when the stopping time satisfies some finiteness condition, and the following example shows that the result can be wrong without such restrictions. Let again $(S_n)_{n\geq 0}$ denote the simple symmetric random walk starting at the origin. Consider the stopping time $T := \inf\{n \geq 1 : S_n = 1\}$. Then we saw that $E[S_T] = 1$, which is obviously different from $E[S_0] = 0$. This is due to the fact that T is not integrable.

The following lemma gives a useful criterium to verify the finiteness of T in concrete situations.

LEMMA 5.3.4. *Assume T is a stopping time with respect to $(\mathcal{F}_n)$ for which there exist an integer $N > 0$ and an $\epsilon > 0$ such that, almost surely,*

$$P(T \leq n + N|\mathcal{F}_n) \geq \epsilon \quad \forall n \geq 0\,. \tag{5.3.6}$$

Then $P(T > kN) \leq (1-\epsilon)^k$ for all $k \geq 1$. In particular, $T \in L^1$.

Condition (5.3.6) means that whatever happened up to a fixed time n, the probability of T occuring before the next N steps is always at least ϵ.

PROOF. If $k = 1$, $P(T \leq N) = E[P(T \leq N|\mathcal{F}_0)] \geq \epsilon$. So assume the result has been shown for k.

$$\begin{aligned} P(T > (k+1)N) &= E[1_{\{T>kN\}}1_{\{T>(k+1)N\}}] \\ &= E[E[1_{\{T>kN\}}1_{\{T>(k+1)N\}}|\mathcal{F}_{kN}]] \\ &= E[1_{\{T>kN\}}P(T > (k+1)N|\mathcal{F}_{kN})]\,. \end{aligned}$$

But $P(T > (k+1)N|\mathcal{F}_{kN}) = 1 - P(T \leq kN + N|\mathcal{F}_{kN}) \leq 1 - \epsilon$. This implies $P(T > (k+1)N) \leq (1-\epsilon)P(T > kN) \leq (1-\epsilon)^{k+1}$. □

We now present applications of the Stopping Theorem.

5.3.1. Application: the Second Heart Problem. In a deck of 52 cards, well shuffled, we turn the cards from the top until the first hearts appears. If we turn one more card, what is the probability that this card is hearts again? Answer: $\frac{1}{4}$.

We describe the experience by a finite sequence $X_0, X_1, \dots, X_{51}$, where X_k is the fraction of hearts left in the deck after the kth card was drawn: $X_0 = \frac{13}{52} = \frac{1}{4}$, and for $k = 1, \dots, 51$ (observe that the number of cards left after the kth card was dran is $52 - k$),

$$X_k = \frac{\#\{\text{hearts remaining in the deck after } k\text{th card was drawn}\}}{52 - k}\,.$$

Observe that since we assume the deck to be well shuffled, X_k is exactly the probability that a hearts is drawn at time $k+1$. We claim that $(X_k)_{k=0}^{51}$ is a martingale with respect to its natural filtration. Namely, by considering separately the cases in which the nth card drawn is a hearts or not,

$$\begin{aligned} E[X_n|\mathcal{F}_{n-1}] &= \frac{X_{n-1}(52-(n-1))-1}{52-n} X_{n-1} + \frac{X_{n-1}(52-(n-1))}{52-n}(1-X_{n-1}) \\ &= X_{n-1}\,. \end{aligned}$$

Define the stopping time T as the first time a hearts is drawn. The probability we are interested in is $E[X_T]$. Since T is bounded ($T \leq 52$) the Stopping Theorem gives: $E[X_T] = E[X_0] = \frac{1}{4}$.

5.3.2. Application: The Gambler's Ruin. This section rederives the results obtained in Exercise 2.9 in a completely different way. The martingales introduced hereafter were apparently introduced first by De Moivre. Consider i.i.d. random variables $Y_1, Y_2, \dots$ with $P(Y_k = +1) = 1 - P(Y_k = -1) = p$, $0 < p < 1$. We consider the simple random walk starting at some $x > 0$: $S_0 := x$, $S_n := x + Y_1 + \cdots + Y_n$, and denote its law by P_x. We fix some $N > x$ and study the probability that the walk reaches 0 before N, that is $P_x(S_T = 0)$. This can be done by introducing the first time at which the walk exits the interval $[1, N-1]$:

$$T := \inf\{n \geq 1 : S_n \in \{0, N\}\}\,.$$

First, we show that $T \in L^1$, which is intuitively true since for T to be very large the walk must stay in the middle of the box for a long period of time, a very rare event. Wherever the walk is at time n, the probability that it exits the interval $[1, N-1]$ during the next N steps is bounded below by the probability that it exits the interval through N, which is itself bounded below by $p^N \equiv \epsilon > 0$. By Lemma 5.3.4, $T \in L^1$.

When $p = \frac{1}{2}$, S_n is a martingale, and $|S_n - S_{n-1}| = 1$. By the Stopping Theorem, $E_x(S_T) = E_x(S_0) = x$. Since $S_T \in \{0, N\}$, we have $E_x(S_T) = 0 \cdot P_x(S_T = 0) + N \cdot P_x(S_T = N)$, and get $P_x(S_T = N) = \frac{x}{N}$, i.e. $P_x(S_T = 0) = 1 - \frac{x}{N}$.

When $p \neq \frac{1}{2}$, S_n we need to find another martingale which "cancels" the asymmetry between p and $1-p$. The following is left as an exercise:

Lemma 5.3.5. *$N_n = S_n - (p-q)n$ and $M_n = \left(\frac{q}{p}\right)^{S_n}$ are martingales.*

As can be verified easily, $\sup_n |M_n - M_{n-1}| \leq \beta_N < \infty$, and therefore by Theorem 5.3.2, $E[M_T] = E[M_0]$. Since $M_T \in \{1, \left(\frac{q}{p}\right)^N\}$ almost surely, we easily get as before

$$P(S_T = 0) = \frac{\left(\frac{q}{p}\right)^x - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N},$$

as we had obtained in Exercise 2.9.

5.3.3. Application: first appearance in a random sequence. The following is a generalization of the problem known as "The first run of three sixes". Let $(X_n)_{n\geq 1}$ be an i.i.d. sequence taking values in the finite set $S = \{\pm 1\}$, with $P(X_k = +1) = p$, $P(X_k = -1) = q$, $p + q = 1$. We always assume that $0 < p < 1$. Denote by $\mathcal{F}_n$ the natural filtration $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$.

Fix a word of size l, $\mathsf{a} = (a_1, a_2, \ldots, a_l)$, with $a_i \in S$. Let T be the first time the word a appears in the sequence $(X_n)_{n\geq 1}$, that is

$$T_{\mathsf{a}} := \inf\{n \geq l : (X_{n-l+1}, \ldots, X_n) = \mathsf{a}\}.$$

Clearly, T_{a} is a stopping time with respect to the natural filtration, and it is easy to show using Lemma 5.3.4 that $T_{\mathsf{a}} \in L^1$. We are interested in computing $E[T_{\mathsf{a}}]$. For example, we will show that for the word $\mathsf{a} = (-1, \ldots, -1)$ (l times the symbol -1),

$$E[T_{\mathsf{a}}] = q^{-1} + q^{-2} + \cdots + q^{-l} \tag{5.3.7}$$

To study this problem, we construct martingales and use the Stopping Theorem. We introduce an auxiliary process as follows. Consider the numbers

$$K_+ := p^{-1}, \quad K_- := q^{-1},$$

which will be rates associated to the appearance of $+1$ or -1 in the sequence X_n. Assume first that a is an infinite sequence $(a_1, a_2, \dots)$. Just before each time n, a gambler enters the game with an initial having of $H_0^{(n)} := 1$ (Brazilian reais, say), and bets $H_0^{(n)}$ reais on the event that X_n will be a_1. If he loses, that is if $X_n \neq a_1$, his new havings become $H_1^{(n)} := 0$, and he leaves the game. If he wins, i.e. if $X_n = a_1$, he wins $K_+ H_0^{(n)}$ if $a_1 = +1$, and $K_- H_0^{(n)}$ if $a_1 = -1$. His havings thus become $H_1^{(n)} := K_{a_1} H_0^{(n)}$. At the next step, just before $n+1$, he bets $H_1^{(n)}$ on the event that X_{n+1} will be a_2. If he loses then $H_2^{(n)} := 0$ (and he quits the game) and if he wins, his new having is $H_2^{(n)} := K_{a_2} H_1^{(n)}$, etc. After k steps,

$$H_{k+1}^{(n)} = \begin{cases} 0 & \text{if } X_{n+k} \neq a_{k+1}\,, \\ K_{a_{k+1}} H_k^{(n)} & \text{if } X_{n+k} = a_{k+1}\,. \end{cases}$$

Observe that whether a_{k+1} is $+1$ or -1, the havings of the player remain equal to 1 on average.

For $n \geq 1$, let Z_n denote the total havings of all players that entered the game up to time n:

$$Z_n := H_n^{(1)} + H_{n-1}^{(2)} + \cdots + H_2^{(n-1)} + H_1^{(n)} \equiv \sum_{j=1}^{n} H_{n-j+1}^{(j)}\,.$$

Define also $Z_0 := 0$. Since the havings of each player remain constant on average, and since at time n exactly n reais have been invested by the players in the game, we expect that the sequence $(M_n)_{n\geq 0}$ defined by $M_n := Z_n - n$ forms a martingale with respect to the natural filtration $\mathcal{F}_n$. Indeed, since for all $j = 1, \dots, n$,

$$E[H_{n+1-j+1}^{(j)} | \mathcal{F}_n] = E[K_{X_{n+1}} H_{n-j+1}^{(j)} | \mathcal{F}_n] = H_{n-j+1}^{(j)} E[K_{X_{n+1}} | \mathcal{F}_n] = H_{n-j+1}^{(j)}\,,$$

and since $E[H_1^{(n+1)} | \mathcal{F}_n] = H_0^{(n+1)} \equiv 1$, we get

$$\begin{aligned} E[Z_{n+1} | \mathcal{F}_n] &= \sum_{j=1}^{n} E[H_{n+1-j+1}^{(j)} | \mathcal{F}_n] + E[H_1^{(n+1)} | \mathcal{F}_n] \\ &= \sum_{j=1}^{n} H_{n-j+1}^{(j)} + 1 \equiv Z_n + 1\,. \end{aligned}$$

This shows that M_n is a martingale.

If $T \in L^1$ is a stopping time with respect to the filtration $\mathcal{F}_n$, we can then stop M_n at T. The Stopping Theorem gives $E[M_T] = E[M_0] = 0$, which implies

$$E[T] = E[Z_T]\,. \tag{5.3.8}$$

Consider for example the case $\mathsf{a} = (-1, -1, \dots, -1)$ (l times the symbol -1), then at time T_{a} the player who entered at time $T_{\mathsf{a}} - l$ lost his first bet and quit the game immediately, but the player who entered at time $T_{\mathsf{a}} - l + 1$ won all his bets up to time T_{a}, and his fortune is thus exactly K_-^l. The fortune of the player who entered at time $T_{\mathsf{a}} - l + 2$ is exactly K_-^{l-1}, etc., until the player who entered just before time T_{a}, whose havings equal K_-. We then have $Z_{T_{\mathsf{a}}} = K_-^l + K_-^{l-1} + \cdots + K_-$. By (5.3.8), this proves (5.3.7).

5.3.4. Application: The Secretary Problem. N candidates present themselves for a job interview. According to the employer's criteria, the ith candidate's suitability for the job is a number between 0 and 1. At time i, the employer interviews candidate i and determines its suitability exactly. Immediately after the interview, he must decide whether to accept or reject the candidate, since no recall is possible. What strategy should the employer adopt in order to maximize his chance of choosing a good candidate?

In probabilistic terms, the suitabilities of the candidates can be modelized by an i.i.d. sequence $X_1, \dots, X_N$ where each X_k has uniform distribution over $[0, 1]$. Finding the best strategy is to find a stopping time T which maximizes $E[X_T]$. This best stopping time is given hereafter.

THEOREM 5.3.3. *Consider the sequence $\alpha_1, \dots, \alpha_N$ defined by $\alpha_N := 0$, and for $k = N, \dots, 1$, $\alpha_{k-1} := \frac{1}{2} + \frac{\alpha_k^2}{2}$. Let $T^* := \inf\{n \geq 1 : X_n > \alpha_n\}$. Then*

$$E[X_T] \leq E[X_{T^*}] \tag{5.3.9}$$

for any stopping time T.

Observe that α_1 depends on N, and that α_k is decreasing. As an example, in the case $N = 5$, we have approximately:

k	1	2	3	4	5
α_k	0.742	0.695	0.625	0.500	0

INCLUIR O ARGUMENTO INTUITIVO DO GUGU.

Proof of Theorem 5.3.3: By a simple computation we have, for all $0 \leq \alpha \leq 1$,

$$E[X_n \vee \alpha] = \frac{1}{2} + \frac{\alpha^2}{2} \quad \forall n = 1, \ldots, N \,. \tag{5.3.10}$$

For each stopping time T (with respect to the natural filtration associated to $(X_k)_{k=1}^N$), we define a sequence $(Y_n^T)_{n=0}^N$, where $Y_0 := \alpha_0$, and

$$Y_n^T := X_{n \wedge T} \vee \alpha_n \,.$$

Claim 1: $(Y_n^T)_{n=0}^N$ is a supermartingale. On $\{T \leq n-1\}$,

$$E[Y_n^T | \mathfrak{F}_{n-1}] = E[X_T \vee \alpha_n | \mathfrak{F}_{n-1}] = X_T \vee \alpha_n \leq X_T \vee \alpha_{n-1} = Y_{n-1}^T \,,$$

and on $\{T > n-1\}$,

$$E[Y_n^T | \mathfrak{F}_{n-1}] = E[X_n \vee \alpha_n | \mathfrak{F}_{n-1}] = E[X_n \vee \alpha_n] = \frac{1}{2} + \frac{\alpha_n^2}{2} = \alpha_{n-1} \leq Y_{n-1}^T \,.$$

Claim 2: $(Y_n^{T^*})_{n=0}^N$ is a martingale. On $\{T^* \leq n-1\}$,

$$E[Y_n^{T^*} | \mathfrak{F}_{n-1}] = E[X_{T^*} \vee \alpha_n | \mathfrak{F}_{n-1}] = X_{T^*} \vee \alpha_n$$

But by the definition of T^*, $X_{T^*} > \alpha_{T^*}$. Since α_k is decreasing, $\alpha_{T^*} \geq \alpha_{n-1} \geq \alpha_n$. This means that $E[Y_n^{T^*} | \mathfrak{F}_{n-1}] = X_{T^*} \vee \alpha_n = X_{T^*} = X_{T^*} \vee \alpha_{n-1} = Y_{n-1}^{T^*}$. Now on $\{T^* > n-1\}$, as before we get $E[Y_n^{T^*} | \mathfrak{F}_{n-1}] = \alpha_{n-1} = Y_{n-1}^{T^*}$.

Now for any stopping time T, Theorem 5.3.2 and Claim 1 give $E[Y_T] \leq E[Y_0]$. On the other hand, by Claim 2, $E[Y_{T^*}] = E[Y_0]$, which proves the theorem. □

5.3.5. More on Optional Stopping. We saw before that looking at a martingale at a random time preserves expectations: $E[X_T] = E[X_0]$. We now show a stronger result stating thaf if a martingale is considered at random times $T_1 \leq T_2 \leq \ldots$, then the sequence $(X_{T_k})_{k \geq 1}$ is again a martingale. This implies, in particular, that $E[X_{T_k}] = E[X_0]$. We first consider a particular case, that of closed martingales. Again, $(\mathfrak{F}_n)_{n \geq 1}$ is a fixed filtration.

Theorem 5.3.4 (General Optional Stopping for Closed Martingales). *Let $(X_n)_{n \geq 1}$ be a closed martingale, i.e. $X_n = E[X | \mathfrak{F}_n]$ for some integrable random variable X. Let $T_1 \leq T_2 \leq \ldots$ be a sequence of a.s. finite stopping times. Then*

$$X_{T_k} = E[X | \mathfrak{F}_{T_k}] \,. \tag{5.3.11}$$

That is, $(X_{T_k})_{k\geq 1}$ is again a closed martingale.

Proof. By a previous lemma, $(X_{T_k})_{k\geq 1}$ is adapted to $(\mathcal{F}_{T_k})_{k\geq 1}$. To verify (5.3.11), fix $k \geq 1$ and denote for simplicity $T := T_k$. Take $A \in \mathcal{F}_T$. Since T is a.s. finite, we can write

$$\int_A X_T \, dP = \sum_{n\geq 0} \int_{A\cap\{T=n\}} X_T \, dP = \sum_{n\geq 0} \int_{A\cap\{T=n\}} E[X|\mathcal{F}_n] \, dP \, .$$

Since $A \cap \{T = n\} = (A \cap \{T \leq n\}) \cap (A \cap \{T \leq n-1\})^c \in \mathcal{F}_n$, we have

$$\sum_{n\geq 0} \int_{A\cap\{T=n\}} E[X|\mathcal{F}_n] \, dP = \sum_{n\geq 0} \int_{A\cap\{T=n\}} X \, dP = \int_A X \, dP \, ,$$

which shows that $X_T = E[X|\mathcal{F}_T]$, i.e. (5.3.11). Then each X_{T_k} is integrable since

$$E[|X_{T_k}|] = E[|E[X|\mathcal{F}_{T_k}]|] \leq E[E[|X||\mathcal{F}_T]] = E[|X|] < \infty \, , \quad (5.3.12)$$

and the martingale property is verified since

$$E[X_{T_{k+1}}|\mathcal{F}_{T_k}] = E[E[X|\mathcal{F}_{T_{k+1}}]|\mathcal{F}_{T_k}] = E[X|\mathcal{F}_{T_k}] = X_{T_k} \, .$$

□

We then turn to the general case. Observe that the difference with the preceding theorem is that the stopping times must now be bounded.

Theorem 5.3.5 (General Optional Stopping Theorem). *Let $(X_n)_{n\geq 1}$ be a martingale (resp. submartingale) and let $T_1 \leq T_2 \leq \dots$ be a sequence of bounded stopping times. Then $(X_{T_k})_{k\geq 1}$ is again a martingale (resp. submartingale). In particular, (in the case of submartingale, replace everywhere $=$ by $\geq$)*

$$E[X_{T_{k+1}}]=E[X_{T_k}] = E[X_1] \quad (5.3.13)$$

Proof. For simplicity, consider the case $k = 1$. To verify that $E[X_{T_2}|\mathcal{F}_{T_1}] = X_{T_1}$, i.e. that

$$\int_A X_{T_2} \, dP = \int_A X_{T_1} \, dP \, , \quad \forall A \in \mathcal{F}_{T_1} \, , \quad (5.3.14)$$

it is sufficient to verify, since T_1 is a.s. finite, that

$$\int_{B_j} X_{T_2} \, dP = \int_{B_j} X_{T_1} \, dP \, , \quad \forall A \in \mathcal{F}_{T_1} \, , \quad (5.3.15)$$

where $B_j := A \cap \{T_1 = j\}$. Then, (5.3.14) follows by summing over $j \geq 1$. To obtain (5.3.15) we shall prove the validity of the following expression for all $k \geq j$:

$$\int_{B_j \cap \{T_2 = k\}} X_{T_2}\, dP = \int_{B_j \cap \{T_2 \geq k\}} X_k\, dP - \int_{B_j \cap \{T_2 \geq k+1\}} X_{k+1}\, dP\,. \tag{5.3.16}$$

Namely from it follows, by summing (5.3.16) over $k = j, \ldots, N$, for some sufficiently large N (larger than $\max\{\max T_1, \max T_2\}$), that

$$\int_{B_j \cap \{T_1 \geq j\}} X_{T_2}\, dP = \int_{B_j \cap \{T_1 \geq j\}} X_{T_1}\, dP\,, \quad \forall A \in \mathfrak{F}_{T_1}\,, \tag{5.3.17}$$

which is exactly (5.3.15) since $T_1 \leq T_2$ implies

$$B_j \cap \{T_1 \geq j\} = A \cap \{T_1 = j\} \cap \{T_1 \geq j\} = A \cap \{T_1 = j\} = B_j\,.$$

To show (5.3.16), first write

$$\int_{B_j \cap \{T_2 \geq k\}} X_k\, dP = \int_{B_j \cap \{T_2 = k\}} X_k\, dP + \int_{B_j \cap \{T_2 \geq k+1\}} X_k\, dP\,.$$

Obviously, the first term equals

$$\int_{B_j \cap \{T_2 = k\}} X_k\, dP = \int_{B_j \cap \{T_2 = k\}} X_{T_2}\, dP\,.$$

For the second, first use the fact that $(X_k, \mathfrak{F}_k)_{k \geq 1}$ is a martingale:

$$\int_{B_j \cap \{T_2 \geq k+1\}} X_k\, dP = \int_{B_j \cap \{T_2 \geq k+1\}} E[X_{k+1}|\mathfrak{F}_k]\, dP\,.$$

Now note that $B_j \in \mathfrak{F}_j$ and $\{T_2 \geq k+1\} = \{T_2 \leq k\}^c \in \mathfrak{F}_k$. Therefore, $B_j \cap \{T_2 \geq k+1\} \in \mathfrak{F}_k$. This yields

$$\int_{B_j \cap \{T_2 \geq k+1\}} E[X_{k+1}|\mathfrak{F}_k]\, dP = \int_{B_j \cap \{T_2 \geq k+1\}} X_{k+1}\, dP\,.$$

We have thus proved (5.3.16). To show (5.3.13) let $N \geq \sup_\omega T_k(\omega)$ and compute

$$\begin{aligned} E[X_{T_k}] = \sum_{j=1}^{N} \int_{\{T_k = j\}} X_{T_k}\, dP &= \sum_{j=1}^{N} \int_{\{T_k = j\}} X_j\, dP \\ &= \sum_{j=1}^{N} \int_{\{T_k = j\}} X_N\, dP = E[X_N] = E[X_1]\,, \end{aligned}$$

where we have used two times the fact that $(X_n)_{n\geq 1}$ is a martingale. □

5.4. The Doob-Kolmogorov Inequality and L^2-martingales

In this section we prove the *Doob-Kolmogorov Inequality*, which will give convergence of L^2-martingales. (In the following section, the *Upcrossing Inequality* will imply convergence of L^1-martingales.)

THEOREM 5.4.1 (Doob-Kolmogorov Inequality). *Let $(S_n, \mathcal{F}_n)_{n\geq 1}$ be a martingale, $\lambda > 0$. Then for all $n \geq 1$,*

$$P\big(\max_{1\leq k\leq n} |S_k| \geq \lambda\big) \leq \frac{E[S_n^2]}{\lambda^2}\,. \tag{5.4.1}$$

PROOF. The proof is analogous to what was done to obtain (2.4.9) Define $A_0 := \Omega$, $A_k := \{|S_j| < \lambda, j = 1, \dots, k\}$. Let $B_k := A_{k-1} \cap \{|S_k| \geq \lambda\}$ be the event in which $|S_j|$ passes over λ for the first time. We have $A_k \cup B_1 \cup \cdots \cup B_k = \Omega$. Therefore,

$$E[S_n^2] \geq \sum_{k=1}^n \int_{B_k} S_n^2 dP\,.$$

Writing, for each k, $S_n^2 = (S_n - S_k)^2 + 2S_k(S_n - S_k) + S_k^2$, we have

$$\int_{B_k} S_n^2 dP \geq 2\int_{B_k} (S_n - S_k) S_k dP + \int_{B_k} S_k^2 dP$$

For the first term, since $B_k \in \mathcal{F}_k$,

$$\begin{aligned}\int_{B_k} (S_n - S_k) S_k dP &= \int_{B_k} E[(S_n - S_k)S_k|\mathcal{F}_k] dP \\ &= \int_{B_k} S_k E[S_n - S_k|\mathcal{F}_k] dP = 0\end{aligned}$$

since $E[S_n - S_k|\mathcal{F}_k] = 0$ by the martingale property. Then observe that on B_k, $|S_k| \geq \lambda$ and so the second term equals

$$\int_{B_k} S_k^2 dP \geq \lambda^2 P(B_k).$$

Therefore, $E[S_n^2] \geq \lambda^2 \sum_{k=1}^n P(B_k) = \lambda^2 P\big(\max_{1\leq k\leq n} |S_k| \geq \lambda\big)$. □

The Doob-Kolmogorov suggests that some control on the convergence of S_n might be obtained by imposing that $E[S_n^2]$ be bounded uniformly in n.

5.4.1. Convergence of L^2-Martingales. In this section we study martingales which have the property of being L^2-bounded, i.e. $\sup_n E[|S_n|^2] < \infty$. For these, convergence follows from the Doob-Kolmogorov Inequality. First,

LEMMA 5.4.1. *A martingale $(X_n)_{n\geq 0}$ is bounded in L^2 if and only if*

$$\sum_{k\geq 1} E[(X_k - X_{k-1})^2] < \infty\,. \tag{5.4.2}$$

PROOF. First, write $X_n = X_0 + \sum_{k=1}^n (X_k - X_{k-1})$ and expand:

$$X_n^2 = X_0^2 + 2\sum_{k=1}^n X_0(X_k - X_{k-1}) + 2\sum_{1\leq i<k\leq n}(X_i - X_{i-1})(X_k - X_{k-1}) + \sum_{k=1}^n (X_k - X_{k-1})^2$$

Now, by the martingale property, all the mixed terms vanish:

$$E[X_0(X_k - X_{k-1})] = E[X_0 E[(X_k - X_{k-1})|\mathfrak{F}_{k-1}]] = 0\,.$$

In the same way, $E[(X_i - X_{i-1})(X_k - X_{k-1})] = 0$. □

THEOREM 5.4.2. *Let $(S_n, \mathfrak{F}_n)_{n\geq 1}$ be a martingale bounded in L^2. Then there exists a random variable S such that*

$$S_n \to S \quad \textit{almost surely and in } L^2\,. \tag{5.4.3}$$

PROOF. Observe that $S_n(\omega)$ converges if and only if $\omega \in A$, where

$$A := \bigcap_{l\geq 1}\bigcup_{m\geq 1}\bigcap_{j\geq 1}\{|S_{m+j} - S_m| \leq l^{-1}\}\,.$$

We show that $P(A^c) = 0$ or, which is equivalent, that $\lim_{m\to\infty} P(A_l(m)) = 0$ for all $l \geq 1$, where

$$A_l(m) := \bigcup_{j\geq 1}\{|S_{m+j} - S_m| > \epsilon_l\}\,,$$

and $\epsilon_l := l^{-1}$. Observe that for a given $m \geq 1$, $P(A_l(m)) = \lim_n P\big(\max_{1\leq k\leq n}|Y_k| \geq \epsilon_l\big)$, where $Y_j := S_{m+j} - S_m$. It is easy to see that $(Y_k)_{k\geq 1}$ is a martingale: for all $B \in \mathfrak{F}_k$,

$$\begin{aligned}\int_B Y_{k+1}dP &= \int_B E[Y_{k+1}|\mathfrak{F}_{m+k}]dP \\ &= \int_B E[S_{m+k+1}|\mathfrak{F}_{m+k}]dP - \int_B E[S_m|\mathfrak{F}_{m+k}]dP \\ &= \int_B S_{m+k}dP - \int_B S_m dP = \int_B Y_k dP\,.\end{aligned}$$

We can thus apply the Doob-Kolmogorov Inequality to Y_k:

$$P\big(\max_{1\le k\le n}|Y_k|\geq \epsilon_l\big)\leq \frac{E[Y_n^2]}{{\epsilon_l}^2}=\frac{E[(S_{m+n}-S_m)^2]}{{\epsilon_l}^2}=\frac{E[S_{m+n}^2]-E[S_m^2]}{{\epsilon_l}^2}, \tag{5.4.4}$$

where we used again the martingale property of S_n:

$$\begin{aligned}E[S_{m+n}^2]&=E[S_m^2]+E[(S_{m+n}-S_m)^2]+2E[S_m(S_{m+n}-S_m)]\\&=E[S_m^2]+E[(S_{m+n}-S_m)^2]+2E[S_mE[S_{m+n}-S_m|\mathcal{F}_m]]\\&=E[S_m^2]+E[(S_{m+n}-S_m)^2].\end{aligned}$$

This also shows that the sequence $E[S_n^2]$ is non-decreasing in n. Since it is at the same time bounded in n, define $M:=\lim_n E[S_n^2]$. Taking $n\to\infty$ in (5.4.4), we get

$$P(A_l(m))\leq\frac{M-E[S_m^2]}{{\epsilon_l}^2},$$

and therefore $\lim_m P(A_l(m))=0$, which shows that $S:=\lim_n S_n$ exists almost surely. To show that the convergence is also in L^2, use Fatou:

$$E[(S-S_n)^2]\leq\liminf_{m\to\infty}E[(S_{n+m}-S_n)^2]=\liminf_{m\to\infty}E[S_{n+m}^2]-E[S_n^2]=M-E[S_n^2],$$

which goes to zero when $n\to 0$. □

As a corollary, we have a convergence result for random series.

THEOREM 5.4.3. *Let $X_1,X_2,\dots$ be a sequence of independent random variables with $E[X_k]=0$ and $\mathrm{var}X_k<\infty$ for all $k\geq 1$. If $\sum_k \mathrm{var}X_k<\infty$, then $\sum_k X_k$ converges a.s.*

PROOF. We consider the natural filtration associated to (X_k), and let $M_n:=X_1+\cdots+X_n$. Since $E[X_k]=0$, (M_n) is a martingale. We also have that $E[(M_k-M_{k-1})^2]=E[X_k^2]=\mathrm{var}X_k$. If $\sum_k\mathrm{var}X_k<\infty$, then (M_n) is bounded in L^2 by Lemma 5.4.1 and by Theorem 5.4.2, $\lim_n M_n$ exists almost surely. □

As an example, consider the random harmonic series $\sum_n\frac{\epsilon_n}{n}$, where the $\epsilon_n=\pm1$ independently with probability $\frac{1}{2}$. Setting $X_n:=\frac{\epsilon_n}{n}$ we have $\mathrm{var}X_n=\frac{1}{n^2}$, and therefore $\sum_n\frac{\epsilon_n}{n}$ converges almost surely.

5.4.2. Application to Markov Chains. As another application of Theorem 5.4.2, let $(X_n)_{n\geq 0}$ be a Markov chain with state space S an transition matrix Q. Assume that the chain is irreducible and recurrrent. It can be shown (Exercise 5.7) that if $\psi : S \to \mathbb{R}$ is harmonic , i.e. $Q\psi = \psi$, then $(\psi(X_n))_{n\geq 0}$ is a martingale. If we assume moreover that ψ is bounded, then $(\psi(X_n))_{n\geq 0}$ is bounded in L^2, and by the previous theorem, $\lim_n \psi(X_n)$ exists almost surely. Since the chain is recurrent, $P(X_n = x \text{ i.o.}) = 1$ for all $x \in S$. But, since $\{X_n = x\} \subset \{\psi(X_n) = \psi(x)\}$, this means that $P(\psi(X_n) = \psi(x) \text{ i.o.}) = 1$. Therefore, we must have

$$\lim_{n\to\infty} \psi(X_n) = \psi(x) \quad \forall x \in S\,,$$

which is possible only if ψ is constant. We have thus shown:

THEOREM 5.4.4. *Let $(X_n)_{n\geq 0}$ be a Markov chain with state space S an transition matrix Q. If the chain is irreducible and recurrent, then the only bounded harmonic functions $\psi : S \to \mathbb{R}$ are the constants.*

5.5. Upcrossings and L^1-boundedness

We now take a closer look at the oscillations of the sequence $X_1(\omega), X_2(\omega), \ldots$ for a fixed $\omega \in \Omega$, where (X_n) is a (sub)martingale.

DEFINITION 5.5.1. *Let $(x_n)_{n\geq 1}$ be a sequence of real numbers, $a < b$ two real numbers. We say that $(x_n)_{n\geq 1}$ upcrosses $[a, b]$ at least k times if there exist $2k$ integers $1 \leq n_1 < n_1' < n_2 < n_2' < \cdots < n_k < n_k'$ such that $x_{n_j} \leq a$, $x_{n_j'} \geq b$ for all $j = 1, \ldots, k$. If $(x_n)_{n\geq 1}$ crosses $[a, b]$ at least one time, the number of upcrossings of $(x_n)_{n\geq 1}$ across $[a, b]$, K, is the largest integer k such that $(x_n)_{n\geq 1}$ upcrosses $[a, b]$ at least k times. If this largest k doesn't exist, we set $K := +\infty$ and we say that $(x_n)_{n\geq 1}$ oscillates endlessly across $[a, b]$.*

Observe that $\liminf_n x_n < \limsup_n x_n$ if and only if there exist two numbers $a < b$ such that $(x_n)_{n\geq 1}$ oscillates endlessly across $[a, b]$. Therefore, $(x_n)_{n\geq 1}$ converges if and only if its number of upcrossings across any interval $[a, b]$ is finite.

THEOREM 5.5.1 (Upcrossing Inequality). *Let $(X_n, \mathcal{F}_n)_{n\geq 1}$ be a submartingale. Let $a < b$ two real numbers, and for each $\omega \in \Omega$, let $U_{a,b}(\omega)$ denote the number of upcrossings of the sequence $(X_n(\omega))_{n\geq 1}$ across*

$[a, b]$. *Then*

$$E[U_{a,b}] \leq \sup_{n \geq 1} \frac{E\big[(X_n - a)^+\big]}{b - a}\,. \tag{5.5.1}$$

PROOF. Obviously, the number of upcrossings of $(X_n)_{n\geq 1}$ across $[a, b]$, is equal to the number of upcrossings of $(Y_n)_{n\geq 1}$ across $[0, 1]$, where

$$Y_n := \frac{(X_n - a)^+}{b - a}\,.$$

By Lemma 5.2.1, $(Y_n)_{n\geq 1}$ is a submartingale. We define a sequence of stopping times $T_1^- < T_1^+ < T_2^- < T_2^+ < \dots$ as follows (see Figure 1):

$$T_1^- := \inf\{n \geq 1 : Y_n = 0\}\,,$$
$$T_1^+ := \inf\{n > T_1^- : Y_n \geq 1\}\,,$$

and, for $k \geq 2$,

$$T_k^- := \inf\{n > T_{k-1}^+ : Y_n = 0\}$$
$$T_k^+ := \inf\{n > T_k^- : Y_n \geq 1\}$$

We call each interval $[T_k^-, T_k^+]$ an upcrossing interval. Since the number

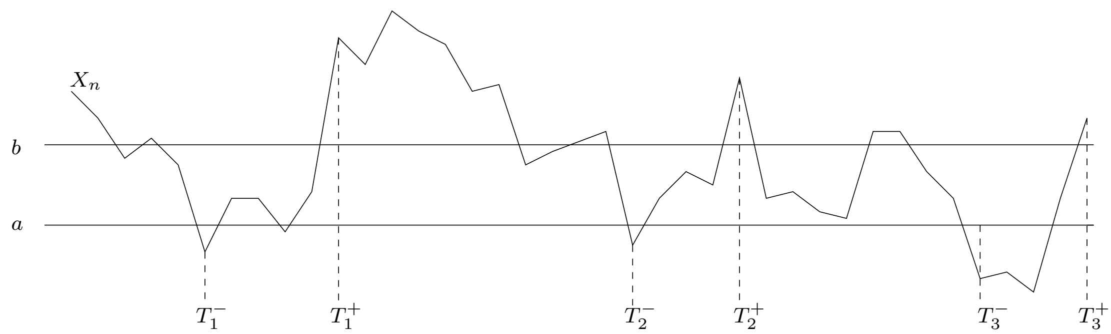


FIGURE 1. The construction of the stopping times T_k^-, T_k^+.

of upcrossings of $(Y_n)_{n\geq 1}$ can be infinite, we take N large and consider $U_{a,b}^{(N)}$, the number of upcrossing intervals of the sequence $(Y_n)_{n\geq 1}$ across $[0, 1]$ entirely contained in $[1, N]$. To express $U_{a,b}^{(N)}$ using $(Y_n)_{n\geq 1}$, we introduce for each $n \geq 1$ a variable C_n which determines whether n belongs to an upcrossing interval:

$$C_n := \begin{cases} 1 & \text{if } n \in (T_k^-, T_k^+] \text{ for some } k\,, \\ 0 & \text{otherwise.} \end{cases}$$

Observe that the sequence C_n is predictable since

$$\{C_n = 1\} = \bigcup_{k\geq 1}\{T_k^- < n\}\cap\{T_k^+ \geq n\} = \bigcup_{k\geq 1}\{T_k^- \leq n-1\}\cap\{T_k^+ \leq n-1\}^c .$$

They key is to observe that one can bound $U_{a,b}^{(N)}$ by (set $Y_0 := 0$)

$$U_{a,b}^{(N)} \leq \sum_{k=1}^{N} C_k(Y_k - Y_{k-1}) .$$

The fact that there is an inequality comes from the fact that we might be considering ks at the end of the interval $[1, N]$ for which $C_k = 1$ but for which the last upcrossing inside $[1, N]$ hasn't been completed. Now this last sum is exactly the Nth term of the transform of (Y_n) by (C_n), considered at time N:

$$\sum_{k=1}^{N} C_k(Y_k - Y_{k-1}) \equiv (C \cdot Y)_N .$$

The Let $\widetilde{C}_n := 1 - C_n$, which is also predictable. By Lemma 5.2.2. $(\widetilde{C} \cdot Y)_n$ is a submartingale, which implies

$$E[(\widetilde{C} \cdot Y)_N] \geq E[(\widetilde{C} \cdot Y)_1] = 0 ,$$

since $C_1 = 0$. Therefore,

$$E[U_{a,b}^{(N)}] \leq E[(C \cdot Y)_N] = E[Y_N] - E[Y_0] - E[(\widetilde{C} \cdot Y)_N] \leq E[Y_N] .$$

We thus have

$$E[U_{a,b}^{(N)}] \leq E[Y_N] \leq \sup_{n\geq 1} E[Y_n] .$$

Since $U_{a,b}^{(N)}$ is increasing in N and $U_{a,b} = \lim_N U_{a,b}^{(N)}$, Monotone Convergence gives (5.5.1). □

The Upcrossing Inequality suggests that convergence of a (sub)martingale $(X_n)_{n\geq 1}$ might hold under some uniformity assumptions on the expectations of $|X_n|$.

A sequence $(X_n)_{n\geq 1}$ is bounded in L^1 if

$$\sup_{n\geq 1} E[|X_n|] < \infty . \tag{5.5.2}$$

Observe that if $(X_n)_{n\geq 1}$ is a submartingale then

$$E[|X_n|] = E[X_n^+ + X_n^-] = 2E[X_n^+] - E[X_n] \leq 2E[X_n^+] - E[X_1] ,$$

and therefore it is L^1-bounded if and only if $\sup_n E[X_n^+] < \infty$. For an arbitrary filtration $(\mathcal{F}_n)_{n\geq 1}$, let $\mathcal{F}_\infty$ be the σ-algebra generated by $(\mathcal{F}_n)_{n\geq 1}$, i.e. $\mathcal{F}_\infty = \sigma\big(\bigcup_n \mathcal{F}_n\big)$.

THEOREM 5.5.2. *Let $(X_n)_{n\geq 1}$ be a submartingale, bounded in L^1. Then there exists $X_\infty \in L^1$, $\mathcal{F}_\infty$-measurable, such that*

$$X_n \to X_\infty\,, \quad a.s. \tag{5.5.3}$$

PROOF. Consider the set M of ωs for which $(X_n(\omega))_{n\geq 1}$ converge. That is,

$$\begin{aligned} M^c &:= \big\{\omega \in \Omega : \liminf_{n\to\infty} X_n(\omega) < \limsup_{n\to\infty} X_n(\omega)\big\} \\ &= \bigcup_{\substack{a,b\in\mathbb{Q} \\ a<b}} \{\omega \in \Omega : U_{a,b}(\omega) = +\infty\}\,. \end{aligned}$$

Since $(X_n)_{n\geq 1}$ is L^1-bounded, we have

$$\sup_{n\geq 1} \frac{E[(X_n-a)^+]}{b-a} \leq \sup_{n\geq 1} \frac{E[|X_n-a|]}{b-a} \leq \frac{a}{b-a} + \sup_{n\geq 1} \frac{E[|X_n|]}{b-a} < \infty\,.$$

By the Upcrossing Inequality, this gives $E[U_{a,b}] < \infty$ and therefore $U_{a,b}$ is finite almost everywhere, i.e. $P(U_{a,b} = +\infty) = 0$ and so $P(M^c) = 0$. Then, define

$$X_\infty(\omega) := \begin{cases} \lim_{n\to\infty} X_n(\omega) & \text{if } \omega \in M\,, \\ 0 & \text{if } \omega \in M^c. \end{cases}$$

Clearly, X_∞ is $\mathcal{F}_\infty$-measurable. Using Fatou's Lemma and L^1-boundedness,

$$E[|X_\infty|] = E[\lim_{n\to\infty} |X_n|] \leq \liminf_{n\to\infty} E[|X_n|] < \infty\,,$$

which shows that $X_\infty \in L^\infty$. □

COROLLARY 5.5.1. *Let $(X_n)_{n\geq 1}$ be a non-negative supermartingale. Then there exists $X_\infty \in L^1$, $\mathcal{F}_\infty$-measurable, such that*

$$X_n \to X_\infty \quad a.s. \tag{5.5.4}$$

PROOF. $Y_n := -X_n$ is a submartingale. Since $E[|Y_n|] = E[X_n] \leq E[X_1] < \infty$, it is bounded in L^1 and so the previous theorem applies. □

If S_n denotes the simple symmetric random walk, then S_n is a martingale but it doesn't converge, since it visits any site of $\mathbb{Z}$ an infinite number of times. Of course, S_n is not bounded in L^1.

Now if this same random walk starts at $x = 1$, Let T denote the time of first visit at the origin. Then $X_n := S_{n\wedge T}$ is a non-negative martingale (Lemma 5.3.3). Therefore, almost surely, $X_\infty := \lim_n S_{n\wedge T}$ exists (by Corollary 5.5.1) and equals to 0 since the symmetric simple walk is recurrent. But $E[X_n] = E[X_0] = 1$, which is different from $E[X_\infty] \equiv 0$. This shows that the a.s. convergence in (5.5.4) may not be in L^1. We will encounter a similar phenomenon in Section 5.6.1.

5.6. Uniformly Integrable Martingales

As we just saw, some stonger hypothesis seems necessary in order to obtain convergence in L^1. This is desirable for various reasons that will become clear in the following sections.

DEFINITION 5.6.1. *A sequence $(X_n)_{n\geq 1}$ is uniformly integrable (UI) if*

$$\lim_{K\to\infty} \sup_{n\geq 1} \int_{|X_n|\geq K} |X_n|\, dP = 0\,. \tag{5.6.1}$$

If $(X_n)_{n\geq 1}$ is UI, then for K large enough,

$$\sup_{n\geq 1} \int_{|X_n|\geq K} |X_n|\, dP \leq 1\,,$$

and therefore

$$E[|X_n|] = \int_{|X_n|<K} |X_n|\, dP + \int_{|X_n|\geq K} |X_n|\, dP \leq K + 1 < \infty\,.$$

This shows that uniform integrability implies L^1-boundedness. The converse can be wrong, as will be seen in Exercise 5.21. Nevertheless, we have

LEMMA 5.6.1. *Let $(X_n)_{n\geq 1}$ be L^1-bounded. Then $(X_n)_{n\geq 1}$ is UI if and only if for all $\epsilon > 0$ there exists $\delta > 0$ such that $P(A) \leq \delta$ implies $\int_A |X_n| dP \leq \epsilon$ for all $n \geq 1$.*

PROOF. Assume first that the sequence is UI and take some $\epsilon > 0$. Let K be such that

$$\sup_{n\geq 1} \int_{|X_n|\geq K} |X_n| dP \leq \frac{\epsilon}{2}\,.$$

Take $\delta := \frac{\epsilon}{2K}$. If $P(A) \leq \delta$, then for any $n \geq 1$,

$$\int_A |X_n| dP = \int_{A\cap\{|X_n|\geq K\}} |X_n| dP + \int_{A\cap\{|X_n|\leq K\}} |X_n| dP \leq \epsilon\,.$$

Going the other way, fix $\epsilon > 0$. Let $\delta > 0$ be such that $P(A) \leq \delta$ implies $\int_A |X_n| dP \leq \epsilon$ for all $n \geq 1$. If $M := \sup_{n\geq 1} E[|X_n|]$, let K be large enough so that $MK^{-1} \leq \delta$. Then, by Chebychev's Inequality, $P(|X_n| \geq K) \leq \delta$. Therefore with $A = \{|X_n| \geq K\}$, $\int_{|X_n|\geq K} |X_n| dP \leq \epsilon$, and so $(X_n)_{n\geq 1}$ is UI. □

Lemma 5.6.2. *Let $(X_n)_{n\geq 1}$, $X_n \in L^1$ be a sequence converging in probability: $X_n \xrightarrow{P} X$. If $(X_n)_{n\geq 1}$ is UI, then $X_n \xrightarrow{L^1} X$.*

Proof. We show that $(X_n)_{n\geq 1}$ is Cauchy in L^1. Decompose $E[|X_n - X_m|]$ as

$$\int_{\{|X_n-X_m|\leq\epsilon\}} |X_n-X_m| dP + \int_{\{\epsilon<|X_n-X_m|\leq K\}} |X_n-X_m| dP + \int_{\{|X_n-X_m|>K\}} |X_n-X_m| dP$$

The first integral is bounded by ϵ, the second by $KP(|X_n - X_m| \geq \epsilon)$. But since $\{|X_n - X_m| \geq \epsilon\} \subset \{|X_n - X| \geq \epsilon/2\} \cup \{|X - X_m| \geq \epsilon/2\}$, and since $X_n \xrightarrow{P} X$, we have $\limsup_{m,n} P(|X_n - X_m| \geq \epsilon) = 0$. To study the third integral, let $\delta > 0$ be as in Lemma 5.6.1 (remember that UI implies L^1-boundedness). Then, take m, n large enough so that $P(|X_n - X_m| \geq K) \leq \delta$. We have with $A = \{|X_n - X_m| > K\}$,

$$\int_{\{|X_n-X_m|>K\}} |X_n - X_m| dP \leq \int_A |X_n| dP + \int_A |X_m| dP \leq 2\epsilon\,.$$

Therefore, $\limsup_{m,n} E[|X_n - X_m|] \leq 3\epsilon$. □

We can now extend the almost everywhere convergence of Theorem 5.5.2 to convergence in L^1 for UI martingales.

Theorem 5.6.1. *Let $(X_n)_{n\geq 1}$ be a UI submartingale. Then there exists $X_\infty \in L^1$, $\mathcal{F}_\infty$-measurable, such that*

$$X_n \to X_\infty \quad \textit{a.s. and in } L^1\,. \tag{5.6.2}$$

Moreover, if $(X_n)_{n\geq 1}$ is a martingale, then $X_n = E[X_\infty|\mathcal{F}_n]$.

Proof. Since (X_n) is UI it is bounded in L^1, so Theorem 5.5.2 guarantees the almost everywhere convergence of $X_n \to X_\infty$. Since this implies convergence in probability, Lemma 5.6.2 implies convergence in L^1. To show that $X_n = E[X_\infty|\mathcal{F}_n]$ if X_n is a martingale, we must show that

$$\int_A X_n dP = \int_A X_\infty dP \quad \forall A \in \mathcal{F}_n\,.$$

But, since $X_n \to X_\infty$ in L^1, we have, for all $A \in \mathcal{F}_n$,

$$\Big|\int_A X_\infty dP - \int_A X_m dP\Big| \le \int_A |X_\infty - X_m| dP \le \int |X_\infty - X_m| dP \to 0$$

when $m \to \infty$. Therefore, using the martingale property,

$$\int_A X_\infty dP = \lim_{m\to\infty} \int_A X_m dP = \int_A X_n dP\,. \quad \square$$

We can now state a representation result for UI martingales.

Proposition 5.6.1. *A martingale is closed if and only if it is UI.*

Proof. Let $X \in L^1$, and $(\mathcal{F}_n)_{n\ge 1}$ be an arbitrary filtration, such that $X_n := E[X|\mathcal{F}_n]$. We have $|X_n| \le E[|X||\mathcal{F}_n]$, which implies

$$\int_A |X_n| dP \le \int_A E[|X||\mathcal{F}_n] dP = \int_A |X| dP \quad \forall A \in \mathcal{F}_n\,. \qquad (5.6.3)$$

With $A = \Omega$, this gives $E[|X_n|] \le E[|X|]$, and so by Chebychev,

$$P(|X_n| \ge K) \le \frac{E[|X_n|]}{K} \le \frac{E[|X|]}{K}\,.$$

Lemma 5.6.3. *Let $X \in L^1$. Given any $\epsilon > 0$, there exists $\delta > 0$ such that if $A \in \mathcal{F}$ is such that $P(A) \le \delta$, then $\int_A |X| dP \le \epsilon$.*

Proof. Assume the result is false: there exists some $\epsilon_0 > 0$ and a sequence of events A_n such that $P(A_n) \le 2^{-n}$ and $\int_{A_n} |X| dP > \epsilon_0$ for all n. Take $A := \limsup_n A_n$. By Borel-Cantelli, $P(A) = 0$. By Fatou,

$$\int_A |X| dP = \int |X| dP - \liminf_{n\to\infty} \int_{A_n^c} |X| dP \ge \epsilon_0\,,$$

which is impossible. $\square$

Fix $\epsilon > 0$. Let $\delta > 0$ be as in the lemma. Take K large enough such that $E[|X|] \le K\delta$. Then $P(|X_n| \ge K) \le \delta$, and using (5.6.3) with $A = \{|X_n| \ge K\}$,

$$\int_{|X_n|\ge K} |X_n| dP \le \int_{|X_n|\ge K} |X| dP \le \epsilon\,.$$

This shows that X_n is UI. On the other hand, if X_n is UI, then we know from Theorem 5.6.1 that there exists $X_\infty \in L^1$ such that $X_n \to X_\infty$ (a.s. and in L^1), and that $X_n = E[X_\infty|\mathcal{F}_n]$. □

5.6.1. Application to the branching process. Consider the branching process (X_n) of Example 5.1.5. The basic properties of extinction and survival were obtained in Exercise 2.11 using moment generating functions. Here we derive the same properties using only martingale theory. As usual, we assume that the number of offsprings $Y_j^{(n)}$ satisfies $p_0 = P(Y_j^{(n)} = 0) > 0$, $E[Y_j^{(n)}] = \lambda > 0$. Moreover, we will assume that

$$\sigma^2 := \operatorname{var} Y_j^{(n)} > 0 .$$

Let X_n denote the size of the population of the nth generation ($X_0 = 1$). As we already saw, $Z_n := \lambda^{-n} X_n$ is a martingale. Since it is non-negative, it converges by Corollary 5.5.1: there exists $Z_\infty \in L^1$ such that

$$Z_n \to Z_\infty \quad a.s.$$

We consider separately the cases $\lambda \leq 1$, $\lambda = 1$ and $\lambda > 1$.

Case $\lambda < 1$*:* In this case, $\lambda^{-n} \nearrow +\infty$, so in order for Z_n to converge to a finite value, and since X_n takes values in $\{0, 1, 2, \dots\}$, the only possibility is that $X_n = 0$ for large n, which means almost sure extinction of the population.

Case $\lambda = 1$*:* In this case, X_n is a non-negative martingale. In particular, $E[X_n] = E[X_0] = 1$, and again $X := \lim_n X_n$ exists a.s. by Corollary 5.5.1. Since X_n is integer-valued, there must almost surely exist some integer $K \geq 0$ such that $X_n = K$ for n sufficiently large. For each $K \geq 0$,

$$P(X_n = K \text{ for large enough } n) \leq \sum_N P(X_n = K \, \forall n \geq N) .$$

We will show that K is necessarily equal to zero, by showing that if $K \geq 1$, then $P(X_n = K \, \forall n \geq N) = 0$ for all N. We have, by the

Markov Property,

$$\begin{aligned} P(X_n = K \,\forall n \geq N) &= \lim_{L\to\infty} P(X_n = K \,\forall n \in \{N, \dots, N+L\}) \\ &= \lim_{L\to\infty} P(X_N = K) \prod_{n=N}^{N+L-1} P(X_{n+1} = K | X_n = K) \\ &\leq \limsup_{L\to\infty} \prod_{n=N}^{N+L} P(X_{n+1} = K | X_n = K)\,. \end{aligned}$$

But

$$P(X_{n+1} \neq K | X_n = K) \geq P(Y_j^{(n)} \geq 2, j = 1, \dots, K) = (1 - p_0 - p_1)^K\,,$$

which is > 0. Namely, $K \geq 1$, and if $1 - p_0 - p_1 = 0$ were true, then necessarily $\lambda = p_1 < 1$. Therefore, $P(X_n = K \,\forall n \geq N) = 0$ when $K \geq 1$, and so the only possibility is $X = K = 0$: the process dies out in a finite time. Observe that in particular, $E[X] = 0$, and therefore $E[X_n]$ does not converge to $E[X]$: this gives another case where the convergence $X_n \to X$ of Theorem 5.5.2 does not necessarily take place in L^1.

Case $\lambda > 1$*:* We show that $P(Z_\infty > 0) > 0$. To do so, we will show that Z_n is uniformly integrable, which implies convergence in L^1 by Theorem 5.6.1, and so

$$E[Z_\infty] = \lim_{n\to\infty} E[Z_n] = 1\,.$$

This implies $P(Z_\infty > 0) > 0$. To check uniform integrability, we show that Z_n is bounded in L^2 (see Exercise 5.21).

$$\begin{aligned} (Z_n - Z_{n-1})^2 &= \lambda^{-2n}(X_n - \lambda X_{n-1})^2 \\ &= \lambda^{-2n}\Big\{\sum_{j=1}^{X_{n-1}} (Y_j^{(n)} - \lambda)\Big\}^2 \\ &= \lambda^{-2n}\Big\{\sum_{j=1}^{X_{n-1}} (Y_j^{(n)} - \lambda)^2 + 2 \sum_{1\leq i<j\leq X_{n-1}} (Y_i^{(n)} - \lambda)(Y_j^{(n)} - \lambda)\Big\} \end{aligned}$$

Now, by independence and since $E[Y_j^{(n)}] = \lambda$,

$$E\Big[\sum_{1\le i<j\le X_{n-1}} (Y_i^{(n)}-\lambda)(Y_j^{(n)}-\lambda)\Big]$$
$$= E\Big[E\big[\sum_{1\le i<j\le X_{n-1}} (Y_i^{(n)}-\lambda)(Y_j^{(n)}-\lambda)|\mathcal{F}_{n-1}\big]\Big]$$
$$= E\Big[\sum_{1\le i<j\le X_{n-1}} E\big[(Y_i^{(n)}-\lambda)(Y_j^{(n)}-\lambda)|\mathcal{F}_{n-1}\big]\Big] = 0\,.$$

Therefore, since $E[(Y_j^{(n)}-\lambda)^2] = \operatorname{var} Y_j^{(n)} = \sigma^2$,

$$\begin{aligned}
E[(Z_n - Z_{n-1})^2] &= \lambda^{-2n} E\Big[\sum_{j=1}^{X_{n-1}} (Y_j^{(n)}-\lambda)^2\Big] \\
&= \lambda^{-2n} E\Big[E\big[\sum_{j=1}^{X_{n-1}} (Y_j^{(n)}-\lambda)^2|\mathcal{F}_{n-1}\big]\Big] \\
&= \lambda^{-2n} E[\sigma^2 X_{n-1}] \\
&= \sigma^2\lambda^{-2n} E[\lambda^{n-1} Z_{n-1}] = \sigma^2\lambda^{-n-1}
\end{aligned}$$

By Lemma 5.4.1, this shows that Z_n is bounded in L^2, which in turn implies that it is also uniformly integrable. Therefore, the event $\{Z_\infty > 0\}$ has positive probability, and on it we have, for large enough n, $X_n = \lambda^n Z_n \ge \frac{Z_\infty}{2}\lambda^n \nearrow +\infty$ exponentially fast. It can be verified by the reader that $Z_\infty > 0$ on the whole non-extinction set.

5.6.2. Application to measure theory. The convergence theorems for martingales allow to give a probabilistic proof of the Radon-Nikodým Theorem. For ease of presentation, we will prove this famous result of measure theory in the particular case where the σ-algebra is countably generated. The minor modifications needed to cover the general case can be found in [**Wil91**].

THEOREM 5.6.2. *Let $(\Omega, \mathcal{F}, P)$ be a probability space in which $\mathcal{F}$ is* countably generated*, i.e. there exists a countable family of sets $A_n \in \mathcal{F}$ such that $\mathcal{F} = \sigma(A_n, n \ge 1)$. Let Q be a finite measure on $(\Omega, \mathcal{F})$ which is absolutely continuous with respect to P. Then there exists $X \in L^1(\Omega, \mathcal{F}, P)$, $X \ge 0$, called the* Radon-Nikodým derivative of Q

with respect to P, such that

$$Q(B) = \int_B X dP\,, \quad \forall B \in \mathcal{F}\,. \tag{5.6.4}$$

Proof. The countability assumption allows to consider the filtration $\mathcal{F}_n := \sigma(A_1, \dots, A_n)$. We denote the atoms of $\mathcal{F}_n$ by $\{C_{n,1}, \dots, C_{n,k_n}\}$. Consider the random variable $X_n : \Omega \to \mathbb{R}$ defined as follows: if $\omega \in C_{n,j}$, then

$$X_n(\omega) := \begin{cases} \frac{Q(C_{n,j})}{P(C_{n,j})} & \text{if } P(C_{n,j}) > 0\,, \\ 0 & \text{if } P(C_{n,j}) = 0\,. \end{cases}$$

Let $n \geq m$ and $A \in \mathcal{F}_m$. Then

$$\int_A X_n dP = \sum_{\substack{j: C_j^{(n)} \subset A \\ P(C_j^{(n)}) > 0}} \int_{C_j^{(n)}} X_n dP = \sum_{\substack{j: C_j^{(n)} \subset A \\ P(C_j^{(n)}) > 0}} Q(C_j^{(n)}) = Q(A)\,.$$

This last equality follows from the absolute continuity of Q with respect to P. We have thus shown that

$$Q(A) = \int_A X_n dP \quad \forall A \in \mathcal{F}_n. \tag{5.6.5}$$

In particular, since X_n is constant on the atoms of $\mathcal{F}_{n+1}$, $(X_n)_{n\geq 1}$ is a martingale with respect to $(\mathcal{F}_n)_{n\geq 1}$ (and to the measure P). We show that $(X_n)_{n\geq 1}$ is uniformly integrable. Fix some $\epsilon > 0$. There exists $\delta > 0$ such that if $A \in \mathcal{F}$ is such that $P(A) \leq \delta$, then $Q(A) \leq \epsilon$. (The existence of this δ can be shown exactly as in Lemma 5.6.3. Namely, assume the claim is wrong. Then there exists some $\epsilon_0 > 0$ and a sequence D_n such that $P(D_n) \leq 2^{-n}$ and $Q(D_n) \geq \epsilon_0$. Let $D := \limsup_n D_n$. Then $P(D) = 0$ by Borel-Cantelli, but by Fatou

$$Q(D) = \int_D dQ = \int \limsup_{n\to\infty} 1_{D_n} dQ \geq 1 - \liminf_{n\to\infty} Q(D_n^c) \geq \epsilon_0\,,$$

a contradiction with the absolute continuity of Q with respect to P.) Let K be large enough such that $Q(\Omega) \leq \delta K$. By Chebychev and (5.6.5) with $A = \Omega$,

$$P(X_n \geq K) \leq \frac{E[X_n]}{K} = \frac{Q(\Omega)}{K} \leq \delta\,.$$

Therefore, using (5.6.5) with $A = \{X_n \geq K\}$,

$$\int_{X_n \geq K} X_n dP = Q(X_n \geq K) \leq \epsilon\,,$$

which shows that X_n is uniformly integrable. By Theorem 5.6.1, there exists $X \in L^1$ such that $X_n \to X$ in L^1. In particular, for all $A \in \mathcal{F}_n$,

$$\int_A X dP = \lim_{n\to\infty} \int_A X_n dP = Q(A)\,.$$

Therefore, the measures Q and $\widetilde{Q} := XP$ coincide on the algebra $\bigcup_{n\geq 1} \mathcal{F}_n$. Since this algebra generates $\mathcal{F}$, we have $Q = \widetilde{Q}$. This proves the theorem. □

We considered in Example 5.1.4 the case where $\Omega = [0,1)$, with the dyadic filtration

$$\mathcal{F}_n := \sigma\Big(\big[\frac{i-1}{2^n}, \frac{i}{2^n}\big) : i = 1, 2, \dots, 2^n\Big)\,.$$

Clearly the Borel σ-field $\mathcal{B}([0,1))$ is generated by $\bigcup_n \mathcal{F}_n$, and the previous theorem applies.

5.6.3. Lévy's Upward Convergence Theorem. We are now ready to state the main convergence result promised at the beginning of our study of martingales, for closed martingales. Nevertheless, we formulate it without using martingales' terminology.

THEOREM 5.6.3 (Lévy's Upward Theorem). *Let $X \in L^1$ on a probability space $(\Omega, \mathcal{F}, P)$. Let $(\mathcal{F}_n)_{n\geq 1}$ be a filtration, $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$. Then*

$$E[X|\mathcal{F}_n] \to E[X|\mathcal{F}_\infty] \quad \textit{a.s. and in } L^1\,. \tag{5.6.6}$$

PROOF. By Proposition 5.6.1, $E[X|\mathcal{F}_n]$ is uniformly integrable. By Theorem 5.6.1, it converges to some $X_\infty \in L^1$, and $E[X|\mathcal{F}_n] = E[X_\infty|\mathcal{F}_n]$. We need to show that $X_\infty = E[X|\mathcal{F}_\infty]$, that is

$$\int_A X_\infty\, dP = \int_A X\, dP\,, \quad \forall A \in \mathcal{F}_\infty\,. \tag{5.6.7}$$

First, assume $X \geq 0$ and define two measures on $(\Omega, \mathcal{F}_\infty)$:

$$\mu(A) := \int_A X_\infty\, dP\,, \qquad \nu(A) := \int_A X\, dP\,.$$

Let $A \in \mathcal{F}_n$. Then,

$$\int_A X\,dP = \int_A E[X|\mathcal{F}_n]\,dP = \int_A E[X_\infty|\mathcal{F}_n]\,dP = \int_A X_\infty\,dP\,.$$

Therefore, μ and ν agree on the algebra $\bigcup_n \mathcal{F}_n$. By Carathéodory's Extension Theorem, they also agree on $\mathcal{F}_\infty$, which shows (5.6.7). In the general case, simply decompose $X = X^+ - X^-$. □

As an application of the Upward Theorem, we prove the 0-1 Law of Theorem 3.2.1, Section 3:

Kolmogorov's 0-1 Law. *If $(X_n)_{n\geq 1}$ is i.i.d., and if*

$$\mathcal{T}_\infty := \bigcap_{n\geq 1} \sigma(X_n, X_{n+1}, \dots)$$

denotes its tail-σ-field, then any $A \in \mathcal{T}_\infty$ is trivial: $P(A) \in \{0,1\}$.

Proof. Consider the natural filtration $\mathcal{F}_n$ associated to $(X_n)_{n\geq 1}$. Take $A \in \mathcal{T}_\infty$, and set $X := 1_A$. Since X is $\mathcal{T}_\infty$-measurable, it is independent of $\mathcal{F}_n$ for all n, and so $E[X|\mathcal{F}_n] = E[X] = P(A)$. Since $X \in L^1$ we get by the Upward Theorem,

$$P(A) = E[X|\mathcal{F}_n] \to E[X|\mathcal{F}_\infty] = X \text{ a.s.}$$

since X is $\mathcal{T}_\infty$-measurable and since $\mathcal{T}_\infty \subset \mathcal{F}_\infty$. Therefore $P(A) = 1_A \in \{0,1\}$. □

5.6.4. Lévy' Backward Convergence Theorem. In a similar way, one can show a version of Lévy' Theorem but for *decreasing* sequence of σ-algebras.

Theorem 5.6.4 (Lévy's Backward Theorem). *Let $X \in L^1$ on a probability space $(\Omega, \mathcal{F}, P)$. Let $(\mathcal{T}_n)_{n\geq 1}$ be decreasing sequence of sub-σ-algebras: $\mathcal{T}_1 \supset \mathcal{T}_2 \supset \dots$. Define $\mathcal{T}_\infty = \bigcap_n \mathcal{T}_n$. Then*

$$E[X|\mathcal{T}_n] \to E[X|\mathcal{T}_\infty] \quad \text{a.s. and in } L^1\,. \tag{5.6.8}$$

Proof. Think backwards: for all large integer $N \geq 1$, define, for $k = 0, 1, \dots, N$, $\mathcal{F}_k := \mathcal{T}_{N-k}$. Then $\mathcal{F}_k \subset \mathcal{F}_{k+1}$. Define $Z_k := E[X|\mathcal{F}_k]$. Clearly, $(Z_k, \mathcal{F}_k)_{k=1}^N$ is a martingale. For each N, let $U_{a,b}^{(N)}$ denote the number of upcrossings of the sequence $Z_0, Z_1, \dots, Z_N$ across $[a,b]$. By the Upcrossing Inequality,

$$E[U_{a,b}^{(N)}] \leq E[(Z_N - a)^+] \leq E[|Z_N|] + a \leq E[|X|] + a < \infty\,.$$

This shows that $Z := \lim_n E[X|\mathcal{T}_n]$ exists almost surely. Exactly as we did in Proposition 5.6.1, we can show that $E[X|\mathcal{T}_n]$ is UI. Therefore, $E[X|\mathcal{T}_n] \to Z$ also in L^1 by Lemma 5.6.2. To verify that $Z = E[X|\mathcal{T}_\infty]$, take any $A \in \mathcal{T}_\infty$. Then $A \in \mathcal{T}_n$ since $\mathcal{T}_\infty \subset \mathcal{T}_n$ for all $n \geq 1$, and so

$$\int_A Z dP = \lim_{n\to\infty} \int_A E[X|\mathcal{T}_n] dP = \lim_{n\to\infty} \int_A X dP = \int_A X dP \,.$$

This finishes the proof. □

As an application of the Backward Theorem, we prove the (pas compris... VOIR LEGALL FIMZINHO)

Strong Law of Large Numbers: *let $X_1, X_2, \dots$ be an i.i.d. sequence such that $X_1 \in L^1$, $S_n := X_1 + \cdots + X_n$. Then*

$$\frac{S_n}{n} \to E[X_1] \quad a.s.$$

Proof. Define $\mathcal{T}_n := \sigma(S_n, S_{n+1}, \dots)$, $\mathcal{T}_\infty := \bigcap_n \mathcal{T}_n$. We have

$$\frac{S_n}{n} = E\big[\frac{S_n}{n}|\mathcal{T}_n\big] = \frac{1}{n}\sum_{k=1}^n E[X_k|\mathcal{T}_n] = E[X_1|\mathcal{T}_n] \,.$$

By the Backward Theorem, $Z = \lim_n \frac{S_n}{n}$ exists almost surely and in L^1. Observe that since,

$$Z = \lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \lim_{n\to\infty} \frac{X_k + \cdots + X_{n+k-1}}{n} \,,$$

Z is $\mathcal{T}_k$-measurable for all k. In particular, it is $\mathcal{T}_\infty$-measurable. By Theorem 3.2.2, Z is therefore almost surely constant: $P(Z = c) = 1$. To compute c,

$$c = E[Z] = \lim_{n\to\infty} E\big[\frac{S_n}{n}\big] = E[X_1] \,.$$

□

5.7. Doob Decomposition (EMPTY)

5.8. Martingales and Markov Chains

In this section we see interesting applications of martingale techniques to the qualitative study of Markov chains, as we already encountered in Section 5.4.2. We follow Varadhan [**Var00**].

Generally speaking, there are various ways of associating martingales to a stochastic process. Let $(X_n)_{n\geq 1}$ be a sequence of random variables and let $(\mathcal{F}_n)_{n\geq 1}$ denote its natural filtration. Let $f : \mathbb{R} \to \mathbb{R}$ be arbitrary but measurable and bounded. For all $k \geq 0$, define

$$h_k(X_0, \dots, X_k) := E_x[f(X_{k+1})|\mathcal{F}_k] - f(X_k)\,.$$

Let $\triangle_k := f(X_{k+1}) - f(X_k) - h_k(X_0, \dots, X_k)$ which, by the definition of h_k satisfies $E_x[\triangle_k|\mathcal{F}_k] = 0$. Let

$$Z_n := \sum_{k=0}^{n-1} \triangle_k = f(X_n) - f(X_0) - \sum_{k=0}^{n-1} h_k(X_0, \dots, X_k)\,, \tag{5.8.1}$$

then $E_x[Z_{n+1} - Z_n|\mathcal{F}_n] = E_x[\triangle_n|\mathcal{F}_n] = 0$, and so Z_n is a martingale. In fact, Z_n is nothing but the martingale obtained by the Doob decomposition of $f(X_n)$ (see previous section).

Consider now the case where X_n is a Markov chain with state space S and transition matrix Q. For simplicity, assume that the chain is irreducible. Remind that if f is a bounded function on S,

$$Qf(x) := \sum_{y\in S} Q(x, y)f(y)\,.$$

Then, on $\{X_n = z\}$ we get

$$\begin{aligned} E_x[f(X_{n+1})|\mathcal{F}_n] &= \sum_{y\in S} f(y)P_x(X_{n+1} = y|\mathcal{F}_n) \\ &= \sum_{y\in S} f(y)Q(z, y) = Qf(X_n)\,, \end{aligned} \tag{5.8.2}$$

and so $h_k(X_0, \dots, X_k) = h_k(X_k) \equiv h(X_k)$, where

$$h := Qf - f\,. \tag{5.8.3}$$

The associated martingale thus takes the form

$$Z_n = f(X_n) - f(X_0) - \sum_{k=0}^{n-1} h(X_k)\,. \tag{5.8.4}$$

Another way to associate martingales to a Markov chain is to consider functions f with special properties. A function f is called harmonic if $Qf = f$, superharmonic if $Qf \leq f$, and subharmonic if $Qf \geq f$. The following lemma follows directly from (5.8.2):

LEMMA 5.8.1. *Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain on S with transition matrix Q. Let $f : S \to \mathbb{R}$ be such that $E_x[|f(X_n)|] < \infty$ for all $n \geq 0$. Then*

$$f \text{ is } \begin{cases} \text{harmonic} \\ \text{superharmonic} \\ \text{subharmonic} \end{cases} \iff (f(X_n))_{n\geq 0} \text{ is } \begin{cases} \text{a martingale} \\ \text{a supermartingale} \\ \text{a submartingale} \end{cases}.$$

5.8.1. The Discrete Dirichlet Problem. Let S be countable, Q be a transition matrix on S. Let $A \subset S$ be a non-empty set. Consider the following problem, called Dirichlet problem: *find a bounded function $f : S \to \mathbb{R}_+$ such that*

$$Qf - f = 0 \text{ on } A^c\,, \tag{5.8.5}$$

$$f = 1 \text{ on } A\,. \tag{5.8.6}$$

A solution to this problem can be given using ideas from Markov chains and martingales. Let X_n the Markov chain on S with transition matrix Q. Define the first visit time at A: $T_A := \inf\{n \geq 0 : X_n \in A\}$. For all $x \in S$, let

$$\phi_A(x) := P_x(T_A < \infty)\,.$$

Clearly, $\phi_A = 1$ on A. If $x \in A^c$, write

$$\phi_A(x) = \sum_{y\in S} P_x(T_A < \infty, X_1 = y)$$

If $y \in A$ then $\{X_1 = y\} \subset \{T_A < \infty\}$ and

$$P_x(T_A < \infty, X_1 = y) = P_x(X_1 = y) = Q(x, y) = Q(x, y)\phi_A(y)\,.$$

If $y \in A^c$, we use the Markov Property at time $n = 1$:

$$\begin{aligned} P_x(T_A < \infty, X_1 = y) &= E_x[1_{\{X_1=y\}} \cdot 1_{\{T_A<\infty\}} \circ \theta] \\ &= E_x[1_{\{X_1=y\}} E_{X_1}[1_{\{T_A<\infty\}}]] \\ &= P_x(X_1 = y)P_y(T_A < \infty) \equiv Q(x, y)\phi_A(y)\,. \end{aligned}$$

This shows that $\phi_A = Q\phi_A$ on A^c. Therefore, ϕ_A is a solution for the Dirichlet problem. The following result shows that any other solution dominates ϕ_A, which is useful in practical situations.

LEMMA 5.8.2. *If f is a solution of the Dirichlet Problem, then $f \geq \phi_A$.*

PROOF. Let f be a solution of the Dirichlet Problem. Define $F := f \wedge 1$. We show that F is superharmonic. Clearly, $F = 1$ on A. Then, if $x \in A^c$,

$$QF(x) = \sum_{y\in S} Q(x,y)F(y) \le \sum_{y\in S} Q(x,y)f(y) = Qf(x) = f(x)\,.$$

Since we always have $QF(x) = \sum_{y\in S} Q(x,y)F(y) \le \sum_{y\in S} Q(x,y) = 1$, this shows that $QF \le F$. Consider the supermartingale $F(X_n)$. Since $F \le 1$, the Optional Stopping Theorem implies that for any bounded stopping time T,

$$E_x[F(X_T)] \le E_x[F(X_0)] = F(x)\,.$$

Since T_A might not be bounded, we consider first its truncated versions $T_A \wedge N$, for large $N \ge 1$. Remembering that $T_A \wedge N$ is also a stopping time (Lemma 5.3.1),

$$f(x) \ge F(x) \ge E_x[F(X_{T_A\wedge N})]\,.$$

We then take the limit $N \to \infty$ in the previous expression:

$$f(x) \ge \limsup_{N\to\infty} \int F(X_{T_A\wedge N})dP_x \ge \limsup_{N\to\infty} \int_{\{T_A<\infty\}} F(X_{T_A\wedge N})dP_x\,.$$

Now observe that on $\{T_A < \infty\}$, $T_A \wedge N \to T_A$ and so $F(X_{T_A\wedge N}) \to 1$. Since F is bounded, Bounded Convergence gives

$$f(x) \ge \int_{\{T_A<\infty\}} F(X_{T_A})dP_x = P_x(T_A < \infty) = \phi_A(x)\,. \quad \square$$

The upper bound $\phi_A(x) \le f(x)$ of Theorem 5.8.1 could be useful if one knew something about the way in which f behaves when $x \to \infty$. As an application, consider the simple symmetric random walk on $S = \mathbb{Z}^d$, $d \ge 3$. Let $0 < \alpha < d-2$ and consider the function

$$\Phi(x) := \begin{cases} \|x\|^{-\alpha} & \text{if } x \neq 0\,, \\ 1 & \text{if } x = 0\,. \end{cases}$$

We claim that Φ is superharmonic far from the origin: if $\|x\|$ is large, then

$$Q\Phi(x) \le \Phi(x)\,. \tag{5.8.7}$$

Namely, let $e_1, \dots, e_d$ denote the unit vectors of $\mathbb{Z}^d$ and write

$$Q\Phi(x) = \frac{1}{2d}\sum_{j=1}^{d}\Big\{\frac{1}{\|x+e_i\|^\alpha} + \frac{1}{\|x-e_i\|^\alpha}\Big\} = \Phi(x)\frac{1}{2d}\sum_{j=1}^{d} g_i(\widehat{x}, \|x\|^{-\alpha})\,,$$

where $\widehat{x} = \frac{x}{\|x\|}$, and

$$g_i(\widehat{x}, t) := \frac{1}{\|\widehat{x} + te_i\|^\alpha} + \frac{1}{\|\widehat{x} - te_i\|^\alpha}\,.$$

We clearly have $g_i(\widehat{x}, 0) = 2$, and it can be verified by direct computation that

$$\frac{\partial}{\partial t} g_i(\widehat{x}, t)\big|_{t=0} = 0\,, \quad \frac{\partial^2}{\partial^2 t} g_i(\widehat{x}, t)\big|_{t=0} = -2\alpha\big(1 - \widehat{x}_i^2(\alpha + 2)\big)\,.$$

Together with $0 < \alpha < d - 2$, these bounds imply that

$$\frac{1}{2d} \sum_{j=1}^{d} g_i(\widehat{x}, t) \leq 1$$

when $|t|$ is small enough, which shows (5.8.7).

Now take some large integer $L \geq 1$, consider the Euclidian sphere $B_L = \{x \in \mathbb{Z}^d : \|x\| \leq L\}$, and run the random walk S_n starting from a point $x \in B_L^c$. Until hitting B_L, $\Phi(S_n)$ is a supermartingale. Consider the first hitting time of B_L: $T_L := \inf\{n \geq 0 : S_n \in B_L\}$. By the Optional Stopping Theorem,

$$E_x[\Phi(S_{T_L \wedge N})] \leq E_x[\Phi(S_0)] = \Phi(x)$$

for all $N \geq 1$. But since

$$E_x[\Phi(S_{T_L \wedge N})] \geq \int_{T_L \leq N} \Phi(S_{T_L}) dP_x \geq P_x(T_L \leq N) \inf_{y \in B_L} \Phi(y)\,.$$

Since $\inf_{y \in B_L} \Phi(y) > 0$, this gives the following upper bound, uniformly in N:

$$P_x(T_L \leq N) \leq \frac{\Phi(x)}{\inf_{y \in B_L} \Phi(y)}$$

which becomes, as $N \to \infty$,

$$P_x(T_L < \infty) \leq \frac{\Phi(x)}{\inf_{y \in B_L} \Phi(y)}\,.$$

Since $\Phi(x) \to 0$ when $\|x\| \to \infty$, this shows in particular that $P_x(T_L < \infty) < 1$ when $\|x\|$ is large enough: the walk is transient in dimension $d \geq 3$.

Consider now the Modified Dirichlet problem: *if* $g : A \to \mathbb{R}_+$ *is bounded, find a bounded function* $f : S \to \mathbb{R}_+$ *such that*

$$Qf - f = 0 \text{ on } A^c\,, \tag{5.8.8}$$

$$f = g \text{ on } A\,. \tag{5.8.9}$$

THEOREM 5.8.1. *Assume* $\phi_A = 1$*, i.e.* $T_A < \infty$ P_x*-almost surely for all* $x \in S$*. Then for any* $g : A \to \mathbb{R}$*, the solution of the modified Dirichlet problem (if any) is unique, and given by*

$$f_*(x) = E_x[g(X_{T_A})]\,. \tag{5.8.10}$$

PROOF. We study the function $x \mapsto E_x[g(X_{T_A})]$, under the assumption that $P_x(T_A < \infty) = 1$ for all $x \in S$, and show that $E_x[g(X_{T_A})] = f(x)$ for any solution f of the modified problem. First, if $x \in A$ then $T_A = 0$ and so $E_x[g(X_{T_A})] = E_x[g(X_0)] = g(x) = f(x)$ since $f = g$ on A. We need thus only consider from now on the case $x \in A^c$. Consider the martingale defined in (5.8.1):

$$Z_n = f(X_n) - f(X_0) + \sum_{k=0}^{n-1} h(X_k)\,, \quad n \geq 1\,.$$

By the Stopping Theorem, $E_x[Z_{T_A \wedge N}] = E_x[Z_1]$, but

$$E_x[Z_1] = E_x[f(X_1) - f(X_0) + h(X_0)] = Qf(x) - f(x) + h(x) = 2h(x) = 0$$

since $h = Qf - f = 0$ on A^c. Now, since $h(X_k) = 0$ for all $k < T_A$, $Z_{T_A \wedge N} = f(X_{T_A \wedge N}) - f(X_0)$. Then, observe that on $\{T_A < \infty\}$ we have $f(X_{T_A \wedge N}) \to f(X_{T_A}) = g(X_{T_A})$. But since we are assuming that $P_x(T_A < \infty) = 1$, Bounded Convergence gives

$$\lim_{N\to\infty} E_x[f(X_{T_A \wedge N})] = \lim_{N\to\infty} \int_{\{T_A < \infty\}} f(X_{T_A \wedge N}) dP_x = E_x[g(X_{T_A})]$$

Since $E_x[f(X_0)] = f(x)$, we have proved that $E_x[g(X_{T_A})] = f(x)$. □

5.8.2. Application: recurrence of birth and death chains. Consider the birth and death chain $(X_n)_{n\geq 0}$ with state space $S = \{0, 1, 2, \dots\}$ and transition matrix

$$Q(x, x+1) = p_x\,,\ Q(x, x-1) = q_x\,,\ Q(x, x) = r_x\,,$$

with $p_x + q_x + r_x = 1$ and $q_0 := 0$. We will assume that $0 < p_x < 1$, $0 < q_x < 1$, so that the chain is irreducible.

We look for a recurrence criterium in function of the triples $\{(q_x, r_x, q_x)\}_{x \geq 0}$. First, we look for a harmonic function $\varphi : S \to \mathbb{R}_+$ which will allow to use the martingale $\varphi(X_n)$. Set $\varphi(0) := 0$, $\varphi(1) := 1$. In order to have $Q\varphi(x) = \varphi(x)$ for each $x \in \{1, 2, \dots\}$, we must have

$$p_x \varphi(x+1) + r_x \varphi(x) + q_x \varphi(x-1) = \varphi(x) \,.$$

Using the fact that $r_x = 1 - p_x - q_x$, this can be written

$$\varphi(x+1) - \varphi(x) = \frac{q_x}{p_x}(\varphi(x) - \varphi(x-1)) \,.$$

Since $\varphi(1) - \varphi(0) = 1$, we get

$$\varphi(x) = \sum_{k=0}^{x-1} \frac{q_k q_{k-1} \cdots q_2 q_1}{p_k p_{k-1} \cdots p_2 p_1} \,,$$

where the fraction is defined to be 1 when $k = 0$.

The harmonic function φ, although it is not necessarily bounded, allows to compute hitting probabilities. For example, fix two integers $0 \leq a < b < \infty$, and consider the probability that the chain starting at x, $a < x < b$, hits a before b, i.e. $P_x(T_a < T_b)$ where $T_y := \inf\{n \geq 0 : X_n = y\}$.

Lemma 5.8.3.

$$P_x(T_a < T_b) = \frac{\varphi(b) - \varphi(x)}{\varphi(b) - \varphi(a)} \,. \tag{5.8.11}$$

Proof. With φ at hand, this is a direct application of Theorem 5.8.1, with $A = \{a, b\}$. We first show that $P_x(T_A < \infty) = 1$. Namely, at any time $n \geq 1$, the chain has a probability of reaching (say) b bounded below by $\prod_{y=a+1}^{b} p_y$. Then $P_x(T_A < \infty) = 1$ follows by Lemma 5.3.4.
We modify φ in such a way that it satisfies the modified Dirichlet problem with g on $\{a, b\}$ defined by $g(a) := 1$, $g(b) := 0$. Since $Q\varphi = \varphi$ on (a, b), the same will hold with $\alpha + \beta\varphi$. As can be seen easily, a proper choice of α and β shows that $\widetilde{\varphi}(x) := \frac{\varphi(b)-\varphi(x)}{\varphi(b)-\varphi(a)}$ coincides with g on $\{a, b\}$; $\widetilde{\varphi}$ solves the modified Dirichlet problem. Since we saw in Theorem 5.8.1 that the solution to this problem is unique, one must have $\widetilde{\varphi}(x) = E_x[g(X_{T_A})]$. But

$$E_x[g(X_T)] = P_x(X_T = a)g(a) + P_x(X_T = b)g(b) = P_x(X_T = a) \equiv P_x(T_a < T_b) \,.$$

□

Observe that one could have obtained (5.8.11) by simply applying the Optional Stopping Theorem to the martingale $\varphi(X_n)$, as we did in Section 5.3.2: $E_x[\varphi(X_T)] = E_x[\varphi(X_0)] = \varphi(x)$, but

$$E_x[\varphi(X_{T_A})] = P_x(X_T = a)\varphi(a) + P_x(X_T = b)\varphi(b)\,.$$

Nevertheless, the method we used applies in more general situations, where A contains more than two points.

Let us consider the particular case $a = 0$, $b = L$. Since $\varphi(0) = 0$, (5.8.11) gives

$$P_x(T_0 < T_L) = 1 - \frac{\varphi(x)}{\varphi(L)}\,.$$

Observe that $T_L \geq L - x$ P_x-a.s., and so we have the following characterization of recurrence for the origin:

THEOREM 5.8.2. *Define* $\varphi(\infty) := \lim_{L\to\infty} \varphi(L)$*, possibly infinite. Then*

$$P_x(T_0 < \infty) = 1 - \frac{\varphi(x)}{\varphi(\infty)}\,. \tag{5.8.12}$$

We apply this result in a particular case in Exercise 5.26.

5.8.3. Application: Exit Times for Random Walk. Let $R > 0$, and $I_R := [-R, +R]$. We consider the simple symmetric random walk started at some $x \in I_R$ and study the time it takes to exit I_R:

$$T_R := \inf\{n \geq 0 : S_n \notin I_R\}\,.$$

Of course, we start looking for a useful martingale. Let $\mathcal{F}_n$ denote the natural filtration associated to S_n. Observe that since

$$E[\cos(\lambda S_{n+1})|\mathcal{F}_n] = \frac{1}{2}\cos(\lambda(S_n + 1)) + \frac{1}{2}\cos(\lambda(S_n - 1)) = \cos(\lambda)\cos(\lambda S_n)\,,$$

the sequence

$$Z_n := \frac{\cos(\lambda S_n)}{(\cos\lambda)^n}$$

is a martingale. We assume that $0 < \cos\lambda < 1$. By Lemma 5.3.4, T_R is P_x-almost surely finite, but since Z_n is unbounded, we use the Stopping Theorem for the bounded stopping time $T_R \wedge N$:

$$E_x[Z_{T_R\wedge N}] = E_x[Z_0] = \cos(\lambda x)\,.$$

If we choose $\lambda < \frac{\pi}{2R}$, then $\cos(\lambda x) \geq \cos(\lambda R)$ for all $x \in I_R$ and so

$$Z_{T_R\wedge N} = \frac{\cos(\lambda S_{T_R\wedge N})}{(\cos\lambda)^{T_R\wedge N}} \geq \frac{\cos(\lambda R)}{(\cos\lambda)^{T_R\wedge N}}$$

By defining $\sigma := -\log(\cos\lambda) > 0$, we can therefore write

$$E_x[e^{\sigma T_R \wedge N}] \leq \frac{\cos(\lambda x)}{\cos(\lambda R)}\,.$$

Since T_R is finite, we can take the limit $N \to \infty$ and obtain

$$E_x[e^{\sigma T_R}] \leq \frac{\cos(\lambda x)}{\cos(\lambda R)}\,.$$

One can then estimate the distribution of the exit time by using Chebychev:

$$P_x(T_R > K) = P_x\big(e^{\sigma T_R} > e^{\sigma K}\big) \leq e^{-\sigma K}\frac{\cos(\lambda x)}{\cos(\lambda R)}\,.$$

Taking for example $\lambda = \lambda_R = \frac{\pi}{4R}$, and considering the associated σ_R, we have

$$P_x(T_R > K) \leq \sqrt{2}e^{-\sigma_R K}\,.$$

As can be easily verified, this gives $E_x[T_R] \leq cR^2$ for some $c > 0$. REFAIRE LA MEME CHOSE AVEC L'IDENTITE DE WALD, (VOIR GRIMMET PAGE 494)

5.9. Exercises

Generalities.

EXERCISE 5.1. Durrett p.229 Let $(X_n, \mathcal{F}_n)_{n\geq 1}$ be a martingale. If $\mathcal{G}_n \subset \mathcal{F}_n$ for all $n \geq 1$, show that $(X_n, \mathcal{G}_n)_{n\geq 1}$ is a martingale.

EXERCISE 5.2. Let $(S_n)_{n\geq 1}$ denote the simple random walk on $\{0, 1, 2, \dots\}$ with a reflecting barrier at the origin: $P(S_{n+1} = 1|S_n = 0) = 1$. Determine, in function of p, whether $(S_n)_{n\geq 1}$ is sub-/super-martingale.

EXERCISE 5.3. [**GS05**] p. 338. Let $X_1, X_2, \dots$ be i.i.d. with zero mean and finite variance, satisfying $E[X_{n+1}|X_1, \dots, X_n] = aX_n + bX_{n-1}$ where $0 < a, b < 1$ and $a + b = 1$. Find a value of α for which $Z_n = \alpha X_n + X_{n-1}$ defines a martingale.

EXERCISE 5.4. Consider an increasing sequence of finite countable measurable partitions of Ω: $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \dots$. Let $\mathcal{F}_n$ be the σ-algebra generated by $\mathcal{B}_n$. Let Q be a probability measure on $(\Omega, \mathcal{F})$. Define

$$X_n := \sum_{\substack{B\in\mathcal{B}_n:\\ P(B)>0}} \frac{Q(B)}{P(B)} 1_B\,.$$

Show that $(X_n, \mathcal{F}_n)_{n\geq 1}$ is a supermartingale, and that if Q is absolutely continuous with respect to P, $Q \ll P$, then $(X_n, \mathcal{F}_n)_{n\geq 1}$ is a martingale.

EXERCISE 5.5. [**RS85**] p. 140. Consider the simple random walk S_n with parameter $p \in [0,1]$. Consider the the convex function $\phi(x) = |x|$.

(1) If $p = \frac{1}{2}$, then S_n is a martingale and $\phi(S_n)$ a submartingale.
(2) If $p > \frac{1}{2}$, then S_n is a submartingale. What about $\phi(S_n)$?

EXERCISE 5.6. [**GS05**] p. 334. Let $(X_n)_{n\geq 0}$ be the Branching process with parameter $\lambda > 0$ (see Example 4.1.8). Is (X_n) a martingale? Show that $W_n := \frac{X_n}{E[X_n]}$ and $V_n := \lambda^{X_n}$ are martingales.

EXERCISE 5.7. [**GS05**] p. 335. Let $(X_n)_{n\geq 0}$ be a Markov Chain with state space S and transition matrix Q. Call $\psi : S \to \mathbb{R}$ harmonic if $Q\psi = \psi$. Show that if ψ is harmonic, then $(\psi(X_n))_{n\geq 0}$ is a martingale.

EXERCISE 5.8. Consider the martingale X_n of Exercise 5.1.4. Show that if $\mu \ll \lambda$, then X_n is closed: $X_n = E_\lambda(f|\mathcal{F}_n)$, where f is the Radon-Nikodým density of μ with respect to λ.

Stopping Times, Optional Stopping.

EXERCISE 5.9. If $T \in L^1$, then $P(T < \infty) = 1$.

EXERCISE 5.10. (1) If T is a stopping time and if $k \geq 1$ is any integer then $T \wedge k$ is a stopping time.
(2) If T_1, T_2 are two stopping times, then $T_1 + T_2$, $T_1 \wedge T_2$, $T_1 \vee T_2$ are stopping times.
(3) $\mathcal{F}_T \subset \mathcal{F}$ is a sub-σ-algebra called the stopped σ-field generated by T,
(4) T is $\mathcal{F}_T$-measurable,
(5) If $T_1 \leq T_2$, then $\mathcal{F}_{T_1} \subset \mathcal{F}_{T_2}$.
(6) X_T is $\mathcal{F}_T$-measurable.

EXERCISE 5.11. Consider the random walk on $\mathbb{Z}$. For $I \subset \mathbb{Z}$, show that $T := \inf\{n \geq 1 : S_n \in I\}$ is a stopping time with respect to the natural filtration. Why is $T' := \sup\{n \geq 1 : S_n \in I\}$ *not* a stopping time?

EXERCISE 5.12. Show Wald's Identity. Show that if T is not integrable, or if it is not integrable, then the result is wrong.

EXERCISE 5.13. If S_n denotes the simple random walk on $\mathbb{Z}$, then $N_n = S_n - (p-q)n$ and $M_n = \left(\frac{q}{p}\right)^{S_n}$ are martingales.

EXERCISE 5.14. [**RS85**] p. 141. Consider an independent sequence D_n, $P(D_n = 4^n) = 1 - P(D_n = -4^n) = \frac{1}{2}$. Set $S_0 := 0$, $S_n := D_1 + \cdots + D_n$.

(1) Show that X_n is a martingale.
(2) Let $T := \{n \geq 0 : S_n < 0\}$. Show that T has geometric distribution $P(T = n) = 2^{-n}$, so that $T \in L^1$.
(3) Show that $E[|X_T|] = +\infty$. Is there a contradiction with the Optional Stopping Theorem?

EXERCISE 5.15. ABRACADABRA

Convergence.

EXERCISE 5.16. Show that the Doob-Kolmogorov Inequality implies the Kolmogorov Inequality: If $X_1, X_2, \ldots$ is a sequence of independent random variables, $S_n := X_1 + \ldots X_n$, then

$$P\big(\max_{1\leq k\leq n} |S_k| \geq \lambda\big) \leq \frac{1}{\lambda^2}\sum_{k=1}^{n} \mathrm{var} X_k \,. \tag{5.9.1}$$

EXERCISE 5.17. Durrett p. 234 Give an example of a martingale X_n with $X_n \to -\infty$. Hint: take $X_n = Y_1 + \cdots + Y_n$ with $E[Y_k] = 0$ (the Y_ks not being identically distributed).

EXERCISE 5.18. Durrett p. 235 Let $Y_1, Y_2, \ldots$ be i.i.d. with $Y_k \geq 0$, $P(Y_k = 1) < 1$.

(1) Show that $X_n := Y_1 Y_2 \cdots Y_n$ is a martingale.
(2) Use Theorem 5.5.2 and a contradiction argument to show that $X_n \to 0$ a.s.
(3) Use the Strong Law of Large Numbers to show that $\frac{1}{n}\log X_n \to -c < 0$ a.s.

EXERCISE 5.19. Durrett p. 235 Let (X_n) and (Y_n) be adapted to $(\mathcal{F}_n)$. Assume

$$E[X_{n+1}|\mathcal{F}_n] \leq (1 + Y_n) X_n \,,$$

and $\sum_n Y_n < \infty$ a.s. Prove that X_n converges a.s. to a finite limit by finding a closely related supermartingale to which Theorem 5.5.2 can be applied.

EXERCISE 5.20. Durrett p. 237 Let (X_n) and (Y_n) be positive, in L^1, adapted to $(\mathcal{F}_n)$. Assume

$$E[X_{n+1}|\mathcal{F}_n] \leq X_n + Y_n \,,$$

and $\sum_n Y_n < \infty$ a.s. Prove that X_n converges a.s. to a finite limit. Hint: let $T := \inf\{k \geq 1 : \sum_{i=1}^k Y_i > M\}$ and stop your supermartingale at time T.

Exercise 5.21. Uniform integrability. Show the following affirmations.

(1) Constant sequences are UI.
(2) If X_n is uniformly integrable, then it is bounded in L^1. Show that the contrary is false (Hint: consider $X_n := n1_{[0,\frac{1}{n}]}$).
(3) If X_n is L^p-bounded for $p > 1$, then it is uniformly integrable.
(4) If X_n is such that there exists $Y \geq 0$, $Y \in L^1$, such that $|X_n| \leq Y$, then X_n is uniformly integrable.

CHERCHER DES CONTRE-EXEMPLES.

Exercise 5.22. Prove the Uprcrossing Inequality without introducing the predictable sequence C_n, by writing

$$U_{a,b}^{(N)} \leq \sum_{j\geq 1} Y_{T_j^+ \wedge N} - Y_{T_j^- \wedge N}\,.$$

Exercise 5.23. Le Gall, série d'exercises (avec corrigé). Soit $X_1, X_2, \ldots$ une suite de v.a. indépendantes, telles que $X_k \in L^2$. Soit $M_n := X_1 + \cdots + X_n$. On note $\mathcal{F}_n$ la filtration canonique associée a M_n.

(1) Montrer que M_n est une martingale si et seulement si $E[X_k] = 0$ pour tout k. Dans la suite on suppose que cette condition est toujours satisfaite.
(2) Soit $\sigma_k := E[X_k^2]^{\frac{1}{2}}$, $s_n := \sigma_1^2 + \cdots + \sigma_n^2$. Montrer que $M_n^2 - s_n$ est une martingale.
(3) En supposant que $s := \sum_k \sigma_k^2 < \infty$, montrer que M_n converge p.s. et dans L^2, et que pour tout $a > 0$,
$$P\big(\sup_n |M_n| > a\big) \leq \frac{s}{a^2}\,.$$
(4) Dans la suite, on suppose que les variables X_k ont même distribution, sont à valeurs dans $\{1, 0, -1, -2, \ldots\}$, et que $P(X_1 = 1) > 0$. Soit $\lambda > 0$ et $\psi(\lambda) = \log E[e^{\lambda X_1}]$. Montrer que $\exp(\lambda M_n - n\psi(\lambda))$ est une martingale. Cette martingale est-elle fermée?
(5) Soit $b \geq 1$ et $T_b := \inf\{n \geq 0 : M_n = b\}$. En considérant la martingale positive $b - M_{n\wedge T_b}$, montrer que $T_b < \infty$ presque sûrement.

(6) Montrer que pour tout $\lambda > 0$,

$$E[\exp(-\psi(\lambda)T_b)] = \exp(-\lambda b)\,.$$

En déduire la valeur de $E[\exp(-\alpha T_b)]$ pour tout $\alpha > 0$, en particulier dans le cas où $P(X_1 = 1) = P(X_1 = -1) = \frac{1}{2}$.

EXERCISE 5.24. Durrett p.271 Consider the assymetric random walk on Z: $S_0 := 0$, $S_n := Y_1 + \cdots + Y_n$, where $P(Y_k = +1) = p$, $P(Y_k = -1) = 1 - p = q$, with $p \neq q$.

(1) Assume $0 < p < 1$, define $\varphi(x) := \left(\frac{1-p}{p}\right)^x$. Show that $\varphi(S_n)$ is a martingale.
(2) Let $T_x := \inf\{n \geq 1 : S_n = x\}$, $a < 0 < b$. Show that

$$P(T_a < T_b) = \frac{\varphi(b) - \varphi(0)}{\varphi(b) - \varphi(a)}\,.$$

Hint: define $T' := T_a \wedge T_b$. Show that T' is almost surely finite and apply the Optional Stopping Theorem.
(3) Assume $\frac{1}{2} < p < 1$. If $a < 0$, show that $P(T_a < \infty) = \left(\frac{1-p}{p}\right)^{|a|}$, if $b > 0$ show that $P(T_b < \infty) = 1$.
(4) Assume $\frac{1}{2} < p < 1$. If $b > 0$, show that $E[T_b] = \frac{b}{2p-1}$. Hint: consider the martingale $Z_n := S_n - (p - q)n$.

Compare these results with Exercise 2.9.

EXERCISE 5.25. Voir Williams. Polya's urn model.

EXERCISE 5.26. Birth and Death chains, Durrett p. 292. Consider the markov chain $(S_n)_{n\geq 0}$ with state space $\{0, 1, 2, \dots\}$ and transition matrix

$$Q(x, x+1) = p_x\,,\ Q(x, x-1) = q_x\,,\ Q(x, x) = r_x\,,$$

with $p_x + q_x + r_x = 1$ and $q_0 := 0$. Let $T := \inf\{n \geq 0 : S_n = 0\}$. We study $P_x(T < \infty)$ in the following situations.

(1) Assume $p_x = p$, $r_x = 0$, $q_x = q = 1 - p$. Let $x \geq 1$. Show that

$$P_x(T < \infty) \begin{cases} = 1 & \text{if } p \leq \frac{1}{2}\,, \\ = \left(\frac{1-p}{p}\right)^x & \text{if } p > \frac{1}{2}\,. \end{cases}$$

(2) Assume $p_x = \frac{1}{2} + \epsilon_x$, $r_x = 0$, $q_x = \frac{1}{2} - \epsilon_x$, where $\epsilon_x \to 0$ as $\epsilon_x \sim \beta x^{-\alpha}$. Show that

$$\text{The chain is } \begin{cases} \text{recurrent if } \alpha > 1\,, \\ \text{transient if } \alpha < 1\,, \\ \text{recurrent if } \alpha = 1 \text{ and } \beta < \frac{1}{4}\,, \\ \text{transient if } \alpha = 1 \text{ and } \beta > \frac{1}{4}\,. \end{cases}$$

EXERCISE 5.27. [**Chu01**] p 371. voir aussi Grimmett p. 504. Likelihood ratios. Let $(X_n)_{n\geq 1}$, $(Y_n)_{n\geq 1}$ be sequences of random variables with respective N-dimensional distributions p_N and q_N. Define

$$Z_n := \frac{q_n(Y_1, \ldots, Y_n)}{p_n(X_1, \ldots, X_n)}$$

if the denominator is positive, zero otherwise. Show that $(Z_n)_{n\geq 1}$ is a supermartingale that converges a.e.

EXERCISE 5.28. Dacunha Castelle Ex. XI.13 p.199., Ex. XI.16 p.203. Martingales and Markov Chains. Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix Q.

(1) Show that if f and g are superharmonic, then $f \wedge g$ is superharmonic.
(2) If f is (super)harmonic, show that $(f(X_n))_{n\geq 0}$ is a (super)martingale with respect to the natural filtration and to P_μ for any initial distribution μ.
(3) If the chain is irreducible recurrent, show that all superharmonic functions are constant.
(4) Assume the chain is transient. Consider the potential

$$U(x, y) := \sum_{n\geq 0} Q^{(n)}(x, y)\,.$$

For a function $f : S \to \mathbb{R}$, we call potential of f the function $Uf : S \to \mathbb{R}$ defined by

$$Uf(x) := \sum_{y\in S} U(x, y)f(y)\,.$$

Show that if f is superharmonic, then $\hat{f} := \lim_{n\to\infty} Q^{(n)}f$ exists and defines a harmonic function. Show that for any superharmonic function f there exists a decomposition $f = h + Ug$, where h is harmonic. Show that this decomposition is unique.

(5) Let $A \subset S$ and let T_A the first visit of the chain at A. Show that $\varphi_A(x) := P_x(T_A < \infty)$ is superharmonic. Interpret the decomposition of $\phi_A = h + Ug$.

EXERCISE 5.29. Let $(X_n)_{n\geq 0}$ be a submartingale adapted to a filtration $(\mathcal{F}_n)_{n\geq 0}$. Define $A_0 := 0$, and

$$A_{n+1} := A_n + E[X_{n+1} - X_n | \mathcal{F}_n] .$$

(1) Show that $(A_n)_{n\geq 0}$ is almost surely increasing, that it is predictable, and that $M_n := X_n - A_n$ is a martingale. Show that these properties and $A_0 = 0$ characterize A_n (up to almost everywhere equivalence). The decomposition

$$X_n = M_n + A_n$$

is called the Doob Decomposition of X_n.

(2) Let $a > 0$ and let $T_a := \inf\{n \geq 0 : A_{n+1}\}$. Show that T_a is a stopping time, and that $E[X_{n\wedge T_a}] \leq a$.

(3) Show that X_n converges almost surely to a finite limit on the set $\{T_a = +\infty\}$. Conclude that if $A_\infty := \lim_n A_n$, then X_n converges to a finite limit, almost surely on the set $\{A_\infty < \infty\}$.

(4) Assume $E\big[\sup_{n\geq 1} |X_{n+1} - X_n|\big] < \infty$. Show that the following three conditions are almost surely equivalent:
(a) X_n converges to a finite limit
(b) X_n is bounded
(c) $A_\infty < \infty$.
Hint: introduce the stopping time $\tau_a := \inf\{n \geq 0 : X_n > a\}$.

EXERCISE 5.30. Durrett p. 235 Let $X_n = \sum_{m\leq n} 1_{B_m}$, where $B_m \in \mathcal{F}_m$. Compute the Doob decomposition of X_n.

EXERCISE 5.31. Bovier. Let X_n be a Markov chain with transition matrix Q on a countable space S. Let $f : S \to \mathbb{R}$ be bounded. Show that that the Doob decomposition of the process $f(X_n)$ is given by $f(X_n) = M_n + A_n$, where

$$M_n = f(X_n) - f(X_0) - \sum_{k=0}^{n-1} h(X_k) , \qquad (5.9.2)$$

where $h := Qf - f$.

EXERCISE 5.32. Bovier. Let S be a countable set, Q a transition matrix on S. Let X_n be an S-valued process adapted to a filtration $\mathcal{F}_n$. Show that X_n is a Markov chain with transition matrix Q if and only if for all bounded function $f : S \to \mathbb{R}$, the process defined in (5.9.2) is a martingale.

CHAPTER 6

Stationary and Ergodic Processes

Let $(\Omega, \mathcal{F}, P)$ denote an arbitrary probability space. In this chapter we consider discrete time stochastic processes $(X_n)_{n\geq 0}$ with the property that the law of $(X_n)_{n\geq 0}$ is the same as that of $(X_{n+k})_{n\geq 0}$. In other words, these sequences are such that the randomness is the same at any place the sequence is looked at.

6.1. Definition and Examples

DEFINITION 6.1.1. *A stochastic process $(X_n)_{n\geq 0}$ is* stationary *if and only if for all $n \geq 0$ and all $k \geq 1$, $(X_n, \ldots, X_{n+k})$ has the same distribution as $(X_0, \ldots, X_k)$. That is, for any Borel set $B \in \mathcal{B}(\mathbb{R}^k)$,*

$$P((X_n, \ldots, X_{n+k}) \in B) = P((X_0, \ldots, X_k) \in B) . \tag{6.1.1}$$

In the case where X_n takes values in a finite or countable alphabet S, stationarity means that for all $n \geq 0$ and all $k \geq 1$,

$$P(X_n = a_0, \ldots, X_{n+k} = a_k) = P(X_0 = a_0, \ldots, X_k = a_k) , \tag{6.1.2}$$

for all $a_0, \ldots, a_k \in S$. Assume (6.1.1) holds for $n = 1$ and all $k \geq 1$. Then for all $B \in \mathcal{B}(\mathbb{R}^k)$,

$$\begin{aligned} P((X_2, \ldots, X_{2+k}) \in B) &= P((X_1, X_2, \ldots, X_{2+k}) \in \mathbb{R} \times B) \\ &= P((X_0, X_1, \ldots, X_{1+k}) \in \mathbb{R} \times B) \\ &= P((X_1, \ldots, X_{1+k}) \in B) \\ &= P((X_0, \ldots, X_k) \in B) . \end{aligned}$$

To verify stationarity, it therefore suffices to test (6.1.1) for $n = 1$. Let us consider a few examples.

EXAMPLE 6.1.1. In the case where the variables X_n are independent and identically distributed, we get

$$\begin{aligned} P(X_n \in B_0, \ldots, X_{n+k} \in B_k) &= P(X_n \in B_0) \ldots P(X_{n+k} \in B_k) \\ &= P(X_0 \in B_0) \ldots P(X_k \in B_k) \\ &= P(X_0 \in B_0, \ldots, X_k \in B_k) , \end{aligned}$$

which implies that the sequence is stationary.

EXAMPLE 6.1.2. Let the sequence $(X_n)_{n\geq 0}$ form a Markov chain taking values in a countable set S, with transition matrix Q. Let assume that there exists an invariant distribution π, $\pi Q = \pi$, and that X_0 has distribution π. Let $a_0, \dots, a_k \in S$. By Lemma 4.1.1,

$$\begin{aligned} P(X_1 = a_0, \dots, X_{1+k} = a_k) &= \sum_{b_0 \in S} P(X_0 = b_0, X_1 = a_0, \dots, X_{1+k} = a_k) \\ &= \sum_{b_0 \in S} \pi(b_0) Q(b_0, a_0) Q(a_0, a_1) \dots Q(a_{k-1}, a_k) \\ &= \pi Q(a_0) Q(a_0, a_1) \dots Q(a_{k-1}, a_k) \\ &= \pi(a_0) Q(a_0, a_1) \dots Q(a_{k-1}, a_k) \\ &\equiv P(X_0 = a_0, \dots, X_k = a_k)\,, \end{aligned}$$

which shows that $(X_n)_{n\geq 0}$ is stationary. Observe that if π isn't invariant, this might not hold.

EXAMPLE 6.1.3. Let $T : \Omega \to \Omega$ be a measurable (i.e. $T^{-1}A \in \mathcal{F}$ for all $A \in \mathcal{F}$) transformation which preserves P, i.e. such that

$$P(T^{-1}A) = P(A) \quad \forall A \in \mathcal{F}\,.$$

Let $Y : \Omega \to \mathbb{R}$ be a random variable. For all $n \in \mathbb{N}$, define $X_n := Y \circ T^n$, where T^0 is the identity, and $T^n := T \circ T^{n-1}$. We verify that $(X_n)_{n\geq 0}$ is stationary:

$$\begin{aligned} P((X_1, \dots, X_{1+k}) \in B) &= P((Y \circ T, \dots, Y \circ T^{1+k}) \in B) \\ &= P(T^{-1}(Y, \dots, Y \circ T^k) \in B) \\ &= P((Y, \dots, Y \circ T^k) \in B) \\ &= P((X_0, \dots, X_k) \in B)\,, \end{aligned}$$

where we used invariance of P in the third equality.

This last example shows that stationary sequences can be constructed from a measure preserving map $T : \Omega \to \Omega$ and from an arbitrary random variable Y. We will now show that this example is actually the only one, since *any stationary sequence* $(X_n)_{n\geq 0}$ *can be considered as constructed on a probability space* $(\Omega', \mathcal{F}', P')$, *such that* $X_n = Y' \circ T'^n$, *where* Y' *is a random variable on* Ω' *and* $T' : \Omega' \to \Omega'$ *is a map preserving* P'. Namely, let Ω' be the product $\Omega' := \mathbb{R}^{\mathbb{N}}$, equipped with the σ-algebra $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ (see Chapter 3). Consider for each $n \geq 1$ the measure μ_n on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ defined on rectangles $B_1 \times \dots \times B_n \in \mathcal{B}(\mathbb{R}^n)$

by

$$\mu_n(B_1 \times \cdots \times B_n) := P(X_1 \in B_1, \dots, X_n \in B_n)\,.$$

By Kolmogorov' s Extension Theorem 3.1.2, there exists a unique P' on $(\Omega', \mathcal{F}')$ having μ_n as marginals. Define T' to be the shift

$$T(\omega'_1, \omega'_2, \omega'_3, \dots) := (\omega'_2, \omega'_3, \dots)\,.$$

By the stationarity of (X_n) under P, P' is invariant under T'. Let $Y'(\omega') := \omega'_1$, $X_n := Y' \circ T^n$. This proves the claim.

Using the above representation, the study of a stationary sequence

$$X_1(\omega'), X_2(\omega'), \dots \tag{6.1.3}$$

can therefore always be reduced to the study of the values taken by a fixed random variable Y along the trajectory of a point $\omega' \in \Omega'$ under iteration of the map T':

$$\omega', T\omega', T^2\omega', \dots \tag{6.1.4}$$

This slightly changes the point of view of a random experiment: in (6.1.3), the variables X_n are realized *simultaneously* as functions of the random experiment ω, but in (6.1.4) ω' is to be considered as a random initial condition, and the sequence of observations is deterministic, along the trajectory of ω'.

One of the objects of interest, in the statistical properties of the sequence X_n, is the sequence of empirical means

$$\frac{X_1 + \cdots + X_n}{n}\,,$$

which, when expressed in the above representation, take the form of a Birkhoff Sum:

$$\frac{1}{n}\sum_{k=0}^{n-1} Y \circ T^k\,. \tag{6.1.5}$$

It turns out that when Y is integrable and when T preserves P, this last sum converges for P-almost all ω, as will be seen in the Ergodic Theorem in the next section.

Since the underlying invariance structure of stationary sequences can be nicely understood by considering them as trajectories of a discrete dynamical system; we will now describe the general setting of measurable dynamics.

Let $(\Omega, \mathcal{F})$ be a measurable space. In this section, we denote a generic probability measure on $(\Omega, \mathcal{F})$ by μ. Consider a map $T : \Omega \to \Omega$, which we assume to be measurable, i.e. such that $T^{-1}A \in \mathcal{F}$ for all $A \in \mathcal{F}$.

Definition 6.1.2. *A measure μ on $(\Omega, \mathcal{F})$ is called* invariant (under T) *if*

$$\mu(T^{-1}A) = \mu(A)\,, \qquad \textit{for all } A \in \mathcal{F}\,. \tag{6.1.6}$$

Observe that if T is invertible, the previous definition is equivalent to

$$\mu(TA) = \mu(A)\,, \qquad \text{for all } A \in \mathcal{F}\,. \tag{6.1.7}$$

The purpose of Ergodic Theory is the study of the existence and properties of measures invariant under a given T. We obtain a discrete time dynamical system by considering the iterations of T, i.e. by constructing the orbit of a point $\omega \in \Omega$: $\omega, T\omega, T^2\omega, \ldots$ Therefore, we call the triple $(\Omega, \mathcal{F}, T)$ a measurable discrete dynamical system.

Example 6.1.4. Let $\Omega = [0, 1)$ with the Borel σ-algebra $\mathcal{B}([0, 1))$ and the Lebesgue measure, denoted λ. If we define the translation modulo 1, $T : [0, 1) \to [0, 1)$ by $Tx := x + 1 \mod 1$, then λ is invariant.

Example 6.1.5. Let A be a finite set. Shift spaces with alphabet A are obtained by considering either of the product spaces $\Omega = \mathsf{A}^{\mathbb{N}}$ (the one-sided shift) or $\Omega = \mathsf{A}^{\mathbb{Z}}$ (the two-sided shift), which we write generically as $\Omega = \mathsf{A}^{\mathbb{T}}$. The σ-algebras on these sets are defined in the usual way. A (thin) cylinder is a subset of Ω of the form

$$C_\Lambda(a_1, \ldots, a_{|\Lambda|}) = \{\omega \in \Omega : \omega_i = a_i,\, \forall i \in \Lambda\}\,, \tag{6.1.8}$$

where Λ is a finite subset of $\mathbb{T}$, and $a_1, \ldots, a_{|\Lambda|} \in \mathsf{A}$. Thin cylinders generate an algebra $\mathcal{C}$ on Ω called the algebra of cylinders, and a σ-algebra $\mathcal{F} = \sigma(\mathcal{C})$. The shift is the map $T : \Omega \to \Omega$, where

$$(T\omega)_i := \omega_{i+1}\,, \quad \forall i \in \mathbb{T}\,. \tag{6.1.9}$$

Observe that when $\mathbb{T} = \mathbb{Z}$, the shift is invertible. The simplest example of a measure μ on $(\Omega, \mathcal{F})$ which is invariant under the shift is that of a product measure: $\mu = p^{\mathbb{T}}$, where p is any probability distribution on A. To verify that this measure is invariant under the shift, it suffices to observe that

$$\mu(T^{-1}C_\Lambda(a_1, \ldots, a_{|\Lambda|})) = \mu(C_\Lambda(a_1, \ldots, a_{|\Lambda|}))$$

for all cylinder. This implies (see Exercise 6.3) that μ is invariant, and the resulting dyanmical system is called a Bernoulli shift. Another example of invariant measure is provided by a Markov shift, which we already enountered in the introduction. Let Q be a stochastic matrix on A. Assume there exists an invariant distribution ν on A, such that $\nu Q = \nu$. For any thin cylinder $C_{[1,n]}(a_1, \dots, a_n)$, define

$$\mu(C_{[1,n]}(a_1, \dots, a_n)) := \nu(a_1) \prod_{i=1}^{n-1} Q(a_i, a_{i+1}) \,. \tag{6.1.10}$$

Then again by Kolmogorov's Extension Theorem, μ extends uniquely to a measure on $\mathsf{A}^{\mathbb{T}}$, which is invariant under the shift T.

Assume that we have a shift space with $\mathbb{T} = \mathbb{N}$, and a measure μ invariant under the shift. For each $n \in \mathbb{N}$, consider the random variable $X_n(\omega) := \omega_n$. By defining the distribution

$$P(X_1 = a_1, \dots, X_n = a_n) := \mu(C_{[1,n]}(a_1, \dots, a_n))$$

we see, together with what was said in the previous section, that constructing an invariant measure on a shift space is equivalent to the construction of a stationary stochastic process (when both have the same finite alphabet).

Throughout the chapter, we denote the measurable space by $(\Omega, \mathcal{F})$, and the dynamic T is considered as fixed. Therefore, invariant objects will always be defined with respect to T. The random variables on $(\Omega, \mathcal{F})$ will be denoted f or g rather than X or Y.

Lemma 6.1.1. *Let μ be invariant under T. Then, for all $f \in L^1(\mu)$, we have the Change of Variable Formula:*

$$\int f d\mu = \int f \circ T d\mu \,. \tag{6.1.11}$$

Proof. If $f = 1_A$, the indicator of an event $A \in \mathcal{F}$, we have

$$\int f d\mu = \mu(A) = \mu(T^{-1}A) = \int f \circ T d\mu \,.$$

For $f \in L^1(\mu)$, use a standard approximation procedure. □

6.2. The Ergodic Theorem

A central question about the orbits of a dynamical system is the following: with what frequence (if any) does the orbit of a point $\omega \in \Omega$

visit a given measurable set $A \in \mathcal{F}$? That is, how does the ratio

$$\frac{1}{n}\sharp\{0 \leq k \leq n-1 : T^k\omega \in A\} \equiv \frac{1}{n}\sum_{k=0}^{n-1} 1_A(T^k\omega)$$

behave for large n? This average is a particular case of (6.1.5) with $Y = 1_A$. Observe that in general the variables $(1_A \circ T^k)_{k\geq 1}$ have no reason of being independent, and so the large-n behaviour of these averages cannot be studied using a simple Law of Large Numbers. But as we will see, when an invariant measure μ exists, the Ergodic Theorem of Birkhoff will imply that these frequencies exist when $n \to \infty$, for μ-almost-all initial condition ω. This theorem actually holds not only for indicator functions but for any integrable function f. Before stating the Ergodic Theorem, we take a look at uniformly distributed sequences, which give an idea of the kind of results we are heading to.

6.2.1. A Parenthesis: Uniformly Distributed Sequences. Uniformly distributed sequences are kind of ideal trajectories for the evolution of a dynamical system, in the sense that they visit the whole phase space in a very regular way. Later, systems whose trajectories behave in this way will be called *ergodic*. We describe these sequences following the first chapter of the books of Kuipers and Niederreiter [**KH74**] and Rauzy [**Rau76**].

Consider a sequence $x_1, x_2, \dots$, with $x_n \in [0,1]$. For $I \subset [0,1]$, denote by $N_I(n)$ the number of elements of the set $\{x_1, \dots, x_n\}$ which are contained in I.

DEFINITION 6.2.1. *A sequence $x_n \in [0,1]$, is uniformly distributed if*

$$\lim_{n\to\infty} \frac{N_{[a,b)}(n)}{n} = b - a\,, \tag{6.2.1}$$

for all $0 \leq a < b \leq 1$.

If one considers a sequence as the trajectory of a discrete dynamical system, then uniform distribution means that the fraction of time the trajectory spends in any interval is equal to the size of this interval. Notice that (6.2.1) can be written as

$$\frac{1}{n}\sum_{k=1}^{n} 1_{[a,b)}(x_k) \to \int 1_{[a,b)}(x)dx\,. \tag{6.2.2}$$

(We always denote $\int_0^1$ by $\int$, and dx is Lebesgue measure.) This suggests the following criterium.

THEOREM 6.2.1. *A sequence $x_n \in [0,1]$, is uniformly distributed if and only if for all continuous $f : [0,1] \to \mathbb{R}$,*

$$\frac{1}{n}\sum_{k=1}^{n} f(x_k) \to \int_a^b f(x)dx\,. \tag{6.2.3}$$

The same holds if $f : [0,1) \to \mathbb{C}$ is continuous and $f(x+1 \mod 1) = f(x)$.

PROOF. Assume (x_n) is uniformly distributed, take f continuous on $[0,1]$. Let $\epsilon > 0$. Approximate f by two step functions $g_1 \le f \le g_2$ such that $\int (g_2 - g_1)dx \le \epsilon$. Then

$$\begin{aligned}
\int f(x)dx - \epsilon \le \int g_1(x)dx &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} g_1(x_k) \\
&\le \liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(x_k) \\
&\le \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(x_k) \\
&\le \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} g_2(x_k) \\
&= \int g_2(x)dx \le \int f(x)dx + \epsilon\,.
\end{aligned}$$

Conversely, assume (6.2.3) holds for all continuous f. Consider some interval $[a,b)$. Fix $\epsilon > 0$. Then the indicator of $[a,b)$ can be approximated by two continuous functions $f_1 \le 1_{[a,b)} \le f_2$ such that $\int (f_2 - f_1)dx \le \epsilon$, and we have

$$\begin{aligned}
b - a - \epsilon \le \int f_2 dx - \epsilon \le \int f_1 dx &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f_1(x_k) \\
\le \liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} 1_{[a,b)}(x_k) &\le \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} 1_{[a,b)}(x_k) \\
\le \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f_2(x_k) &= \int f_2 dx \le \int f_1 dx + \epsilon \le b - a + \epsilon\,.
\end{aligned}$$

Since ϵ is arbitrary, this shows that (x_n) is uniformly distributed. $\square$

The previous theorem allows, in principle, to verify if a sequence is uniformly distributed. Nevertheless, the fact that one must test (6.2.3) for *all* continuous functions makes the criterium rather difficult to use in the practice. The following result shows that it is actually sufficient to verify convergence for functions of the form $f_k(x) = e^{2\pi i k x}$. These satisfy $\int f_k dx = 0$ for all $k \neq 0$.

THEOREM 6.2.2 (Weyl's Criterium). *A sequence $x_1, x_2, \dots$, where $x_n \in [0,1]$, is uniformly distributed if and only if for all integer $l \neq 0$,*

$$\frac{1}{n}\sum_{k=1}^{n} e^{2\pi i l x_k} \to 0\,. \tag{6.2.4}$$

PROOF. If (x_n) is uniformly distributed then (6.2.4) follows from Theorem 6.2.1. So suppose (6.2.4) holds for all integer $l \neq 0$. Take some continuous $f : [0,1) \to \mathbb{C}$ satisfying $f(x+1 \mod 1) = f(x)$. Fix $\epsilon > 0$. By Weierstrass' Approximation Theorem, there exists a finite linear combination of functions of the type $e^{2\pi i l x}$, denoted $\phi(x)$, such that $\sup_{0\leq x\leq 1}|f(x) - \phi(x)| \leq \epsilon$. Observe that $\int \phi dx = 0$.

$$\Big|\int f(x)dx - \frac{1}{n}\sum_{k=1}^{n} f(x_k)\Big| \leq \Big|\int f(x)dx - \int \phi(x)dx\Big| \\ + \Big|\int \phi(x)dx - \frac{1}{n}\sum_{k=1}^{n}\phi(x_k)\Big| + \Big|\frac{1}{n}\sum_{k=1}^{n}\phi(x_k) - \frac{1}{n}\sum_{k=1}^{n} f(x_k)\Big|\,.$$

The first and last term are $\leq \epsilon$, and the second goes to zero because of (6.2.4). $\square$

We will use Weyl's Criterium to study the following simple problem. Let $c \in S^1 = \{z \in \mathbb{C} : |z| = 1\}$. Define $z_n := c^n$. Is $z_1, z_2, \dots$ uniformly distributed on S^1? Observe that $c = e^{2\pi i\alpha}$, where $\alpha \in [0,1]$. One can thus write $z_n = e^{2\pi i n\alpha} = e^{2\pi i x_n}$, where $x_n = \alpha n - \lfloor \alpha n\rfloor \in [0,1]$. Now, we show

LEMMA 6.2.1. *If α is irrational, then x_n is uniformly distributed.*

PROOF. We use Weyl's Criterium.

$$\Big|\frac{1}{n}\sum_{k=1}^{n} e^{2\pi i l\{\alpha k\}}\Big| = \Big|\frac{1}{n}\sum_{k=1}^{n} e^{2\pi i l\alpha k}\Big| = \frac{1}{n}\Big|\frac{e^{2\pi i l\alpha(n+1)} - 1}{e^{2\pi i l\alpha} - 1}\Big| \leq \frac{1}{n}\frac{2}{|e^{2\pi i l\alpha} - 1|}\,, \tag{6.2.5}$$

which tends to zero when $n \to \infty$ for all integer l since $e^{2\pi i l\alpha} - 1$ is never zero, whatever the value of l, by the irrationality of α. $\square$

The sequence z_n above is obtained by considering the orbit of 1 under multiplication by $c = e^{2\pi i\alpha}$, which is a rotation of angle α on the unit circle. In general, many sequences of interest are obtained in this way, i.e. by iterating a map: $x_{n+1} = Tx_n$, which is exactly the situation we are interested in. The difference is that we consider orbits on a general measurable space $(\Omega, \mathcal{F})$. The analog of uniformity will be defined in terms of an invariant measure with respect to T (for sequences in $[0, 1]$, this is Lebesgue measure $\lambda(a, b) = b - a$), and the Ergodic Theorem will be the ingredient giving conditions under which uniformity is guaranteed.

6.2.2. The Ergodic Theorem of Birkhoff. We go back to the study of the general study of measures invariant under a measurable map T. Call a set $A \in \mathcal{F}$ strictly invariant if $T^{-1}A = A$. The collection of strictly invariant sets forms a σ-algebra which we denote by $\mathcal{I}$. One of the consequences of the following theorem is a construction of the conditional expectation with respect to $\mathcal{I}$.

THEOREM 6.2.3 (Birkhoff's Ergodic Theorem). *Let μ be invariant under T. For any $f \in L^1(\mu)$, the limit*

$$\widehat{f}(\omega) := \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (f \circ T^k)(\omega) \tag{6.2.6}$$

exists for μ-almost all $\omega \in \Omega$ and satisfies the following properties:

(1) *$\widehat{f} \in L^1(\mu)$, and is μ-almost-invariant: $\widehat{f} \circ T = \widehat{f}$ μ-almost surely.*
(2) *The convergence in (6.2.6) holds also in $L^1(\mu)$.*
(3) *For each strictly invariant $A \in \mathcal{F}$,*

$$\int_A \widehat{f}\, d\mu = \int_A f\, d\mu\,. \tag{6.2.7}$$

In particular, $\widehat{f}$ is a version of $E[f|\mathcal{I}]$, and $\int \widehat{f} d\mu = \int f d\mu$.

The proof relies on the following result, usually called Maximal Ergodic Theorem:

PROPOSITION 6.2.1. *Let μ be invariant under T. Let $f \in L^1(\mu)$ and define, for each $\lambda \in \mathbb{R}$, the event*

$$N_\lambda = N_\lambda(f) := \Big\{\omega \in \Omega : \sup_{n\geq 1} \frac{1}{n}\sum_{k=0}^{n-1}(f \circ T^k)(\omega) > \lambda\Big\}. \tag{6.2.8}$$

Then the following holds:

$$\int_{N_\lambda} f \, d\mu \geq \lambda\mu(N_\lambda). \tag{6.2.9}$$

Moreover, if A is such that $T^{-1}A = A$, then

$$\int_{A\cap N_\lambda} f \, d\mu \geq \lambda\mu(A \cap N_\lambda). \tag{6.2.10}$$

The inequality (6.2.9) can be understood by assuming $f \geq 0$; then, (6.2.9) yields

$$\mu(N_\lambda) \leq \frac{1}{\lambda}\|f\|_1, \tag{6.2.11}$$

which is a useful concentration property: the set of initial conditions ω for which the empirical averages $\frac{1}{n}\sum_{k=0}^{n-1}(f \circ T^k)(\omega)$ take large values (along the orbit of ω) has small probability.

PROOF. To obtain (6.2.10), we use (6.2.9) with $\lambda = 0$ and $\tilde{f} := (f - \eta)1_A$. As can be seen verified easily, $T^{-1}A = A$ implies $N_0(\tilde{f}) = N_\eta(f) \cap A$, which gives

$$0 \leq \int_{N_0(\tilde{f})} \tilde{f} \, d\mu = \int_{N_\eta(f)\cap A} f \, d\mu - \eta\mu(N_\eta(f) \cap A).$$

This shows (6.2.10). To obtain (6.2.9), it suffices to show that

$$\int_E f \, d\mu \geq 0, \tag{6.2.12}$$

where

$$E = E[f] := \Big\{\omega \in \Omega : \sup_{n\geq 1} \sum_{k=0}^{n-1}(f \circ T^k)(\omega) > 0\Big\}. \tag{6.2.13}$$

Namely, by observing that $E[f - \lambda] = N_\lambda(f)$, (6.2.9) follows immediately by taking $f - \lambda$ in place of f in (6.2.12). To show (6.2.12), we express E as a limit of increasing events E_n. Set $S_0 f(\omega) := 0$ and, for all $k \geq 1$,

$$S_k f(\omega) := \sum_{j=0}^{k-1} (f \circ T^j)(\omega)\,.$$

Then, for all $n \geq 0$, set

$$M_n f(\omega) := \max_{0 \leq k \leq n} S_k f(\omega)\,.$$

We clearly have $M_{n+1} f \geq M_n f \geq \cdots \geq 0$ and so by defining

$$E_n := \left\{ \omega \in \Omega : M_n f(\omega) > 0 \right\},$$

one has $E_n \subset E_{n+1}$. Moreover, $E_n \subset E$, which implies $\bigcup_n E_n \subset E$. Inversely, if $\omega \in E$ then there exists some integer n such that $S_n f(\omega) > 0$, i.e. $\omega \in E_n$. That is $E \subset \bigcup_n E_n$, i.e. $E_n \nearrow E$. Therefore, if one can show that

$$\int_{E_n} f \, d\mu \geq 0\,, \tag{6.2.14}$$

then Dominated Convergence yields (6.2.12). Now, since $M_n f \geq S_k f$ for all $0 \leq k \leq n$,

$$M_n f(T\omega) \geq S_k f(T\omega) = S_{k+1} f(\omega) - f(\omega) \quad \forall 0 \leq k \leq n\,.$$

This last inequality can be rewritten

$$f(\omega) + M_n f(T\omega) \geq S_k f(\omega) \quad \forall 1 \leq k \leq n+1\,.$$

In particular,

$$f(\omega) + M_n f(T\omega) \geq \max_{1 \leq k \leq n} S_k f(\omega)\,.$$

Let us integrate this last expression over E_n:

$$\int_{E_n} f \, d\mu + \int_{E_n} M_n f \circ T \, d\mu \geq \int_{E_n} \max_{1 \leq k \leq n} S_k f \, d\mu = \int_{E_n} M_n f \, d\mu\,,$$

where in the last expression we used the fact that when $\omega \in E_n$, then $S_k f(\omega) > 0$ for some $0 \leq k \leq n$, which in turn, since $S_0 f(\omega) = 0$ by definition, implies that $\max_{1 \leq k \leq n} S_k f(\omega) = \max_{0 \leq k \leq n} S_k f(\omega) =$

$M_n f(\omega)$. Therefore,

$$\begin{aligned}\int_{E_n} f\,d\mu &\geq \int_{E_n} M_n f\,d\mu - \int_{E_n} M_n f \circ T\,d\mu \\ &= \int M_n f\,d\mu - \int_{E_n} M_n f \circ T\,d\mu \\ &\geq \int M_n f\,d\mu - \int M_n f \circ T\,d\mu = 0\end{aligned}$$

In the first equality we used the fact that if $\omega \in E_n^c$ then $M_n f(\omega) = 0$. In the inequality we used the fact that $M_n f \geq 0$. Finally, we used the invariance of μ with respect to T and the Change of Variable Formula of Lemma 6.1.1. This shows (6.2.14), and consequently, (6.2.9), which finishes the proof of the proposition. □

PROOF OF THE ERGODIC THEOREM: We study the set of points $\omega \in \Omega$ at which the limit (6.2.6) doesn't exist, i.e. at which $f_*(\omega) < f^*(\omega)$, where

$$f_*(\omega) := \liminf_n \mathcal{A}_n(\omega)\,, \quad f^*(\omega) := \limsup_n \mathcal{A}_n(\omega)\,,$$

are the empirical averages (or Birkhoff sums)

$$\mathcal{A}_n f := \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k\,.$$

As can be seen easily, these functions are invariant: $f_* \circ T = f_*$, $f^* \circ T = f^*$. For two rationals $\alpha < \beta$, consider the invariant set

$$A(\alpha,\beta) := \big\{\omega \in \Omega : f_*(\omega) < \alpha < \beta < f^*(\omega)\big\}\,.$$

Consider the set $N_\beta(f)$, defined as in (6.2.8). Since $A(\alpha,\beta) \subset N_\beta(f)$, we have, by the Maximal Ergodic Theorem,

$$\begin{aligned}\beta\mu(A(\alpha,\beta)) &= \beta\mu(A(\alpha,\beta) \cap N_\beta(f)) \\ &\leq \int_{N_\beta(f)\cap A(\alpha,\beta)} f\,d\mu = \int_{A(\alpha,\beta)} f\,d\mu \qquad (6.2.15)\end{aligned}$$

Doing the same with the set $N_{-\alpha}(-f)$ we get

$$\begin{aligned}-\alpha\mu(A(\alpha,\beta)) &= -\alpha\mu(A(\alpha,\beta) \cap N_{-\alpha}(-f)) \\ &\leq -\int_{N_{-\alpha}(-f)\cap A(\alpha,\beta)} f\,d\mu = -\int_{A(\alpha,\beta)} f\,d\mu \qquad (6.2.16)\end{aligned}$$

Therefore,

$$\beta\mu(A(\alpha,\beta)) \leq \int_{A(\alpha,\beta)} f\, d\mu \leq \alpha\mu(A(\alpha,\beta))\,,$$

which, since $\alpha < \beta$, is possible only if $\mu(A(\alpha,\beta)) = 0$. Since this holds for all rationals $\alpha < \beta$, this shows the first part of the Ergodic Theorem, namely the almost sure convergence of the means defining $\widehat{f}$ in (6.2.6). At points ω' where the limit doesn't exist, we set $\widehat{f}(\omega') := f_*(\omega')$. To see that $\widehat{f}$ is integrable, we use Fatou's Lemma and the Change of Variable Formula:

$$\begin{aligned}\int |\widehat{f}|\, d\mu &\leq \int \liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} |f\circ T^k|\, d\mu \\ &\leq \liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \int |f\circ T^k|\, d\mu = \int |f|\, d\mu\,.\end{aligned}$$

To see that $\widehat{f}$ is almost surely invariant, write

$$\frac{n+1}{n}\mathcal{A}_{n+1}f(\omega) - \mathcal{A}_n f(T\omega) = \frac{f(\omega)}{n}$$

Since $f \in L^1(\mu)$, this tends to zero almost everywhere. Therefore, $\widehat{f}\circ T$ exists and equals $\widehat{f}$ almost everywhere. To verify that $\mathcal{A}_n f \to \widehat{f}$ in L^1, i.e. that $\|\mathcal{A}_n f - \widehat{f}\|_1 \to 0$, we proceed as follows:

$$\int |\widehat{f} - \mathcal{A}_n f|\, d\mu \leq \int_{M_\lambda^c} |\widehat{f} - \mathcal{A}_n f|\, d\mu + \int_{M_\lambda} |\widehat{f}| + |\mathcal{A}_n f|\, d\mu\,,$$

where $M_\lambda = \{\omega : \sup_{n\geq 0} |\mathcal{A}_n f(\omega)| > \lambda\}$, and $\lambda > 0$ is a parameter which will be choosen large later on. Since $|\mathcal{A}_n f| \leq \lambda$ on M_λ^c, Dominated Convergence gives

$$\lim_{n\to\infty} \int_{M_\lambda^c} |\widehat{f} - \mathcal{A}_n f|\, d\mu = 0\,.$$

Next, observe that $M_\lambda \subset N_\lambda(|f|)$. And therefore, by the Maximal Ergodic Theorem,

$$\mu(M_\lambda) \leq \mu(N_\lambda(|f|)) \leq \frac{1}{\lambda}\int_{N_\lambda(|f|)} |f|\, d\mu \leq \frac{1}{\lambda}\int |f|\, d\mu\,,$$

which converges to zero when $\lambda \to \infty$. As a consequence, $\int_{M_\lambda} |\widehat{f}|\, d\mu$ converges to zero when $\lambda \to \infty$. For the last term,

$$\int_{M_\lambda} |\mathcal{A}_n f|\, d\mu \leq \frac{1}{n} \sum_{k=0}^{n-1} \int_{M_\lambda} |f| \circ T^k\, d\mu\,.$$

We decompose the integrals as follows:

$$\int_{M_\lambda} |f| \circ T^k\, d\mu \leq \int_{|f|\circ T^k > R} |f| \circ T^k\, d\mu + \int_{M_\lambda \cap \{|f|\circ T^k \leq R\}} |f| \circ T^k\, d\mu\,.$$

By the Change of Variable Formula (Exercise 6.1),

$$\int_{|f|\circ T^k > R} |f| \circ T^k\, d\mu = \int_{|f| > R} |f|\, d\mu\,,$$

which is small when R is large. On the other hand,

$$\int_{M_\lambda \cap \{|f|\circ T^k \leq R\}} |f| \circ T^k\, d\mu \leq R\mu(M_\lambda)\,,$$

which is small for large λ, as seen before. Combined with the previous bounds, this shows that $\mathcal{A}_n f \to \widehat{f}$ in L^1. In particular, by the Change of Variable Formula,

$$\int \widehat{f}\, d\mu = \lim_{n\to\infty} \int \mathcal{A}_n f\, d\mu = \lim_{n\to\infty} \int f\, d\mu = \int f\, d\mu\,. \qquad (6.2.17)$$

To show (6.2.7), we take a strictly invariant $A \in \mathcal{F}$ and apply the first part of the theorem to the function $f1_A \in L^1(\mu)$. Since the strict invariance of A gives $\widehat{f}1_A = \widehat{f1_A}$, we have

$$\int_A \widehat{f}\, d\mu = \int \widehat{f1_A}\, d\mu \overset{(6.2.17)}{=} \int f1_A\, d\mu = \int_A f\, d\mu\,,$$

which finishes the proof of the Ergodic Theorem. □

6.3. Ergodicity

By the Ergodic Theorem, the averages $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k\omega)$ in (6.2.6) converge almost surely to a limit which is, in general random, i.e. which is has a non-trivial dependence on the initial condition ω: $\omega \mapsto E[f|\mathcal{I}](\omega)$ is a random variable. One can wonder what kind of extra condition can be imposed on μ in order to the averages converge to some deterministic limit, i.e. to a constant (almost surely) independent of ω. The simplest way in which a conditional expectation $E[f|\mathcal{I}]$ is guaranteed to be constant almost everywhere is when the measure satisfies a 0-1 Law on $\mathfrak{I}$.

LEMMA 6.3.1. *Assume μ is* **trivial on** $\mathfrak{I}$*, i.e. $\mu(A) \in \{0,1\}$ for all $A \in \mathfrak{I}$. Then, for all $f \in L^1(\mu)$,*

$$E[f|\mathfrak{I}] = E[f] \quad \mu - a.s. \tag{6.3.1}$$

PROOF. Consider first an indicator $f = 1_B$, $B \in \mathfrak{F}$. For any $A \in \mathfrak{I}$,

$$\int_A \mu(B) d\mu = \mu(B)\mu(A) = \mu(B \cap A) = \int_A \mu[B|\mathfrak{I}] d\mu\,.$$

This implies that $E[1_B]$ is a version of $E[1_B|\mathfrak{I}]$, i.e. that they are equal almost everywhere. Let f be positive, bounded, and let f_n be a sequence of simple functions converging pointwise to f. Then for all $A \in \mathfrak{I}$,

$$\int_A E[f] d\mu = \lim_{n\to\infty} \int_A E[f_n] d\mu = \lim_{n\to\infty} \int_A E[f_n|\mathfrak{I}] d\mu = \int_A E[f|\mathfrak{I}] d\mu\,.$$

The general case now follows easily. □

DEFINITION 6.3.1. *A T-invariant measure μ is called* **ergodic** [1] *if it is* **trivial on** $\mathfrak{I}$*, i.e. if $\mu(A) = 0$ or 1 for each $A \in \mathfrak{I}_\mu$.*

In the case where the invariant measure is ergodic, the Ergodic Theorem thus says that *temporal averages of observables are deterministic.* In mathematical terms:

THEOREM 6.3.1. *Let μ be invariant under T and ergodic. For any $f \in L^1(\mu)$,*

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to E[f] \quad \mu - a.s. \tag{6.3.2}$$

As a concrete example, consider the indicator function $f = 1_A$. If μ is ergodic,

$$\frac{1}{n}\sum_{k=0}^{n-1} 1_A \circ T^k \to \mu(A)\,, \quad \mu - a.s.\,,$$

which is a first precise information about the orbit of almost-all point ω: it spends a fraction of time in a measurable set A exactly equal to the probability of A, $\mu(A)$. It is important to notice that there

[1] Sometimes, an ergodic measure is also called undecomposable, since Ω cannot be decomposed into $\Omega = A \cup A^c$ with $0 < \mu(A) < 1$, $A \in \mathfrak{I}$.

can sometimes exist different ergodic measures for the same transformation T, and this previous description of trajectories is different in each case. In particular, the "almost sureness" of convergence for the averages depends on the measure under consideration.

We will explore later the deep consequences of the Ergodic Theorem, but before this we give a few ways in which ergodicity can be tested.

6.3.1. Testing Ergodicity.

THEOREM 6.3.2. *Let μ be invariant under T. The following are equivalent:*

(1) *μ is ergodic.*
(2) *If $A \in \mathfrak{F}$ is* almost surely invariant, *i.e. $\mu(A \triangle T^{-1}A) = 0$, then $\mu(A) \in \{0,1\}$.*
(3) *For all $A, B \in \mathfrak{F}$ with $\mu(A) > 0$, $\mu(B) > 0$, there exists an integer $n \geq 1$ such that $\mu(A \cap T^n B) > 0$.*

PROOF. (1) implies (2): Assume $\mu(A \triangle T^{-1}A) = 0$. This clearly implies that $\mu(A \triangle T^{-1}A \triangle \dots \triangle T^{-k}A) = 0$ for all $k \geq 1$. Define $A_* := \limsup_n T^{-n}A$. Then A_* is strictly invariant, $T^{-1}A_* = A_*$, i.e. $\mu(A_*) \in \{0,1\}$ since μ is ergodic. Then,

$$\mu(A_*) = \lim_{n\to\infty} \mu\Big(\bigcap_{m\geq n} T^{-m}A\Big) = \lim_{n\to\infty} \mu\Big(T^{-n}\bigcap_{m\geq 0} T^{-m}A\Big)$$
$$= \mu\Big(\bigcap_{m\geq 0} T^{-m}A\Big) = \lim_{k\to\infty} \mu\Big(\bigcap_{m=0}^{k} T^{-m}A\Big) = \mu(A)\,.$$

This implies that $\mu(A) \in \{0,1\}$.
(2) implies (3): assume (2) holds. Take A, B with $\mu(A) > 0$, $\mu(B) > 0$ and assume that (3) is false, i.e. that $\mu(A \cap T^{-n}B) = 0$ for all $n \geq 1$. Then of course $\mu(A \cap B') = 0$, where $B' := \bigcup_{n\geq 1} T^{-n}B$. Then $T^{-1}B' \subset B'$, which implies $\mu(B' \triangle T^{-1}B') = \mu(B' \backslash T^{-1}B') = 0$, and so $\mu(B') \in \{0,1\}$ by (2). But $B' \supset T^{-1}B$, and so $\mu(B') \geq \mu(T^{-1}B) = \mu(B) > 0$, i.e. $\mu(B') = 1$. Since $\mu(A) > 0$, this contradicts $\mu(A \cap B') = 0$.
(3) implies (1): assume (3) holds, but that μ is not ergodic, that is that there exists a strictly invariant $B \in \mathfrak{F}$ such that $0 < \mu(B) < 1$. Then $T^{-n}B \cap B^c = \emptyset$ and $\mu(T^{-n} \cap B^c) = 0$ for all $n \geq 1$, which contradicts (3). □

THEOREM 6.3.3. *Let μ be invariant under T. Then μ is ergodic if and only if*

$$\frac{1}{n}\sum_{k=0}^{n-1}\mu(A\cap T^{-k}B)\to\mu(A)\mu(B)\,,\qquad \text{for all } A,B\in\mathcal{F}\,. \tag{6.3.3}$$

Actually, μ is ergodic if and only if (6.3.3) *holds for all $A,B\in\mathcal{C}$, where $\mathcal{C}$ is any algebra which generates $\mathcal{F}$.*

PROOF. Assume μ is ergodic. Take any pair of events $A,B\in\mathcal{F}$. By the Ergodic Theorem,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}1_A\cdot 1_B\circ T^k=1_A\cdot\mu(B)\qquad \mu-\text{a.s.}$$

Integrating on both sides with respect to μ and using Dominated Convergence gives (6.3.3). Then, assume (6.3.3) holds for all $A,B\in\mathcal{F}$. If A is invariant, then

$$\mu(A)=\mu(A\cap A)=\frac{1}{n}\sum_{k=0}^{n-1}\mu(A\cap T^{-k}A)\to\mu(A)\mu(A)\,,$$

and therefore $\mu(A)\in\{0,1\}$: μ is ergodic. Now assume that (6.3.3) holds for all $A,B\in\mathcal{C}$, where $\mathcal{C}$ is an algebra generating $\mathcal{F}$. Take $E,F\in\mathcal{F}$. Let $\epsilon>0$. By the Approximation Lemma 3.2.2, there exists $A\in\mathcal{C}$ such that $\mu(E\triangle A)\le\epsilon$, and $B\in\mathcal{C}$ such that $\mu(F\triangle B)\le\epsilon$. A direct calculation gives, for all $n\ge 1$,

$$\mu\big((T^{-n}E\cap F)\triangle(T^{-n}A\cap B)\big)\le\mu(E\triangle A)+\mu(F\triangle B)\le 2\epsilon\,.$$

This implies $|\mu(T^{-n}E\cap F)-\mu(T^{-n}A\cap B)|\le 2\epsilon$. So

$$\Big|\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-n}E\cap F)-\mu(E)\mu(F)\Big|\le\Big|\frac{1}{n}\sum_{k=0}^{n-1}[\mu(T^{-n}E\cap F)-\mu(T^{-n}A\cap B)]\Big|$$
$$+\Big|\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-n}A\cap B)-\mu(A)\mu(B)\Big|+|\mu(A)\mu(B)-\mu(E)\mu(F)|\,.$$

Therefore,

$$\Big|\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-n}E\cap F)-\mu(E)\mu(F)\Big|\le 4\epsilon+\Big|\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-n}A\cap B)-\mu(A)\mu(B)\Big|\,,$$

which tends to zero when $n\to\infty$. This shows the theorem. $\square$

The criterium for ergodicity given in (6.3.3) is sometimes called mixing. We say that μ is strongly mixing if

$$\mu(A \cap T^{-n}B) \to \mu(A)\mu(B) \tag{6.3.4}$$

when $n \to \infty$ for all $A, B \in \mathcal{F}$. By (6.3.3), strong mixing implies ergodicity. This criterium will be used often in the practice.

6.4. Maps on the Interval; Borel's Theorem on Normal Numbers

6.5. Ergodic Sequences

As we saw at the beginning of the chapter, a stationary sequence $(X_n)_{n\geq 1}$ can always be considered as constructed on the product space $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}), \mu)$ of sequences $\omega = (\omega_1, \omega_2, \dots)$, by taking

$$X_k(\omega) := Y(T^k\omega)\,,$$

where $Y(\omega) := \omega_1$. Remember that $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ is generated by the algebra of cylinders $\mathcal{C}$, and that the measure μ is obtained by the marginals of the sequence $(X_n)_{n\geq 1}$ and Kolmogorov's Extension Theorem. Then, $(X_n)_{n\geq 1}$ is stationary if and only if μ is invariant with respect to the shift T: $\mu \circ T^{-1} = \mu$.

We say that the sequence $(X_n)_{n\geq 0}$ is ergodic (resp. strongly mixing) if μ is ergodic (resp. strongly mixing) with respect to the shift T. The same definition holds when the variables take values in a finite or countable set S, like Markov chains. We discuss two important examples of such sequences.

6.5.1. IID Sequences and the Strong Law of Large Numbers. Assume $(X_n)_{n\geq 1}$ is i.i.d., integrable and with distribution ν. This means that μ is the product measure $\nu^{\mathbb{N}}$. Set $Y := X_1$. By the Ergodic Theorem,

$$\frac{X_1 + \cdots + X_n}{n} \to E[Y|\mathcal{I}] \quad \mu\text{-a.s.} \tag{6.5.1}$$

Let us show that μ is strongly mixing. Take two cylinders $A, B \in \mathcal{C}$. Then clearly, for large enough n, the cylinders A and $T^{-n}B$ have disjoint bases, and therefore become independent: $\mu(A \cap T^{-n}B) = \mu(A)\mu(B)$ for large enough n. By Theorem 6.3.3, this implies that μ

is ergodic, and so (6.5.1) reads

$$\frac{X_1 + \cdots + X_n}{n} \to E[X_1] \quad \mu\text{-a.s.}\,, \tag{6.5.2}$$

which is the Strong Law of Large Numbers.

6.5.2. Stationary Markov Chains. Let $(X_n)_{n\geq 0}$ be a Markov chain taking values in a countable set S, with transition matrix Q. Let π be a distribution on S which is invariant with respect to Q, i.e. $\pi Q = \pi$. If X_0 has distribution π, then we saw in Example 6.1.2 that the chain $(X_n)_{n\geq 0}$ is stationary. What extra condition should it satisfy in order to be ergodic? strongly mixing?

Consider the canonical construction of the chain on the product space $S^{\{0,1,\dots\}}$, with the σ-algebra generated by thin cylinders (see Section 4.1.1). Assume the initial distribution is π, and that the measure associated to the chain is denoted μ. For notational convenience, we don't indicate the dependence of μ on π.

The study of ergodicity for Markov chain is greatly simplified by the introduction of a matrix Π on $S \times S$, whose properties are similar to those of a transition matrix. Remember by Theorem 6.3.3 that the chain is ergodic if and only if

$$\frac{1}{n}\sum_{k=0}^{n-1} \mu(A \cap T^{-k}B) \to \mu(A)\mu(B) \tag{6.5.3}$$

for all cylinders $A, B \in \mathcal{C}$. Since S is countable, we can even consider the previous convergence only for thin cylinders. A particular case is the one in which these thin cylinders are $A = \{\omega : \omega_0 = x\}$, $B = \{\omega : \omega_0 = y\}$. Then

$$\mu(A \cap T^{-k}B) = \mu(X_k = y, X_0 = x) = Q^{(k)}(x,y)\pi(x)\,, \tag{6.5.4}$$

and (6.5.3) becomes

$$\pi(x) \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} Q^{(k)}(x,y) = \pi(x)\pi(y)\,. \tag{6.5.5}$$

Therefore, if we assume that $\pi(x) > 0$ for all $x \in S$, the following matrix appears naturally:

$$\Pi := \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} Q^{(k)} .$$

Observe that the existence of Π is guaranteed by the Ergodic Theorem [2]. A few properties of Π are given hereafter.

LEMMA 6.5.1. Π *is a transition matrix on* S*, and satisfies*

$$Q\Pi = \Pi Q = Q\,, \quad \Pi^{(2)} = \Pi\,.$$

Moreover, if π *denotes an invariant distribution with respect to* Q*, then* π *is also an invariant distribution with respect to* Π*:* $\pi\Pi = \pi$*.*

PROOF. The proof is a straightforward repeated use of the Dominated Convergence Theorem. Clearly, $0 \le \Pi(x,y) \le 1$, and

$$\sum_{y\in S} \Pi(x,y) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1}\sum_{y\in S} Q^{(k)}(x,y) = 1\,,$$

which shows that Π is a transition matrix.

$$\begin{aligned} Q\Pi(x,y) = \sum_{z\in S} Q(x,z)\Pi(z,y) &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1}\sum_{y\in S} Q(x,z)Q^{(k)}(z,y) \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} Q^{(k+1)}(x,y) = \Pi(x,y)\,. \end{aligned}$$

In the same way, $\Pi Q = \Pi$, and

$$\begin{aligned} \Pi^{(2)}(x,y) = \sum_{z\in \S} \Pi(x,z)\Pi(z,y) &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1}\sum_{z\in S} Q^{(k)}(x,z)\Pi(z,y) \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} Q^{(k)}\Pi(x,y) \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \Pi(x,y) = \Pi(x,y)\,. \end{aligned}$$

[2]Namely, the Ergodic Theorem implies that $\frac{1}{n}\sum_{k=0}^{n-1} 1_A \cdot 1_B \circ T^k$ has a limit when $n \to \infty$ exists μ-almost surely and in $L^1(\mu)$. By integrating with respect to μ and using (6.5.4), this expression equals $\pi(x)\Pi(x,y)$.

If π is invariant with respect to Q, i.e. $\pi Q = \pi$, then $\pi Q^{(k)} = \pi$ for all $k \geq 1$, and by the same type of argument one gets $\pi \Pi = \pi$. $\square$

We have already proved part of

LEMMA 6.5.2. *Assume π is a invariant with respect to Q, and that $\pi(x) > 0$ for all $x \in S$. Then*

(1) *μ is strongly mixing if and only if*

$$Q^{(n)}(x, y) \to \pi(y) \quad \forall x, y \in S\,.$$

(2) *μ is ergodic if and only if $\Pi(x, y) = \pi(y)$, i.e.*

$$\frac{1}{n}\sum_{k=0}^{n-1} Q^{(n)}(x, y) \to \pi(y) \quad \forall x, y \in S\,.$$

PROOF. Consider two thin cylinders $C = [x_0, \ldots, x_p]$, $D = [y_0, \ldots, y_q]$. Then for large k, $\mu(C \cap T^{-k}D)$ equals

$$\pi(x_0)Q(x_0, x_1)\cdots Q(x_{p-1}, x_p)Q^{(k-p+1)}(x_p, y_0)Q(y_0, y_1)\cdots Q(y_{q-1}, y_p).$$

This proves (1). Averaging over $k = 0, 1, \ldots, n-1$ and since

$$\frac{1}{n}\sum_{k=0}^{n-1} Q^{k-p+1}(x_p, y_0) \to \Pi(x_p, y_0) = \pi(y_0)\,,$$

we have that $\frac{1}{n}\sum_{k=0}^{n-1}\mu(C \cap T^{-k}D)$ converges to

$$\pi(x_0)Q(x_0, x_1)\cdots Q(x_{p-1}, x_p)\pi(y_0)Q(y_0, y_1)\cdots Q(y_{q-1}, y_p) \equiv \mu(C)\mu(D)\,,$$

which shows that μ is ergodic. $\square$

The following shows that the chain is ergodic if and only if its transition matrix is irreducible.

THEOREM 6.5.1. *Consider the canonical representation of a stationary Markov chain with transition matrix Q and invariant distribution $\pi > 0$. Consider the following statements.*

(1) *μ is ergodic with respect to the shift T.*
(2) *$\Pi(x, y)$ does not depend on x.*
(3) *Q is irreducible*
(4) *$\Pi(x, y) > 0$ for all $x, y \in S$.*

Then (1) $\Leftrightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4). *If $|S| < \infty$, then* (4) $\Rightarrow$ (2).

PROOF. (1)⇔(2): By Lemma 6.5.1, $\sum_y \pi(y)\Pi(y,x) = \pi(x)$. Therefore, $\Pi(y,x)$ does not depend on y if and only if it equals $\pi(x)$. By the Lemma 6.5.2 this is equivalent to μ being ergodic.
(1)⇒(3): Assume that μ is ergodic but that Q is not irreducible. Then there exists some nonempty proper subset $S' \subset S$ such that $Q(x,y) = 0$ for all $x \in S'$, $y \in S \backslash S'$. Define $A := \{X_0 \in S'\}$. We have $\mu(A) = \pi(S') \in (0,1)$ (remember that $\pi > 0$). On the other side, $T^{-1}A = \{X_1 \in S'\}$, and so $A \triangle T^{-1}A = \{X_0 \in S', X_1 \notin S'\} \cup \{X_0 \notin S', X_1 \in S'\}$, so clearly $\mu(A \triangle T^{-1}A) = 0$. By Theorem 6.3.2, μ can therefore not be ergodic.
(3)⇒(4): Fix some $x \in S$, and let $S_x := \{y : \Pi(x,y) > 0\}$. We claim that S_x is closed. Namely, take $y \in S_x$, $z \in S \backslash S_x$. Assume $Q(y,z) > 0$. Then

$$\Pi(x,z) = \Pi Q(x,y) = \sum_{y' \in S} \Pi(x,y')Q(y',z) \geq \Pi(x,y)Q(y,z) > 0\,,$$

i.e. $z \in S_x$, a contradiction. Since the chain is irreducible, this implies $S_x \equiv S$, and shows that $\Pi(x,y) > 0$ for all y.
Finally, assume $|S| < \infty$ and that (4) holds. Observe that since $\Pi^{(2)} = \Pi$, any column of Π is a solution of the linear system $\Pi f = f$. Namely, by fixing y and defining $f(x) := \Pi(x,y)$, we have

$$\Pi f(x) = \sum_z \Pi(x,z) f(z) = \sum_z \Pi(x,z)\Pi(z,y) = \Pi(x,y) = f(x)\,.$$

Let $m := \max_{x \in S} f(x)$. Assume there exists x_* such that $f(x_*) < m$. Then for all other x we would have $f(x) = \Pi f(x) < m$, which is impossible since we are assuming that S is finite. Therefore, $f(x) = m$ for all x, which is equivalent to saying that $\Pi(x,y)$ does not depend on x, and therefore (4)⇒(2). □

Finally, let us see how the Ergodic Theorem for Markov chains can be obtained as a corollary of the Ergodic Theorem of Birkhoff, at least in the case where S is finite. Let $f : S \to \mathbb{R}$ be integrable with respect to the invariant distribution π: $\int |f| d\pi < \infty$. This implies that $\widetilde{f}$ defined on $S^{\{0,1,2,\dots\}}$ by $\widetilde{f}(\omega) := f(\omega_0)$ is integrable with respect to μ. By the Ergodic Theorem of Birkhoff,

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \equiv \frac{1}{n}\sum_{k=0}^{n-1} \widetilde{f} \circ T^k \to E[\widetilde{f}|\mathfrak{I}]$$

μ-almost surely. In particular, if we assume that the chain is irreducible, then by the previous theorem, μ is ergodic with respect to the shift, and therefore $E[\widetilde{f}|\mathcal{I}] = E[\widetilde{f}] = \int f d\pi$. This is Theorem 4.6.1 of Section 4.6.

6.6. Convex Structure of Invariant Measures

In this section we relate ergodicity to a certain undecomposability with respect to convex combination of probability measures. Consider a measurable space $(\Omega, \mathcal{F})$. Denote the set of measures on $(\Omega, \mathcal{F})$ by $\mathcal{M}^+$, and the set of probability measures by $\mathcal{M}_1^+ \subset \mathcal{M}^+$. $\mathcal{M}_1^+$ is a convex subset of $\mathcal{M}^+$. Now if $T : \Omega \to \Omega$ is measurable, we denote by $\mathcal{M}_{1,T}^+ \subset \mathcal{M}_1^+$ the set of measures which are invariant under T. Again, $\mathcal{M}_{1,T}^+$ is convex, and we will see that its extreme elements are exactly the invariant probability measures which are ergodic with respect to T. Before that, let us show that distinct ergodic measures which are invariant with respect to the same transformation have disjoint supports.

Lemma 6.6.1. *Let $\mu, \nu \in \mathcal{M}_{1,T}^+$ be both ergodic. Then either they coincide, or they are singular (in that there exists $A \in \mathcal{F}$ with $\mu(A) = 1$, $\nu(A) = 0$).*

Proof. For any $A \in \mathcal{F}$, consider the set $A_\mu \subset \Omega$ on which $\frac{1}{n}\sum_{k=0}^{n-1} 1_A$ converges μ-almost surely to $\mu(A)$. Since μ is ergodic, we have $\mu(A_\mu) = 1$. In the same way, we have the set A_ν for which $\nu(A_\nu) = 1$. Now if μ and ν are distinct, there exists some A such that $\mu(A) \neq \nu(A)$. But then the sets A_μ and A_ν are disjoint, which implies that μ and ν are singular. □

Theorem 6.6.1. *A probability measure $\mu \in \mathcal{M}_{1,T}^+$ is extremal if and only if it is ergodic.*

Proof. If μ is not ergodic, then there exists some invariant set A such that $0 < \mu(A) < 1$. Consider the measures

$$\nu_1 := \mu(\cdot|A), \quad \nu_2 := \mu(\cdot|A^c).$$

These measures are invariant. Namely, by the invariance of A and μ,

$$\nu_1(T^{-1}B) = \frac{\mu(T^{-1}B \cap A)}{\mu(A)} = \frac{\mu(T^{-1}(B \cap A))}{\mu(A)} = \frac{\mu(B \cap A)}{\mu(A)} = \nu_1(B).$$

Since $\mu = \mu(\cdot \cap A) + \mu(\cdot \cap A^c) = \lambda\nu_1 + (1-\lambda)\nu_2$ where $\lambda := \mu(A)$, μ is not extreme. On the other hand, assume μ is ergodic and that there exist two invariant measures ν_1, ν_2 and some $0 < \lambda < 1$ such that $\mu = \lambda\nu_1 + (1-\lambda)\nu_2$. Then, ν_1 is absolutely continuous with respect to μ. In particular, ν_1 is ergodic. By Lemma 6.6.1, μ and ν_1 either coincide, or are singuler. But they can't be singular since there would exist some event A such that $\mu(A) = 1$, $\nu_1(A) = 0$, which is impossible since

$$1 = \mu(A) = \lambda\nu_1(A) + (1-\lambda)\nu_2(A) = (1-\lambda)\nu_2(A) \leq 1-\lambda < 1\,.$$

Therefore $\mu = \nu_1$. In the same way, $\mu = \nu_2$. Therefore, μ is extremal. □

Since extreme elements of $\mathcal{M}^+_{1,T}$, a natural question is to know if there exists a way in which any invariant probability measure can be decomposed into a convex combination of ergodic measures, and if this decomposition is unique. It happens that such decompositions are guaranteed when the measurable space $(\Omega, \mathcal{F})$ satisfies some countability assumptions. We refer to [**Geo88**], Section 7, where a general statement can be found.

6.7. Exercises

VOIR DAJANI

EXERCISE 6.1. Prove the Change of Variable Formula: if μ is invariant under T, then for all $f \in L^1(\Omega, \mathcal{F}, \mu)$,

$$\int f \circ T\, d\mu = \int f\, d\mu\,. \tag{6.7.1}$$

EXERCISE 6.2. Show that the set function μ defined by

$$\mu(C_\Lambda(a_1, \dots, a_{|\Lambda|})) := \prod_{i\in\Lambda} p(a_i) \tag{6.7.2}$$

is σ-additive on the algebra of cylinders $\mathcal{C}$.

EXERCISE 6.3. Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, where $\mathcal{F}$ is generated by an algebra $\mathcal{C}$. If $T^{-1}A \in \mathcal{F}$ and $\mu(T^{-1}A) = \mu(A)$ for all $A \in \mathcal{C}$, then μ is invariant under T. (Hint: use the Monotone Class Theorem.)

EXERCISE 6.4. Define ergodicity for measures on $\mathbb{Z}^d$. Show that product measures are ergodic.

EXERCISE 6.5. Call μ weakly mixing if

$$\frac{1}{n}\sum_{k=0}^{n-1}|\mu(A\cap T^{-k}B)-\mu(A)\mu(B)|\to 0\,,\qquad \text{for all } A,B\in\mathcal{F}\,.$$

Show that μ is weakly mixing if and only if

$$\frac{1}{n}\sum_{k=0}^{n-1}(\mu(A\cap T^{-k}B)-\mu(A)\mu(B))^2\to 0\,,\qquad \text{for all } A,B\in\mathcal{F}\,.$$

EXERCISE 6.6. [**KH74**] Uniformly distributed sequences in $[0,1]$

(1) show that any uniformly distributed sequence is dense
(2) show that if r is rational, then the sequence $x_n := (rn)$, $n\geq 1$ is not uniformly distributed. Is there a proper interval of $[0,1]$ which is visited with the appropriated frequency by the sequence?
(3) show that the sequence $0/2, 1/2, 2/2, 0/3, 1/3, 2/3, 3/3$, etc. is uniformly distributed.

EXERCISE 6.7. [**Pet00**] p. 33. Identify $\widehat{f}$ in the following cases.

(1) The two sided shift $\Omega = \mathsf{A}^{\mathbb{Z}}$ with invariant measure $\mu = p^{\mathbb{Z}}$, and $f(\omega) = 1_{a_0}(\omega_0)$ for some $a_0 \in \mathsf{A}$.
(2) $\Omega = [0,1)$, $T\omega = \omega + \alpha \mod 1$, μ is the Lebesgue measure, and $f = 1_I$ for some interval $I \subset [0,1)$.
(3) $\Omega = \mathbb{R}$, $T\omega = \omega + 1$, μ is the Lebesgue measure, and $f \in L^1$.

EXERCISE 6.8. Show that if μ is trivial on $\mathcal{I}_\mu$, then for all $f \in L^1(\mu)$, $E_\mu(f|\mathcal{I}_\mu) = E_\mu(f)$ μ-almost surely. Hint: start with simple functions.

EXERCISE 6.9. Grimmett p. 410, ex. 14. Consider $([0,1),\mathcal{B}([0,1))\lambda)$. Show that the shift $T:[0,1)\to[0,1)$ is measurable, preserves λ, and ergodic. VOIR BILL. P. 11. VOIR LA REMARQUE 1.2.15 de Dajani. VOIR AUSSI SON EXERCICE 1.2.22 SUR LA TRANSFORMATION DU BOULANGER. Let $X(\omega) = \omega$. Show that the proportion of $1s$ in in the expansion of X is in base to equals $\frac{1}{2}$.

EXERCISE 6.10. dAJANI P. 29. Show that Gauss' transformation doesn't preserve the lebesgue measure, but preserves

$$\mu(A) := \frac{1}{\log 2}\int_A \frac{1}{1+x}dx\,.$$

VOIR BILLINGSLEY p.43 ou RAUZY

EXERCISE 6.11. Verify that $\mathcal{I}_\mu$ is a σ-algebra. Show that $g : \Omega \to \mathbb{R}$ is $\mathcal{I}_\mu$-measurable if and only if it is μ-almost invariant.

EXERCISE 6.12. Varadhan p. 184. Prove that any any almost invariant set differs by a set of measure zero from an invariant set.

EXERCISE 6.13. Varadhan p. 184. Prove that any product measure is ergodic for the shift

EXERCISE 6.14. Varadhan p. 187. Show that any two distinct ergodic invariant measures are orthogonal on $\mathcal{I}$.

CHAPTER 7

Brownian Motion

This chapter is devoted to the construction of the Brownian motion and to the study of its basic properties. Most of this chapter, in particular Section 7.2 on weak convergence, is taken from the book of Billingsley [**Bil68**].

7.1. Introduction

Consider the simple symmetric random walk on $\mathbb{Z}$ starting at the origin, denoted $(S_n)_{n\geq 0}$. We denote its increments by Y_k, and the underlying probability space by $(\Omega, \mathcal{F}, P)$.

The properties of the position of the walk at time n are described by the Law of Large Numbers (LLN) and by the Central Limit Theorem (CLT). In the CLT for instance, the description of the distribution of $\frac{S_n}{\sqrt{n}}$ satisfies

$$P\Big(\frac{S_n}{\sqrt{n}} \in A\Big) \to \frac{1}{\sqrt{2\pi}} \int_A e^{-\frac{x^2}{2}} dx \tag{7.1.1}$$

for all Borel set $A \in \mathcal{B}(\mathbb{R})$. The CLT is a statement about the *weak convergence in distribution of* $\frac{S_n}{\sqrt{n}}$ *to the centered normal Gaussian.* A sequence of probability measures μ_n on the real line $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is said to converge weakly to a probability measure μ if

$$\mu_n(A) \to \mu(A)\,, \quad \text{for all } A \in \mathcal{B}(\mathbb{R}) \text{ with } \mu(\partial A) = 0\,. \tag{7.1.2}$$

In our case, the distribution of $\frac{S_n}{\sqrt{n}}$ is the probability measure on the real line defined by $\mu_n(A) := P(\frac{S_n}{\sqrt{n}} \in A)$ and μ is the centered normal Gaussian distribution $\mu(dx) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})dx$. The fact that in (7.1.1) the convergence holds for *all* Borel set A stems from the fact that μ is absolutely continuous with respect to the Lebesgue measure, which is non-atomic.

The CLT describes the behaviour of the position of the random walk *at time n*. A natural question is to ask about the statistical properties of the *trajectory as a whole*, that is considering the values of S_k at *all* times $k \in [0,n]$, for large n. To keep track of the whole trajectory, we consider a double rescaling, in space and time, and consider the trajectory up to time n as a function of a real variable t on the interval $[0,1]$. The rescaling in time is thus a division by n:

$$\{0,1,\dots,,n-1,n\} \to \{0,\frac{1}{n},\dots,\dots,1-\frac{1}{n},1\},$$

and since S_n has typical fluctuations of order $\sqrt{n}$ by the CLT, a spatial rescaling of order $\sqrt{n}$ is necessary in order to obtain a limiting object (if any) which is bounded. More precisely, for each time n we consider the rescaled trajectory $X_n : [0,1] \to \mathbb{R}$ defined by interpolating, on each time interval $t \in [\frac{k-1}{n}, \frac{k}{n}]$, between the values of $\frac{S_{k-1}}{\sqrt{n}}$ and $\frac{S_k}{\sqrt{n}}$ (see Figure 1):

$$X_n(t) := \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} + (nt - \lfloor nt \rfloor)\frac{1}{\sqrt{n}} Y_{\lfloor nt \rfloor + 1}\,. \qquad (7.1.3)$$

X_n belongs to the set of continuous functions on $[0,1]$, denoted by C.

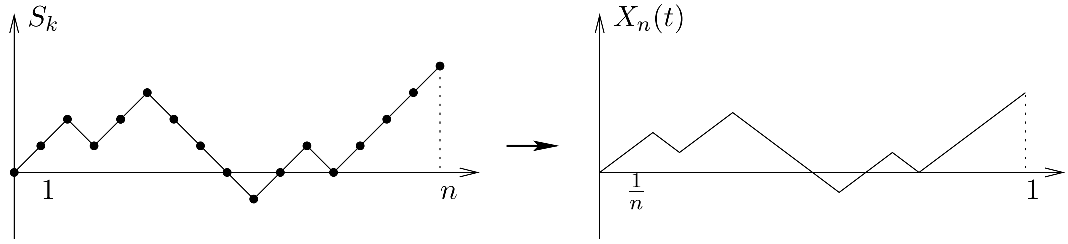


FIGURE 1. The rescaling of the simple random walk.

Since each X_n is a random function, it has two inputs: the time variable $t \in [0,1]$ and the random realization $\omega \in \Omega$. We will sometimes write $X_n : [0,1] \times \Omega \to \mathbb{R}$ to indicate this double dependence.

To understand the statistical properties of the random functions X_n when n becomes large we need to define their distribution, which requires the definition of a σ-algebra on C. The natural distance on C is the sup-norm: for $x, y \in \mathsf{C}$,

$$\rho(x,y) := \sup_{t \in [0,1]} |x(t) - y(t)|\,.$$

We consider on C the Borel σ-algebra generated by the open sets defined by ρ. The distribution of X_n is defined as follows: for any Borel set $B \subset \mathsf{C}$,

$$\mu_n(B) := P(X_n \in B)\,. \qquad (7.1.4)$$

We say that a sequence of probability measures μ_n converges weakly to a probability measure μ if

$$\mu_n(B) \to \mu(B)\,, \quad \text{for all Borel set } B \text{ with } \mu(\partial B) = 0\,. \qquad (7.1.5)$$

The rest of the chapter is essentially devoted to show that the sequence μ_n defined in (7.1.4) converges weakly to a measure $\mathcal{W}$ called Wiener measure. $\mathcal{W}$ is a measure on the set of continuous functions over $[0, 1]$. Proving its existence will require a full exposition of weak convergence of probability measures in metric spaces.

The rest of the chapter is organized as follows. In Section 7.2 we introduce weak convergence for sequences of probability measures on a general metric space (S, ρ). After describing some of its basic properties, we give a few criteria, among which the Portmanteau Theorem, which allow to test weak convergence in concrete cases, for example when $S = \mathbb{R}^d$, $\mathbb{R}^{\mathbb{N}}$, or C. We then prove the Theorem of Prohorov, in Section 7.3, which gives a necessary and sufficient condition under which a sequence μ_n is guaranteed to converge weakly to *some* probability measure: *tightness*. These results are then used, together with the Arzelà-Ascoli Theorem, to construct the Wiener measure in Section 7.4. We prove Donsker's Invariance Principle in Section 7.5. Some properties of the Brownian motion, such as almost-everywhere non-differentiability and the Law of the Iterated Logarithm, are then described in Section 7.6.

The whole of Section 7.2 is taken from the book of Billingsley [**Bil68**].

7.2. Weak Convergence in Metric Spaces

Let (S, ρ) be a metric space. An open sphere of radius $r > 0$ centered at $x \in S$ is denoted $B_r(x) = \{y \in S : \rho(y, x) < r\}$. The interior of $A \subset S$ is denoted $\text{int}A$, its closure by $\bar{A}$. Let $A \subset S$, and define $\rho(x, A) := \inf\{\rho(x, y) : y \in A\}$. For $\epsilon > 0$, denote the open ϵ-thickening of A by $[A]_\epsilon := \{y : \rho(y, A) < \epsilon\}$, and the closed ϵ-thickening of A, by $[A]^\epsilon := \{y : \rho(y, A) \leq \epsilon\}$. Observe that if A is closed, then $[A]^\epsilon \searrow A$ when $\epsilon \searrow 0^+$. Let $C(S)$ denote the family of continuous bounded functions $f : S \to \mathbb{R}$.
LEMBRETE DE TOPOLOGIA?

7.2.1. Borel Sets and Probability Measures on S. The open sets of S generate the σ-algebra of Borel sets, which we denote by $\mathcal{S}$. Observe that $\mathcal{S}$ contains all open, closed and compact sets. $\mathcal{S}$ can therefore be generated equivalently by either of these classes. If S, S' are two metric spaces and if $f : S \to S'$ is continuous, then it is also measurable: $f^{-1}(B') \in \mathcal{S}$ for all $B' \in \mathcal{S}'$ [1]. This holds in particular for the continuous functions $f : S \to \mathbb{R}$, which we denote $C(S)$. Usually, we will denote the elements of $\mathcal{S}$ by $A, B, C, \dots$ (unless otherwise specified, these letters will be used to denote Borel sets), and probability measures on $(S, \mathcal{S})$ will be denoted by μ or ν.

We first show how Borel sets can be approximated, in a measure-theoretic sense, by closed or open sets. Call a probability measure μ on $(S, \mathcal{S})$ regular if, for all A and all $\epsilon > 0$ there exists a closed set F and an open set G such that $F \subset A \subset G$ and $\mu(G \backslash F) \leq \epsilon$.

LEMMA 7.2.1. *Any probability measure on $(S, \mathcal{S})$ is regular.*

PROOF. Take some probability measure μ. Let $\mathcal{D}$ denote the class of all sets A for which there exists, for all $\epsilon > 0$, a closed set F and an open set G such that $F \subset A \subset G$ and $\mu(G \backslash F) \leq \epsilon$. If A is closed, take $F := A$, and $G_\delta := [A]_\delta$. G_δ is open. Since A is closed, $G_\delta \searrow A$ when $\delta \to 0^+$, and so $\mu(G_\delta) \searrow \mu(A)$. Therefore, $\mathcal{D}$ contains all closed sets. If we show that $\mathcal{D}$ is a σ-algebra, then we are done. Clearly, $\mathcal{D}$ is stable under complementation, and contains $\emptyset, S$. Let then $A_n \in \mathcal{D}$. It suffices to show that $A := \bigcup_n A_n \in \mathcal{D}$. Let $\epsilon > 0$. For each n, take a closed set F_n and an open set G_n such that $F_n \subset A_n \subset G_n$, and such that $\mu(G_n \backslash F_n) \leq \epsilon 2^{-(n+1)}$. Set $F_N := \bigcup_{n=1}^N F_n$, $G := \bigcup_{n \geq 1} G_n$. Take N large enough so that $\mu(F_\infty \backslash F_N) \leq \frac{\epsilon}{2}$, and let $F := F_N$. We have $\mu(G \backslash F) \leq \mu(G \backslash F_\infty) + \mu(F_\infty \backslash F)$. But

$$\mu(G \backslash F_\infty) \leq \sum_{n \geq 1} \mu(G_n \backslash F_\infty) \leq \sum_{n \geq 1} \mu(G_n \backslash F_n) \leq \frac{\epsilon}{2},$$

and by definition $\mu(F_\infty \backslash F) \leq \frac{\epsilon}{2}$. This shows that $A \in \mathcal{D}$. □

[1] Let S, S' be two metric spaces and denote their Borel σ-algebras by $\mathcal{S}$, $\mathcal{S}'$ respectively. Let $f : S \to S'$ be continuous. Take an open set $A' \subset S'$. Since $f^{-1}(A')$ is an open set of S, it belongs to $\mathcal{S}$. Denote by $\widetilde{\mathcal{S}}$ the set of subsets $A \subset S'$ for which $f^{-1}(A) \in \mathcal{S}$. It is easy to verify that $\widetilde{\mathcal{S}}$ is a σ-algebra, and we have just seen that it contains all the open sets of S'. Since these generate $\mathcal{S}'$, $\widetilde{\mathcal{S}}$ contains $\mathcal{S}'$. f is therefore measurable.

DEFINITION 7.2.1. *A class of sets $\mathcal{D} \subset \mathcal{S}$ is a* **determining class** *if any two probability measures which coincide on $\mathcal{D}$ are equal. That is, if $\mu(D) = \nu(D)$ for all $D \in \mathcal{D}$ then $\mu = \nu$*

Lemma 7.2.1 has an immediate corollary:

COROLLARY 7.2.1. *The closed sets form a determining class.*

PROOF. Namely, assume μ, ν are such that $\mu(F) = \nu(F)$ for all closed F. Then by the previous lemma, for any $A \in \mathcal{S}$, there exists a sequence of closed sets $F_n \subset A$ such that $\mu(A) = \lim_n \mu(F_n)$. But $\lim_n \mu(F_n) = \lim_n \nu(F_n) \leq \nu(A)$. In the same way, we show that $\nu(A) \leq \mu(A)$. Therefore, $\mu = \nu$. □

The next result shows that indicators can be approximated by uniformly continuous functions. Then, as can be easily verified [2],

$$|\rho(x, A) - \rho(y, A)| \leq \rho(x, y)\,, \quad \forall x, y \in S\,, \tag{7.2.1}$$

which implies that $\rho(\cdot, A)$ is uniformly continuous.

THEOREM 7.2.1. *Let $A \in \mathcal{S}$, $\epsilon > 0$. Then there exists $f \in C(S)$, uniformly continuous, such that $0 \leq f \leq 1$, $f(x) = 0$ if $\rho(x, A) \geq \epsilon$, and $f(x) = 1$ if $x \in A$.*

PROOF. Let $\phi(t) = 1$ if $t \leq 0$, $\phi(t) = 1 - t$ if $t \in [0, 1]$, and $\phi(t) = 0$ if $t \geq 1$. ϕ is uniformly continuous. For each $\epsilon > 0$, define $f(x) := \phi(\epsilon^{-1}\rho(x, A))$. It is easy to verify that f satisfies the wanted properties. □

Since indicators of measurable sets can be approximated by continuous functions, we can now show that a measure is uniquely determined by the way in which it integrates continuous functions.

THEOREM 7.2.2. *Two probability measures μ, ν on $(S, \mathcal{S})$ are equal if and only if*

$$\int f d\mu = \int f d\nu\,, \quad \forall f \in C(S)\,. \tag{7.2.2}$$

PROOF. Assuming (7.2.2) holds, take $F \in \mathcal{S}$ closed. Consider the sequence

$$f_n(x) := \phi(n\rho(x, F)) \tag{7.2.3}$$

[2] Write, for all $x_0 \in F$, $\rho(x, A) \leq \rho(x, x_0) \leq \rho(x, y) + \rho(y, x_0)$. This implies $\rho(x, A) \leq \rho(x, y) + \rho(y, A)$. Interchanging the roles of x and y proves (7.2.1).

where ϕ is the function of Theorem 7.2.1. We have $f_n \geq 1_F$ and since F is closed, f_n converges pointwise to the indicator of F: $f_n(x) \searrow 1_F(x)$. By Dominated Convergence,

$$\mu(F) = \lim_{n\to\infty} \int f_n(x) d\mu = \lim_{n\to\infty} \int f_n(x) d\nu = \int 1_F d\nu = \nu(F) . \tag{7.2.4}$$

By Corollary 7.2.1, $\mu = \nu$. □

7.2.2. Weak Convergence: Definition and Testing Criteria. The following is the classical definition of weak convergence. The Portmanteau Theorem will show that it is actually equivalent to the definition we gave in (7.1.5).

DEFINITION 7.2.2. *Let $(\mu_n)_{n\geq 1}$ and μ be probability measures on $(S, \mathcal{S})$. We say that μ_n **converges weakly to** μ if*

$$\int f d\mu_n \to \int f d\mu , \quad \forall f \in C(S) . \tag{7.2.5}$$

When (7.2.5) *holds, we write $\mu_n \stackrel{\mathrm{w}}{\Rightarrow} \mu$.*

The main result of this section is Theorem 7.2.1 and its two corollaries, which say that weak convergence can be tested on classes of sets which are strictly smaller than $\mathcal{S}$. Such classes will be easy to handle in concrete situations. Call a set $A \in \mathcal{S}$ μ-continuous if $\mu(\partial A) = 0$. (Observe that $\partial A = \bar{A} \backslash \mathrm{int} A$, so $\partial A \in \mathcal{S}$.)

THEOREM 7.2.3 (The Portmanteau Theorem). *Let μ_n, μ be probability measures on $(S, \mathcal{S})$. The following conditions are equivalent.*

1. $\mu_n \stackrel{\mathrm{w}}{\Rightarrow} \mu$,
2. $\int f d\mu_n \to \int f d\mu$ *for all $f \in C(S)$ uniformly continuous,*
3. $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ *for all closed F,*
4. $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$ *for all open G,*
5. $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *for all μ-continuous set A.*

PROOF. (1) *implies* (2): trivial.
(2) *implies* (3): Let F be closed. Consider the uniformly continuous function f_k defined in (7.2.3). For all k,

$$\limsup_{n\to\infty} \mu_n(F) = \limsup_{n\to\infty} \int 1_F d\mu_n \leq \limsup_{n\to\infty} \int f_k d\mu_n = \int f_k d\mu$$

As we saw in (7.2.4), $\int f_k d\mu \to \mu(F)$ in the limit $k \to \infty$. This yields (3).

(3) *implies* (1): Take $f \in C(S)$. We will show that $\limsup_n \int f d\mu_n \leq \int f d\mu$; the result then follows by doing the same with $-f$. Since f is bounded, we can assume for example that $f(x) \in [0,1)$. For each $N \geq 1$, we divide the interval $[0,1)$ into N intervals of equal size and consider the approximation of f by the functions

$$f_N^-(x) := \sum_{i=1}^N \frac{i-1}{N} 1_{\{x:\frac{i-1}{N} \leq f(x) < \frac{i}{N}\}}(x)\,, \quad f_N^+(x) := \sum_{i=1}^N \frac{i}{N} 1_{\{x:\frac{i-1}{N} \leq f(x) < \frac{i}{N}\}}(x)\,.$$

Clearly, $f_N^- \leq f \leq f_N^+$ and so for any probability measure ν,

$$\int f_N^- d\nu \leq \int f d\nu \leq \int f_N^+ d\nu\,.$$

By noting that $\nu\big(\{x : \frac{i-1}{N} \leq f(x) < \frac{i}{N}\}\big) = \nu(F_{i-1}) - \nu(F_i)$, where F_i are the closed sets $F_i := \{x : f(x) \geq i/N\}$, $F_0 := S$, $F_N := \emptyset$, we easily obtain

$$\int f_N^+ d\nu = \frac{1}{N} + \frac{1}{N}\sum_{i=1}^N \nu(F_i)\,, \quad \int f_N^- d\nu = \frac{1}{N}\sum_{i=1}^N \nu(F_i)\,.$$

Using these expressions for the measures μ_n, μ,

$$\begin{aligned}\limsup_{n\to\infty} \int f d\mu_n &\leq \limsup_{n\to\infty} \Big\{\frac{1}{N} + \frac{1}{N}\sum_{i=1}^N \mu_n(F_i)\Big\} \\ &\leq \frac{1}{N} + \frac{1}{N}\sum_{i=1}^N \mu(F_i) = \frac{1}{N} + \int f_N^- d\mu \leq \frac{1}{N} + \int f d\mu\,,\end{aligned}$$

and the needed inequality then follows by taking $N \to \infty$.
(3) *is equivalent to* (4): Trivial since $\mu(F^c) = 1 - \mu(F)$.
(3) *and* (4) *imply* (5): Let $A \in \mathcal{S}$. Then, by (3) and (4),

$$\mu(\bar{A}) \geq \limsup_{n\to\infty} \mu_n(\bar{A}) \geq \limsup_{n\to\infty} \mu_n(\text{int}A) \geq \liminf_{n\to\infty} \mu_n(\text{int}A) \geq \mu(\text{int}A)\,.$$

But if $\mu(\partial A) = 0$, then $\mu(\bar{A}) = \mu(\text{int}A)$.
(5) *implies* (3): Let F be closed. Assume for a while we can find, for any small $\epsilon > 0$, some closed set $F' \supset F$ such that $\mu(\partial F') = 0$, and $P(F' \backslash F) \leq \epsilon$. Then clearly

$$\limsup_{n\to\infty} \mu_n(F) \leq \limsup_{n\to\infty} \mu_n(F') = \mu(F') \leq \mu(F) + \epsilon\,,$$

and the result follows by taking $\epsilon \to 0$. A natural candidate for F' is the closed ϵ-thickening of F, $[F]^\epsilon$, which is closed.
Claim: there exists a sequence $\epsilon_k \searrow 0$ *such that* $\mu(\partial[F]^{\epsilon_k}) = 0$ *for all*

k. Namely, assume that $\mu(\partial[F]^\epsilon) > 0$ for all $\epsilon \in (0,1)$. Lemma 7.2.2 hereafter implies that there exists some sequence $\epsilon_j' \in (0,1)$ of distinct values of ϵ such that $\sum_j \mu(\partial[F]^{\epsilon_j'}) = +\infty$. Since the sets $\{\partial[F]^\epsilon\}_{\epsilon\in(0,1)}$ are disjoint, μ wouldn't be a probability. Therefore, there exists at least one $\epsilon_1 \in (0,1)$ such that $\mu(\partial[F]^{\epsilon_1}) = 0$. Then, we start again: assume that $\mu(\partial[F]^\epsilon) > 0$ for all $\epsilon \in (0,\epsilon_1)$. Lemma 7.2.2 hereafter implies that there exists some sequence $\epsilon_j' \in (0,\epsilon_1)$ of distinct values of ϵ such that $\sum_j \mu(\partial[F]^{\epsilon_j'}) = +\infty$. Since the sets $\{\partial[F]^\epsilon\}_{\epsilon\in(0,\epsilon_1)}$ are disjoint, μ wouldn't be a probability. Therefore, there exists at least one $\epsilon_2 \in (0,\epsilon_1)$ such that $\mu(\partial[F]^{\epsilon_2}) = 0$, etc. This proves the claim. Now, for all k,

$$\limsup_{n\to\infty} \mu_n(F) \leq \limsup_{n\to\infty} \mu_n([F]^{\epsilon_k}) = \mu([F]^{\epsilon_k}) .$$

As we know, $\mu([F]^{\epsilon_k}) \searrow \mu(F)$ since F is closed. This finishes the proof of the Portmanteau Theorem □

Lemma 7.2.2. *Let $g : (0,a) \to (0,+\infty)$. Then there exists a sequence $t_k \in (0,a)$, $t_k \neq t_{k'}$, such that $\sum_k g(t_k) = +\infty$.*

Proof. Write $(0,\infty) = \bigcup_{j\geq 1} R_j$, where $R_1 := (1,\infty)$, and $R_j := (\frac{1}{j}, \frac{1}{j-1}]$ for $j \geq 2$. There exists at least one $j_0 \geq 1$ such that $g^{-1}(R_{j_0})$ contains an uncountable number of points (otherwise, $(0,a)$ would be countable). Let $\{t_1, t_2, \dots\} \subset g^{-1}(R_{j_0})$ be distinct. Since $g(t_k) \geq \frac{1}{j_0}$, we have $\sum_k g(t_k) = +\infty$. □

Condition (5) of the Portmanteau Theorem implies that weak convergence can be verified by testing if $\mu_n(A) \to \mu(A)$ for the μ-continuous Borel sets A. In concrete situations, one will want to test this convergence on classes of sets A whose structure is in general simpler than the whole σ-algebra $\mathcal{S}$. Later, we shall call these *convergence-determining classes*.

Proposition 7.2.1. *Let $\mathcal{U} \subset \mathcal{S}$ be such that*

1. *$\mathcal{U}$ is stable under finite intersections,*
2. *each open set $G \in \mathcal{S}$ can be written as a finite or countable union of elements of $\mathcal{U}$.*

If $\mu_n(A) \to \mu(A)$ for all $A \in \mathcal{U}$, then $\mu_n \overset{w}{\Rightarrow} \mu$.

PROOF. First observe that if $A_1, \dots, A_m \in \mathcal{U}$,

$$\mu_n(A_1 \cup \cdots \cup A_m) = \sum_{k=1}^{m} (-1)^{k+1} \sum_{I \subset \{1,\dots,k\}} \mu_n\Big(\bigcap_{j \in I} A_j\Big) .$$

Since $\bigcap_{j\in I} A_j \in \mathcal{U}$, this gives

$$\lim_{n\to\infty} \mu_n(A_1 \cup \cdots \cup A_m) = \mu(A_1 \cup \cdots \cup A_m) .$$

Let $G \in \mathcal{S}$ be open, $\epsilon > 0$. Since by assumption G can be written as a finite or countable union $G = \bigcup_{k\geq 1} A_k$, $A_k \in \mathcal{U}$, there exists some m such that

$$\mu(G) - \epsilon \leq \mu(A_1 \cup \cdots \cup A_m) = \lim_{n\to\infty} \mu_n(A_1 \cup \cdots \cup A_m) \leq \liminf_{n\to\infty} \mu_n(G) .$$

Since ϵ is arbitrary, the Portmanteau Theorem implies that $\mu_n \overset{w}{\Rightarrow} \mu$. □

We then give a criterium for testing weak convergence which has the advantage of involving a family $\mathcal{U}$ of sets which is close to being a base for the metric topology. It holds under a separability assumption.

THEOREM 7.2.4. *Let $\mathcal{U} \subset \mathcal{S}$ be such that*

(1) *$\mathcal{U}$ is stable under finite intersections,*
(2) *for all $x \in S$ and all $\epsilon > 0$, $\exists A \in \mathcal{U}$ such that $x \in \mathrm{int}A \subset A \subset B_\epsilon(x)$.*

If S is separable and if $\mu_n(A) \to \mu(A)$ for all $A \in \mathcal{U}$, then $\mu_n \overset{w}{\Rightarrow} \mu$.

First, we prove a topological lemma.

LEMMA 7.2.3. *Let S be separable. Then there exists a countable family $\mathcal{B}$ of open spheres such that any open set A can be expressed as a countable union of elements of $\mathcal{B}$ ($\mathcal{B}$ is called a* **base**).

PROOF. Let $T \subset S$ be dense and countable. Let $\mathcal{B} = \{B_1, B_2, \dots\}$ denote the set of open spheres centered at points of T, with radii $r \in \mathbb{Q}$. Let A be open. Let $A' := \bigcup_{k:B_k \subset A} B_k$. Clearly, $A' \subset A$. We then show that $A \subset A'$. Take any $x \in A$. Since A is open, there exists $r > 0$, such that $B_r(x) \subset A$. Let $t \in T \cap B_{\frac{r}{3}}(x)$. Take any rational $\frac{r}{3} < r_* < \frac{2r}{3}$. Then $x \in B_{r_*}(t) \subset B_r(x) \subset A$. Therefore $A \subset A'$. □

PROOF OF THEOREM 7.2.4. Let $G \subset S$ be open. The result will follow by Proposition 7.2.1 if one can show that G can be written as a countable union of elements of $\mathcal{U}$. Since G is open, there exists for all $x \in G$ some $r_x > 0$ such that $B_{r_x}(x) \subset G$, and by (2) there exists $A_x \in \mathcal{U}$ such that $x \in \text{int}A_x \subset A_x \subset B_{r_x}(x) \subset G$. Therefore, $G = \bigcup_{x \in G} A_x$. Consider the countable set of open spheres $\mathcal{B} = \{B_1, B_2, \dots\}$ of Lemma 7.2.3. For each $k \in \{1, 2, \dots\}$, let $A(k)$ be any of the A_xs which is such that $\text{int}A_x$ contains B_k (if any; otherwise, set $A(k) = \emptyset$). Then clearly $G = \bigcup_k A(k)$. □

DEFINITION 7.2.3. *A family of sets $\mathcal{U} \subset \mathcal{S}$ is a convergence-determining class if $\mu_n(A) \to \mu(A)$ for all μ-continuity set $A \in \mathcal{U}$ implies $\mu_n \overset{\text{w}}{\Rightarrow} \mu$.*

As well known, the class $\mathcal{U}$ of semi-infinite intervals $(-\infty, x]$ forms a convergence-determining class for weak convergence of probability measures on the real line. See Section 7.2.3 hereafter.

LEMMA 7.2.4. *Any convergence-determining class is determining. (The converse is false, see Section 7.2.5.)*

PROOF. Assume $\mu(A) = \nu(A)$ for all A belonging to a convergence determining class $\mathcal{U}$. Define $\mu_n := \mu$. Since $\mu_n(A) \to \nu(A)$ for all $A \in \mathcal{U}$ (in particular if A is μ-continuous), then $\mu_n \overset{\text{w}}{\Rightarrow} \nu$. This means that for all continuous f ,

$$\int f d\mu = \int f d\mu_n = \lim_{n\to\infty} \int f d\mu_n = \int f d\nu \,.$$

By Theorem 7.2.2, this implies that $\mu = \nu$. □

As a last criterium to test weak convergence, we have:

THEOREM 7.2.5. *$\mu_n \overset{\text{w}}{\Rightarrow} \mu$ if and only if each subsequence $\{\mu_{n_k}\} \subset \{\mu_n\}$ has a further subsequence $\{\mu_{n'_j}\} \subset \{\mu_{n_k}\}$ such that $\mu_{n'_j} \overset{\text{w}}{\Rightarrow} \mu$.*

PROOF. The result follows by the equivalent property for sequences of real numbers, which can be easily verified: a real sequence $\{x_n\}$ converges to x if and only if any subsequence $\{x_{n_k}\} \subset \{x_n\}$ has a further subsequence $\{x_{n'_j}\} \subset \{x_{n_k}\}$ which converges to x. □

In the following three sections, we consider particular cases of metric spaces which will be used later.

7.2.3. The metric space $\mathbb{R}^d$. Consider $S = \mathbb{R}^d$, whose elements are d-tuples $x = (x_1, \dots, x_d)$, $x_k \in \mathbb{R}$, with the Euclidian metric

$$\rho(x, y) := \Big(\sum_{k=1}^{d} |x_k - y_k|^2 \Big)^{\frac{1}{2}} .$$

$(\mathbb{R}^d, \rho)$ is separable and complete. The Borel sets generated by the open sets of the metric ρ are denoted $\mathcal{B}(\mathbb{R}^d)$, and are the same as those obtained by considering the product σ-algebra generated by rectangles.

As well known when proving the Central Limit Theorem in $d = 1$, weak convergence of probability measures is equivalent to weak convergence of distribution functions. With $d \geq 2$, the same holds, as we now show. For $x, y \in \mathbb{R}^d$, write $x \leq y$ if $x_k \leq y_k$ for all $k = 1, \ldots, d$. An interval is a set of the form $(a, b] = \{y : a_k < y_k \leq b_k, k = 1, \ldots, d\}$. For any μ, probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, define the **distribution function**

$$F(x) := \mu(\{y : y \leq x\}).$$

By definition, F is non-decreasing in x, and it is easy to see that F is continuous from above, i.e. for all $\epsilon > 0$ there exists a $\delta > 0$ such that $x \leq y \leq x + \delta e$ ($e = (1, 1, \ldots, 1)$) implies $F(x) \leq F(y) \leq F(x) + \epsilon$. Therefore, F is continuous at x if and only if it is continuous from below at x, that is if for all $\epsilon > 0$ there exists a $\delta > 0$ such that $x - \delta e \leq y \leq x$ implies $F(x) - \epsilon \leq F(y) \leq F(x)$ or, equivalently, if $F(x) = \sup_{\delta > 0} F(x - \delta e) = \mu(\{y : y < x\})$. Therefore, F is continuous at x if and only if $\mu(\partial\{y : y \leq x\}) = 0$, i.e. if $\{y : y \leq x\}$ is a continuity set of μ.

Let F_n denote the distribution function of μ_n.

THEOREM 7.2.6. *The class of sets of the form $\{y : y \leq x\}$, $x \in \mathbb{R}^d$, form a convergence-determining class for weak convergence in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. In other words, $\mu_n \overset{w}{\Rightarrow} \mu$ if and only if $F_n(x) \to F(x)$ at each continuity point of F.*

PROOF. If $\mu_n \overset{w}{\Rightarrow} \mu$ then by (5) of the Portmanteau Theorem and by what we just said, $F_n(x) \to F(x)$ at each x where $\mu(\partial\{y : y \leq x\}) = 0$, i.e. where F is continuous.
On the other hand, let $\mathcal{U}$ denote the set of bounded intervals $(a, b]$. Then $\mathcal{U}$ satisfies (1) and (2) of Theorem 7.2.4, and so $\mathcal{U}$ is a convergence determining class: if $\mu_n((a, b]) \to \mu((a, b])$ for all $(a, b] \in \mathcal{U}$ with $\mu(\partial(a, b]) = 0$, then $\mu_n \overset{w}{\Rightarrow} \mu$. To transform this into a condition on the sets $\{y : y \leq x\}$, first observe that for any $r = 1, \ldots, d$, there can exist at most countably many hyperplanes $H_t := \{y : y_r = t\}$ with $\mu(H_t) > 0$ (repeat an argument similar to the one of Lemma 7.2.2).

Now since each $\mu_n((a,b])$ can be expressed as a sum $\sum_{k=1}^{2^d} \pm F_n(x_k)$, where the x_ks are the 2^d corners of $(a,b]$, and since when $(a,b] \in \mathcal{U}$, F is continuous at each of these points, we have that $\mu_n((a,b]) \to \mu((a,b])$ for all $(a,b] \in \mathcal{U}$ if and only if $F_n(x_k) \to F(x)$ for each k. □

7.2.4. The metric space $\mathbb{R}^{\mathbb{N}}$. Consider $S = \mathbb{R}^{\mathbb{N}}$, the set of real sequences $x = (x(1), x(2), \dots)$ where $x(k) \in \mathbb{R}$. We first introduce a metric on $\mathbb{R}$ equivalent to the Euclidian metric $|\cdot|$: for all $\alpha, \beta \in \mathbb{R}$,

$$\rho_0(\alpha, \beta) := \frac{|\alpha - \beta|}{1 + |\alpha - \beta|}\,.$$

The advantage is that $0 \le \rho_0 < 1$. Then define, for $x, y \in \mathbb{R}^{\mathbb{N}}$,

$$\rho(x,y) := \sum_{k \ge 1} \frac{1}{2^k} \rho_0(x(k), y(k))\,. \tag{7.2.6}$$

Lemma 7.2.5. $(\mathbb{R}^{\mathbb{N}}, \rho)$ *is complete, separable.*

Proof. Let $(x_n)_{n\ge1}$ be a Cauchy sequence, i.e. $\rho(x_n, x_m) \to 0$ when $n, m \to \infty$. Then for each component k, $\rho_0(x_n(k), x_m(k)) \le 2^k \rho(x_n, x_m)$, and therefore $x(k) := \lim_{n\to\infty} x_n(k)$ exists. Let $x := (x(1), x(2), \dots)$. We have, by the Lemma of Fatou,

$$\rho(x, x_n) \le \liminf_{m\to\infty} \sum_{k\ge1} \frac{1}{2^k} \rho_0(x_m(k), x_n(k)) = \liminf_{m\to\infty} \rho(x_m, x_n)\,,$$

which is arbitrarily small when n is large. Therefore, $x_n \to x$, which shows that $(\mathbb{R}^{\mathbb{N}}, \rho)$ is complete. To see that it is separable, consider the set $T := \bigcup_{n\ge1} T_n$, where T_n is the set of elements $x = (x(1), x(2), \dots)$ whose first n coordinates are rational, and $x(k) = 0$ for all $k > n$. Clearly, T is dense in $\mathbb{R}^{\mathbb{N}}$. □

Remember from Chapter 3 that cylinders $\mathcal{C} \subset \mathbb{R}^{\mathbb{N}}$ are sets of the form $\pi_n^{-1}(B)$, for some Borel set $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$, where $\pi_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^n$ is the canonical projection (3.1.2).

Lemma 7.2.6. $\mathcal{C}$ *is a convergence-determining, hence determining.*

Proof. Clearly, the projections $\pi_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^n$ are continuous (when $\mathbb{R}^{\mathbb{N}}$ and $\mathbb{R}^m$ are equipped with their respective metrics), hence measurable. Therefore, $\mathcal{C} \subset \mathcal{B}(\mathbb{R}^{\mathbb{N}})$. Consider an open sphere $B_\epsilon(x)$. Let K be large enough so that $\sum_{k>K} 2^{-k} < \frac{\epsilon}{2}$. Consider the cylinder $C = \{y : |y(k) - x(k)| < \frac{\epsilon}{4}, k = 1, \dots, K\}$. Clearly $x \in C \equiv \text{int}C \subset$

$\bar{C} \subset B_\epsilon(x)$. Since $\mathbb{R}^{\mathbb{N}}$, Theorem 7.2.4 implies that $\mathcal{C}$ is a convergence-determining class. □

7.2.5. The metric space C. Consider C (a short notation for $C[0,1]$), the set of continuous (hence bounded) functions $x : [0,1] \to \mathbb{R}$. Consider the metric

$$\rho(x,y) := \sup\{|x(t) - y(t)| : 0 \le t \le 1\}. \tag{7.2.7}$$

LEMMA 7.2.7. (C, ρ) *is complete, separable.*

PROOF. Let $(x_n)_{n\ge1}$ be a Cauchy sequence in C. Then for all $t \in [0,1]$, $|x_n(t) - x_m(t)| \le \rho(x_n, x_m)$. Therefore, $x(t) := \lim_{n\to\infty} x_n(t)$ exists. We show that $x \in \mathsf{C}$. Take $\epsilon > 0$. Let then m, n be such that $\rho(x_m, x_n) \le \epsilon$. For all $t \in [0,1]$, $|x(t) - x_n(t)| = \lim_m |x_m(t) - x_n(t)| \le \epsilon$. Therefore, $\rho(x, x_n) \le \epsilon$. This shows that the convergence $x_n \to x$ is uniform. Since each x_n is continuous, x also is. So C is complete. Then, let $\mathcal{D}_n$ be the set of functions which take rational values at the points $\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\}$, and which is linear on each of the intervals $[\frac{k-1}{n}, \frac{k}{n}]$. Let $\mathcal{D} := \bigcup_{n\ge1} \mathcal{D}_n$. Since the elements of C are uniformly continuous, it is easy to see that $\mathcal{D}$ is dense in C: C is separable. □

Let $\mathcal{B}(\mathsf{C})$ denote the Borel σ-algebra generated by ρ. For each ordered set $N = (t_1, t_2, \dots, t_n)$, $0 \le t_1 < \cdots < t_n \le 1$, consider the projection $\pi_N : \mathsf{C} \to \mathbb{R}^n$ defined by

$$\pi_N(x) := (x(t_1), \dots, x(t_n)).$$

As in the previous example, we consider the family $\mathcal{C}$ of cylinders, that is sets of the form $\pi_N^{-1}(B)$, with $B \in \mathcal{B}(\mathbb{R}^n)$. It is clear that $\mathcal{C}$ is an algebra. Let us see that they also generate the Borel σ-algebra generated by ρ.

LEMMA 7.2.8. $\sigma(\mathcal{C}) = \mathcal{B}(\mathsf{C})$.

PROOF. Since C is separable, each open set can be written as a countable set of open spheres $B_\epsilon(x)$ (Lemma 7.2.3). But each such sphere can be written as

$$B_\epsilon(x) = \bigcup_{n\ge1} \big\{y \in \mathsf{C} : \rho(y,x) \le \epsilon - 1/n\big\}.$$

But, since each $x \in \mathsf{C}$ is also uniformly continuous, for all $\delta > 0$,

$$\big\{y \in \mathsf{C} : \rho(y,x) \le \delta\big\} = \bigcup_{m\ge1} \big\{y \in \mathsf{C} : |y(i/m) - x(i/m)| \le \delta\big\},$$

which is an intersection of cylinders. That is, each open set belongs to $\sigma(\mathsf{C})$, and therefore $\mathcal{B}(\mathsf{C}) \subset \sigma(\mathcal{C})$. Then, since the projections π_N are obviously continuous, $\mathcal{C} \subset \mathcal{B}(\mathsf{C})$, and so $\sigma(\mathcal{C}) \subset \mathcal{B}(\mathsf{C})$. □

LEMMA 7.2.9. *Cylinders are determining,* but not *convergence-determining.*

PROOF. As we just saw, $\mathcal{C}$ generates $\mathcal{B}(\mathsf{C})$. Since it is an algebra, it forms a determining class (Carathéodory's Theorem). To see that $\mathcal{C}$ is not convergence-determining, consider the sequence μ_n on $(\mathsf{C}, \mathcal{B}(\mathsf{C}))$ defined as follows: μ_n is the Dirac mass at x_n, which is defined by

$$x_n(t) := \begin{cases} nt & \text{if } 0 \le t \le n^{-1}\,, \\ 2 - nt & \text{if } n^{-1} \le t \le 2n^{-1}\,, \\ 0 & \text{if } 2n^{-1} \le t \le 1\,. \end{cases}$$

Let μ be the Dirac mass at $x \equiv 0$. Then clearly, $\mu_n(C) \to \mu(C)$ for any cylinder C. Nevertheless, let $A := B_{\frac{1}{2}}(0)$, where 0 denotes the function $x \equiv 0$. Then $\mu_n(A) = 0$ for all n, and $\mu(A) = 1$. But $\partial A = \{y \in \mathsf{C} : \sup_{t\in[0,1]} |y(t)| = \frac{1}{2}\}$, and so $\mu(A) = 0$. Therefore, μ_n does not converge weakly to μ. □

7.3. Prohorov's Theorem (EMPTY)

THEOREM 7.3.1. *Let $(\mu_n)_{n\ge1}$ be a sequence of probability measures on a separable metric space S. If $(\mu_n)_{n\ge1}$ is tight, then it is relatively compact.*

7.4. The Wiener Measure (EMPTY)

We described the basic metric properties of C in Section 7.2.5.

7.5. The Invariance Principle (EMPTY)

7.6. Sample Path Properties (EMPTY)

7.7. Exercises

EXERCISE 7.1. Billingsley ex. 6 p. 11. Show that $\mathcal{S}$ contains all compact sets. Show that $\mathcal{S}$ is generated by either of the following classes: open sets, closed sets, compact sets.

EXERCISE 7.2. Find a counter-example where the indicator of an open set cannot be appxoximated by a uniformly continuous functions.

EXERCISE 7.3. Find an example of a sequence x_n in $\mathbb{R}^{\mathbb{N}}$ such that $x(k) := \lim_n x_n(k)$ exists for all k, but

EXERCISE 7.4. Is $C(\mathbb{R})$ separable?

EXERCISE 7.5. Show that P is tight if and only if it has σ-compact support.

EXERCISE 7.6. Shiryayev p. 318. Let $(\mu_\alpha)_{\alpha \in I}$ be a family of Gaussian measures on the line with parameters m_α and σ_α^2. Show that $(\mu_\alpha)_{\alpha \in I}$ is tight if and only if $|m_\alpha| \leq m$ and $\sigma_\alpha^2 \leq \sigma$ for all $\alpha \in I$.

EXERCISE 7.7. Shiryayev p. 318. Construct examples of tight and non-tight families on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$.

APPENDIX A

Entropy

In this section we introduce the notion of entropy associated to a random experiment. Our aim is to be as broad as possible since later the concept of entropy will be used in various different situations, for example in the case where the random experiment is the joint realization of n random variables with stationary distribution.

Entropy is a number which gives a convenient quantization of the *predictability* or *unpredictability* of a given random experiment. Equivalently, entropy is a measure of *randomness*: the more random, the less predictable.

Consider a random experiment modelized by some probability space $(\Omega, \mathcal{F}, P)$. Suppose our aim is to make a reasonable prediction about the outcome of the experiment. We will do so *assuming we know the probability* P; there is no inference here. Clearly, a reasonable *a priori* prediction about the outcome $\omega \in \Omega$ is possible when the measure P concentrates inhomogeneously on certain subsets of Ω. The extreme case is when P is a Dirac mass at some $\omega_0 \in \Omega$; the absence of randomness allows to make an essentially perfect prediction about the result: "the outcome will be ω_0". Since the outcome is almost surely equal to the a priori prediction, nothing interesting is learnt from the result. At the other extreme, the most unpredictable experience is when the measure P is uniform over Ω. In this case, the outcome of the experiment will be most probably very different from any a priori prediction; one says that the outcome of the experiment *produces information.*

Entropy allows to quantify precisely this information production, but is defined naturally for experiments with a finite number of possible outcomes. Therefore, some approximation procedure is necessary in the case where Ω has an infinite number of outcomes [1]. A natural way

[1] This will be the case in particular when Ω has a continuous structure, for example when $\Omega = [0, 1]$, in which case each outcome usually has zero probability.

to do this approximation is to introduce a coarsening of the possible results of the experiment. This is done by introducing what is commonly called a finite scheme [**Khi57**]. A finite scheme is nothing but a partition $\mathcal{A}$ of Ω into a finite number of sets $A_k \in \mathcal{F}$: $\mathcal{A} = \{A_1, \dots, A_n\}$, with $\bigcup_{k=1}^n A_k = \Omega$, $A_k \cap A_{k'} = \emptyset$ if $k \neq k'$. Rather than the result of the random experiment itself, i.e. ω, one is interested in the atom A_k of the partition $\mathcal{A}$ to which ω belongs. The coarse-grained result of the experience is therefore an index $k = k(\omega) \in \{1, 2, \dots, n\}$, giving the unique atom $A_k \ni \omega$, and can be considered as partial knowledge about the result of the experiment. To the finite scheme $\mathcal{A} = \{A_1, \dots, A_n\}$ corresponds a set of probabilities $p_1 := P(A_1), \dots, p_n := P(A_n)$, satisfying $\sum_{k=1}^n p_k = 1$ (remember we are assuming that P is known).

With this coarse-grained description in mind, we can move on to the definition of entropy associated to a finite scheme. The following definition was proposed by C.E. Shannon in [**Sha48**] as a measure of the average information produced by one realization of a random experiment with outcomes of respective probabilities $p_1, \dots, p_n$:

DEFINITION A.0.1. *The entropy of a probability distribution* $(p_1, \dots, p_n)$ *associated to a finite scheme is defined by*

$$H(p_1, p_2, \dots, p_n) := -\sum_{k=1}^{n} p_k \log p_k \,, \tag{A.0.1}$$

where it is assumed that the logarithm is with respect to the base 2*, and where we make the convention that* $0 \log 0 := 0$.

When we wish to express explicitely that the entropy is associated to the scheme $\mathcal{A}$, each atom A_k of which has probability $p_k = P(A_k)$, we will write

$$H_P(\mathcal{A}) := H(P(A_1), \dots, P(A_n)) = -\sum_{A \in \mathcal{A}} P(A) \log P(A) \,. \tag{A.0.2}$$

Let us verify that this definition suits our requirements for production of information, as discussed above. First, observe that H is a positive quantity which attains its minimal value $H = 0$ exactly when all p_is are zero except one (when the measure is concentrated on a single event, the outcome doesn't produce any information). Then, as expected, we show that entropy is maximal for the uniform distribution $(p_1, \dots, p_n) = (\frac{1}{n}, \dots, \frac{1}{n})$:

$$H(p_1, \dots, p_n) \leq H\big(\frac{1}{n}, \dots, \frac{1}{n}\big) = \log n \,. \tag{A.0.3}$$

Namely, by introducing the concave function $\psi(x) := -x \log x$ for $x \in (0, 1)$, $\psi(0) = \psi(1) := 0$, one can write

$$\begin{aligned} H(p_1, \dots, p_n) = \sum_{k=1}^{n} \psi(p_k) &= n \sum_{k=1}^{n} \frac{1}{n} \psi(p_k) \\ &\leq n\psi(\frac{1}{n}) = H\big(\frac{1}{n}, \dots, \frac{1}{n}\big) \,. \end{aligned}$$

Therefore, (A.0.3) fullfills our previous requirement: unpredictability is largest for equiprobable events. Moreover, in Shannon's own words, *any change towards equalization of the probabilities* $(p_1, \dots, p_n)$ *increases* $H(p_1, \dots, p_n)$. This can be seen by explicit calculation, by considering the variation of H when, say $p_1(s) = p_0 + s$, $p_2(s) = p_0 - s$, and where all the other $n - 2$ variables are kept fixed:

$$\begin{aligned} \frac{d}{ds} H &= \frac{d}{ds} \big[-(p_0 + s) \log(p_0 + s) - (p_0 - s) \log(p_0 - s) \big] \\ &= \log \frac{p_0 - s}{p_0 + s} \,, \end{aligned}$$

which is < 0 when $s > 0$. This means that if s decreases, i.e. when p_1 and p_2 tend to equalize, then the entropy increases.

A further property of H is that it is a concave function of $(p_1, \dots, p_n)$. To gain geometric intuition, observe that we are only interested in the restriction of H to the simplex

$$\mathcal{P} = \Big\{ \mathbf{p} = (p_1, \dots, p_n) : p_k \geq 0, \sum_{k=1}^{n} p_k = 1 \Big\} \subset \mathbb{R}^n \,.$$

Any $\mathbf{p} \in \mathcal{P}$ can thus be considered as a convex combination of the extreme elements of $\mathcal{P}$, which are the unit vectors of the canonical basis of $\mathbb{R}^n$: $\mathbf{e}_1, \dots, \mathbf{e}_n$. At each extreme element $\mathbf{e}_i$, $H(\mathbf{e}_i) = 0$. Moreover, for any $\mathbf{p}_1, \mathbf{p}_2 \in \mathcal{P}$, $0 \leq \lambda \leq 1$,

$$H(\lambda \mathbf{p}_1 + (1 - \lambda)\mathbf{p}_2) \geq \lambda H(\mathbf{p}_1) + (1 - \lambda) H(\mathbf{p}_2) \,.$$

This follows from the concavity of the function $\psi(x)$ introduced before. One thus has a picture of H as a concave function on $\mathcal{P}$ which attains its maximum ($\log n$) at the barycenter of $\mathcal{P}$, namely $(\frac{1}{n}, \ldots, \frac{1}{n})$.

Further properties of H are most naturally formulated in terms of schemes. Consider two schemes, say $\mathcal{A}$ and $\mathcal{B}$, giving a coarse-grained description of the same random experiment. Consider the scheme $\mathcal{A} \vee \mathcal{B}$ defined by

$$\mathcal{A} \vee \mathcal{B} := \{A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\} .$$

This scheme is a refinement of both $\mathcal{A}$ and $\mathcal{B}$, since each of its atoms contains a double information ($\omega \in A$ *and* $\omega \in B$ for some couple (A, B)); its entropy equals, by definition,

$$H_P(\mathcal{A} \vee \mathcal{B}) = -\sum_{(A,B)} P(A \cap B) \log P(A \cap B) . \tag{A.0.4}$$

When $\mathcal{A}$ and $\mathcal{B}$ are independent, that is if $P(A \cap B) = P(A)P(B)$ for all pair $A \in \mathcal{A}$, $B \in \mathcal{B}$, a simple computation yields

$$H_P(\mathcal{A} \vee \mathcal{B}) = H_P(\mathcal{A}) + H_P(\mathcal{B}) . \tag{A.0.5}$$

This property is called extensivity. In the general case (i.e. without assuming independence) we always have

$$H_P(\mathcal{A} \vee \mathcal{B}) \geq -\sum_{(A,B)} P(A \cap B) \log P(A) = H_P(\mathcal{A}) .$$

The exact excess in the previous inequality is measured by the relative entropy of $\mathcal{B}$ with respect to $\mathcal{A}$:

$$H_P(\mathcal{B}|\mathcal{A}) := H_P(\mathcal{A} \vee \mathcal{B}) - H_P(\mathcal{A}) , \tag{A.0.6}$$

which measures the average excess of information produced by the outcome of the experiment $\mathcal{A} \vee \mathcal{B}$ over the information produced by the experiment $\mathcal{A}$. As can be verified explicitely,

$$H_P(\mathcal{B}|\mathcal{A}) = -\sum_{(A,B)} P(A \cap B) \log P(B|A) \equiv \sum_{A} P(A) H_{P_A}(\mathcal{B}) , \tag{A.0.7}$$

where $P_A(\cdot)$ is the conditional probability $P(\cdot|A)$. When written as:

$$H_P(\mathcal{A} \vee \mathcal{B}) = H_P(\mathcal{A}) + H_P(\mathcal{B}|\mathcal{A}) , \tag{A.0.8}$$

(A.0.6) is seen to be (up to a logarithm) the information-theoretic equivalent of the well known probabilistic expression

$$P(A \cap B) = P(A)P(B|A) .$$

We will define more notions related to entropy in subsequent sections.

A.1. Entropy as Average of Pointwise Information

Until now we introduced entropy as an *average* quantity which quantifies the unpredictability of a coarse-grained random experiment. Let us see here how Definition (A.0.1) arises naturally from a basic set of conditions a measure of information should satisfy. To start with, assume a partial information about the outcome of the experiment $\omega \in \Omega$ is that ω belongs to some measurable set $A \in \mathcal{A}$, where $\mathcal{A}$ is a finite scheme. We want to define a number I_A which quantifies the partial information "$\omega \in A$". Once this will be done, we will define a random variable giving the information produced by the outcome of ω

$$I_{\mathcal{A}}(\omega) := \sum_{A \in \mathcal{A}} I_A 1_A(\omega) \,.$$

How should I_A be defined? We shall naturally ask for I_A to be positive, and at our coarse-grained level of description, we have no reason to make I_A depend on other characteristics of A other than its probability $P(A)$ (remember that this number will be the *same* for any other outcome $\omega' \in A$). Therefore we must have

$$I_A = \varphi(P(A)) \,,$$

for some real non-negative function $\varphi = \varphi(x)$. To find a proper function φ, we turn to the essential property that I_A should satisfy in order to properly represent information, namely *extensivity*. Assume we have a double information of the form "$\omega \in A$ *and* $\omega \in B$". In the case where A and B are independent, it seems reasonable to impose that [2]

$$I_{A \cap B} = I_A + I_B \,.$$

In terms of φ, this means $\varphi(P(A)P(B)) = \varphi(P(A)) + \varphi(P(B))$. A convenient choice for the function φ is thus $\varphi(x) := -\log x$, which yields

$$I_{\mathcal{A}}(\omega) = \sum_{A \in \mathcal{A}} \big(- \log P(A) \big) 1_A(\omega) \,.$$

Finally, we can define **entropy** as the expected information produced by a realization of the coarse-grained experiment, namely the expectation

[2] This extensivity condition is made clear by considering a simple example. Suppose we are drawing a Brasilian ω at random. Two possible partial informations for this experience are for example $A = \{\omega \text{ is Carioca}\}$ and $B = \{\omega \text{ is Atleticano}\}$. Then assuming these two events are independent, the double information $I_{A \cap B}$ should thus be additive and equal the sum of I_A with I_B.

of $I_{\mathcal{A}}$, which is

$$H_P(\mathcal{A}) := E_P(I_{\mathcal{A}}) \equiv -\sum_{A\in\mathcal{A}} P(A)\log P(A)\,.$$

We thus recover (A.0.1).

A.2. Entropy of Discrete Random Variables

Consider a discrete random variable $X : \Omega \to \mathbb{R}$ taking values in a finite alphabet $\mathsf{A} = \{x_1, \dots, x_n\}$. The entropy of X, denoted $H(X)$, is just the entropy of the scheme $\mathcal{A}_X = \{A_1, \dots, A_n\}$ whose atoms are $A_k := \{\omega : X(\omega) = x_k\}$, with probability $p_k = P(A_k) = P(X = x_k)$. For example, in the case where X takes two values, say 0 and 1, with probabilities $P(X = 1) = p$, $P(X = 0) = 1 - p$, then

$$H(X) = -p\log p - (1-p)\log(1-p)\,,$$

which is concave in p, and attains its maximum at $p = \frac{1}{2}$.

The joint entropy of a couple of two discrete random variables (X, Y) is naturally defined by the entropy of the scheme $\mathcal{A}_X \vee \mathcal{A}_Y$, and is given explicitely by

$$H(X,Y) = -\sum_{k,l} P(X = x_k, Y = y_l)\log P(X = x_k, Y = y_l)\,.$$

When X and Y are independent,

$$H(X,Y) = H(X) + H(Y)\,.$$

In general, the relative entropy of X with respect to Y is defined by

$$H(X|Y) := H(X,Y) - H(Y)\,. \tag{A.2.1}$$

We now turn to the simplest use of entropy in the study of stochastic processes. Namely, let $X_1, X_2, \dots$ be a sequence of i.i.d. random variables taking values in a finite alphabet A. The Strong Law of Large Numbers (SLLN) reads

$$\frac{X_1 + \cdots + X_n}{n} \to E[X_1] \qquad \text{a.s.}\,,$$

and the Weak Law of Large Numbers (WLLN) states that for large n, most of the outcomes of the random variables $X_1, \dots, X_n$ have an empirical mean which is close to $E[X_1]$: for all $\epsilon > 0$,

$$P\Big(\Big|\frac{X_1 + \cdots + X_n}{n} - E[X_1]\Big| \geq \epsilon\Big) \to 0\,.$$

The WLLN gives a concentration property concerning the typical outcomes of a random realization of a sequence $(X_1, \dots, X_n)$: with high probability, the empirical average $\frac{X_1+\dots+X_n}{n}$ is close to its theoretical mean $E[X_1]$. More precisely one has, for any $\epsilon > 0$, any $\delta > 0$ and large enough n, a set of **mean-typical sequences** of size n denoted $A_n(\epsilon, \delta)$, with the following properties:

(1) $P(A_n(\epsilon, \delta)) \geq 1 - \delta$,
(2) Each $(X_1, \dots, X_n) \in A_n(\epsilon, \delta)$ satisfies

$$\Big|\frac{X_1 + \dots X_n}{n} - E[X_1]\Big| < \epsilon\,.$$

The Asymptotic Equipartition Property (AEP) which we present hereafter gives another look at this concentration phenomenon, but from the point of view of the typical probability of the sequence $X_1, \dots, X_n$. Define the joint distribution $p(x_1, \dots, x_n) := P(X_1 = x_1, \dots, X_n = x_n)$, and following random variable

$$p(X_1, \dots, X_n)(\omega) := p(X_1(\omega), \dots, X_n(\omega))\,,$$

called the **empirical joint distribution** of the outcome $X_1, \dots, X_n$.

THEOREM A.2.1 (Asymptotic Equipartition Property for i.i.d. random variables). *Assume the sequence $X_1, X_2, \dots$ is i.i.d. Then*

$$-\frac{1}{n}\log p(X_1, \dots, X_n) \to H(X_1)\,, \qquad a.e.$$

PROOF. Since the variables X_k are independent, we have $p(x_1, \dots, x_n) = P(X_1 = x_1) \cdots P(X_n = x_n)$. By applying the SLLN to the sequence $Y_k(\omega) := -\log P(X_k = X_k(\omega))$,

$$-\frac{1}{n}\log p(X_1, \dots, X_n) = \frac{1}{n}\sum_{k=1}^{n} Y_k \to E[Y_1]\,, \text{ a.e.}$$

But $E[Y_1] \equiv H(X_1)$. □

Since almost everywhere convergence implies convergence in probability, a weaker form of the AEP is that for all $\epsilon > 0$ and all $\delta > 0$,

$$P\Big(\Big| -\frac{1}{n}\log p(X_1, \dots, X_n) - H(X_1)\Big| \geq \epsilon\Big) \leq \delta$$

for large enough n. This basically means that most of the sequences $X_1, \dots, X_n$ have equal probability, roughly equal to $2^{-H(X_1)n}$. More precisely one has, for any $\epsilon > 0$, any $\delta > 0$ and large enough n, a set of probability-typical sequences of size n denoted $B_n(\epsilon, \delta)$, with the following properties:

(1) $P(B_n(\epsilon, \delta)) \geq 1 - \delta$,
(2) Each $(X_1, \ldots, X_n) \in B_n(\epsilon, \delta)$ satisfies

$$2^{(-H(X_1)-\epsilon)n} \leq p(X_1, \ldots, X_n) \leq 2^{(-H(X_1)+\epsilon)n}.$$

(3) $B_n(\epsilon, \delta)$ contains at most $2^{(H(X_1)+\epsilon)n}$ sequences.

This last property of $B_n(\epsilon, \delta)$ has fundamental consequences in information theory (see [**CT06**]).

APPENDIX B

Dynkin Systems

Let Ω be any non-empty set. We denote by 2^Ω the family of all subsets of Ω, including the emptyset.

DEFINITION B.0.1. *A collection $\mathcal{D} \subset 2^\Omega$ is called a* Dynkin System *(or simply* D-system*) if the following conditions hold:*

(1) $\Omega \in \mathcal{D}$.
(2) *If* $A, B \in \mathcal{D}$, $A \subset B$, *then* $B \backslash A \in \mathcal{D}$.
(3) *If* $A_n \in \mathcal{D}$ *for all* $n \geq 1$, $A_n \nearrow A$, *then* $A \in \mathcal{D}$

Observe that D-systems are stable by complementation since $A \in \mathcal{D}$ implies $A^c = \Omega \backslash A \in \mathcal{D}$. Since $B \backslash A = B \cap A^c$, σ-algebras are D-systems, but since D-systems are not necessarily stable under intersections.

LEMMA B.0.1. *A collection $\mathcal{F} \subset 2^\Omega$ is a σ-algebra if and only if it is a D-system stable under intersection.*

PROOF. The "only if" part is trivial. Then, assume $\mathcal{F}$ is a D-system stable under intersection. Let $A, B \in \mathcal{F}$. We have $A \cup B = (A^c \cap B^c)^c = \Omega \backslash (A^c \cap B^c) \in \mathcal{F}$. Let $A_n \in \mathcal{F}$, $B_n := \bigcup_{k=1}^n A_k$. Since $B_n \in \mathcal{F}$ and $B_n \nearrow \bigcup_{n \geq 1} B_n$, we have that $\bigcup_{n \geq 1} B_n \in \mathcal{F}$. This shows that $\mathcal{F}$ is a σ-algebra. □

As can be easily verified, the intersection of an arbitrary family of D-systems is a D-system. Therefore, given any collection $\mathcal{C} \subset 2^\Omega$, one can define the smallest D-system containing $\mathcal{C}$, called the D-system generated by $\mathcal{C}$, denoted $\mathcal{D}(\mathcal{C})$. In practice, it is interesting to compare the D-system $\mathcal{D}(\mathcal{C})$ with the σ-algebra $\sigma(\mathcal{C})$. One clearly has $\mathcal{D}(\mathcal{C}) \subset \sigma(\mathcal{C})$.

THEOREM B.0.1. *If $\mathcal{C} \subset 2^\Omega$ is stable under intersection, then $\mathcal{D}(\mathcal{C}) = \sigma(\mathcal{C})$.*

PROOF. To simplify the notations, denote $\mathcal{D}(\mathcal{C})$ by $\mathcal{D}$ and $\sigma(\mathcal{C})$ by $\mathcal{F}$. We already saw that $\mathcal{D} \subset \mathcal{F}$. To show that $\mathcal{D} \supset \mathcal{F}$, it suffices to verify that $\mathcal{D}$ is a σ-algebra. By Lemma B.0.1, it suffices to verify that

$\mathcal{D}$ is stable under intersection.
Define $\mathcal{D}_1 := \{B \in \mathcal{D} : B \cap C \in \mathcal{D}\, \forall C \in \mathcal{C}\}$. We verify that $\mathcal{D}_1 = \mathcal{D}$. By definition, $\mathcal{D}_1 \subset \mathcal{D}$. To verify that $\mathcal{D}_1 \supset \mathcal{D}$, it suffices to see that $\mathcal{D}_1$ is a D-system containing $\mathcal{C}$. Now $\mathcal{D}_1 \supset \mathcal{C}$ follows from the fact that $\mathcal{C}$ is closed under intersection. This also implies that $\Omega \in \mathcal{D}_1$. Let $B_1, B_2 \in \mathcal{D}_1$, $B_1 \subset B_2$, $C \in \mathcal{C}$. Then

$$(B_2 \backslash B_1) \cap C = B_2 \cap C \cap (B_1^c \cup C^c) = (B_2 \cap C) \backslash (B_1 \cap C) \in \mathcal{D}$$

Then, if $B_n \in \mathcal{D}_1$, $B_n \nearrow B$, then $B \cap C = \bigcup_n (B_n \cap C) \in \mathcal{D}$, logo $B \in \mathcal{D}_1$. This proves that $\mathcal{D}$ is a D-system.
Define $\mathcal{D}_2 := \{A \in \mathcal{D} : A \cap B \in \mathcal{D}\, \forall B \in \mathcal{D}\}$. We verify that $\mathcal{D}_2 = \mathcal{D}$, which will show that $\mathcal{D}$ is stable under intersection. By the first step, $\mathcal{D}_2$ contains $\mathcal{C}$. As before, one can show that $\mathcal{D}_2 = \mathcal{D}$. This shows that $\mathcal{D}$ is stable under intersection, and finishes the proof of the theorem. □

The previous result is usually used in the following form:

Corollary B.0.1. *Let $\mathcal{C} \subset 2^\Omega$ be stable under intersection. If $\mathcal{D}$ is a D-system containing $\mathcal{C}$, then $\mathcal{D} \supset \sigma(\mathcal{C})$.*

The last result is useful to show that the measurable sets of some σ-algebra $\mathcal{F}$ satisfy particular property. An example of application is given in the following proposition and its corollary.

Proposition B.0.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\mathcal{A}_1, \ldots, \mathcal{A}_n$ ($\mathcal{A}_k \subset \mathcal{F}$) be independent collections* [1]*, each of which is stable under intersection. Then the σ-algebras $\sigma(\mathcal{A}_1), \ldots, \sigma(\mathcal{A}_n)$ are independent.*

Proof. Without loss of generality, we can suppose that each $\mathcal{A}_k$ contains Ω. We will show that if $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ are independent and stable under intersection, then $\sigma(\mathcal{A}_1), \mathcal{A}_2, \ldots, \mathcal{A}_n$ are independent (and stable under intersection). The proof then follows by induction. Fix $A_2 \in \mathcal{A}_2, \ldots, A_n \in \mathcal{A}_n$, set $F := A_2 \cap \cdots \cap A_n$ and let $\mathcal{D}_F := \{A \in \mathcal{A}_1 : P(A \cap F) = P(A)P(F)\}$. We have $\mathcal{D}_F \ni \Omega$. Then, let $A, B \in \mathcal{D}_F$ with $A \subset B$: $P((B \backslash A) \cap F) = P(B \backslash A)P(F)$, and so $B \backslash A \in \mathcal{D}_F$. Finally, if $A_n \in \mathcal{D}_F$, $A_n \nearrow A$, then $P(A \cap F) = \lim_n P(A_n \cap F) = \lim_n P(A_n)P(F) = P(A)P(F)$, and so $A \in \mathcal{D}_F$. This shows that $\mathcal{D}_F$ is a D-system. Since $\mathcal{D}_F \supset \mathcal{A}_1$, Corollary B.0.1 gives $\mathcal{D}_F \supset \sigma(\mathcal{A}_1)$. Since this holds for all choice of F, we have shown that $\sigma(\mathcal{A}_1), \mathcal{A}_2, \ldots, \mathcal{A}_n$ are independent. □

[1] Remember that $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are independent if for all $I \subset \{1, 2, \ldots, n\}$, any family $A_i, i \in I$ is independent: $P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$.

COROLLARY B.0.2. *Let $\mathcal{F}_1, \ldots, \mathcal{F}_n$ be independent sub-σ-algebras ($\mathcal{F}_k \subset \mathcal{F}$ for all k). Then, for any $1 \le k \le n$, $\sigma(\mathcal{F}_1, \ldots, \mathcal{F}_k)$ and $\sigma(\mathcal{F}_{k+1}, \ldots, \mathcal{F}_n)$ are independent.*

PROOF. Let $\mathcal{A}$ be the collection of all intersections $\bigcap_{j=1}^{k} A_j$ with $A_j \in \mathcal{F}_j$, and $\mathcal{B}$ be the collection of all intersections $\bigcap_{j=k+1}^{n} B_j$ with $B_j \in \mathcal{F}_j$. Clearly, $\mathcal{A}$ and $\mathcal{B}$ are stable under intersection. By Proposition B.0.1, $\sigma(\mathcal{A})$ and $\sigma(\mathcal{B})$ are independent. But $\sigma(\mathcal{A}) = \sigma(\mathcal{F}_1, \ldots, \mathcal{F}_k)$ and $\sigma(\mathcal{B}) = \sigma(\mathcal{F}_{k+1}, \ldots, \mathcal{F}_n)$. □

COROLLARY B.0.3. *Assume the variables $(X_n)_{n\ge 1}$ are independent. Then for all $k \ge 1$, $\sigma(X_1, \ldots, X_k)$ and $\sigma(X_{k+1}, \ldots)$ are independent.*

PROOF. Let $\mathcal{A} := \sigma(X_1, \ldots, X_k)$, $\mathcal{B} := \bigcup_{j\ge 1} \sigma(X_{k+1}, \ldots, X_{k+j})$. Clearly, both $\mathcal{A}$ and $\mathcal{B}$ are stable under intersection. Now by Corollary B.0.2, $\sigma(X_1, \ldots, X_k)$ and $\sigma(X_{k+1}, \ldots, X_{k+j})$ are independent for all $j \ge 1$. Therefore, $\mathcal{A}$ and $\mathcal{B}$ are independent. By Proposition B.0.1, $\sigma(\mathcal{A})(\equiv \mathcal{A})$ and $\sigma(\mathcal{B})$ are independent. But $\sigma(\mathcal{B}) = \sigma(X_{k+1}, \ldots)$, which proves the lemma. □

Bibliography

[Bau88] H. Bauer, *Probability theory and elements of measure theory*, New-York: Academic Press, 1988.

[Bil65] P. Billingsley, *Ergodic theory and information*, Wiley, New York, 1965.

[Bil68] ______, *Convergence of probability measures*, Wiley, New York, 1968.

[Bil95] ______, *Probability and measure*, Wiley, New York, 1995.

[Bor09] E. Borel, *Les probabilités dénombrables et leurs applications arithmétiques*, Supplemento ai rendiconti delo Circolo Matematico di Palermo **27** (1909), no. 247-271.

[Chu01] K. L. Chung, *A course in probability theory*, Academic Press, 2001.

[CT06] T.M. Cover and J.A. Thomas, *Elements of information theory*, Wiley and Sons, 2006.

[Doe38] W. Doeblin, *Exposé de la théorie des chaînes simples constantes de Markov à nombre fini d'états*, Rev. Math. Union Interbalkanique **2** (1938), 77–105.

[Geo88] H.-O. Georgii, *Gibbs measures and phase transitions*, de Gruyter Studies in Mathematics, 1988.

[GS05] G.R. Grimmett and D.R. Stirzaker, *Probability and random processes*, Oxford University Press, 2005.

[GS06] ______, *One thousand exercises in probability*, Oxford University Press, 2006.

[KH74] L. Kuipers and H.Niederreiter, *Uniform distribution of sequences*, Dover, 1974.

[Khi57] A.I. Khinchin, *Mathematical foundations of information theory*, Dover, 1957.

[Lin92] T. Lindvall, *Lectures on the coupling method*, Dover, 1992.

[Nev70] J. Neveu, *Chaînes de Markov*, Textos de matemática vol. 19, Univ. Federal de Pernambuco, 1970.

[Pet00] K. Petersen, *Ergodic theory*, Cambridge Studies in Advanced Mathematics, 2000.

[R.88] Durrett R., *Probability: Theory and examples*, Duxbury Press, 1988.

[Rév05] P. Révész, *Random walk in random and non-random environments*, World Scientific, 2005.

[RS85] J.R. Romano and A.F. Siegel, *Counterexamples in probability and statistics*, Wadsworth & Brooks/Cole Statistics and Probability Series, 1985.

[Sha48] C.E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal **27** (1948), 379–423.

[Shi84] A. N. Shiryayev, *Probability*, Springer-Verlag, 1984.

[Str93] D. W. Stroock, *Probability theory; an analytic view*, Cambridge University Press, 1993.

[Str05] ______, *An introduction to markov processes*, Springer, 2005.

[Var00] S.R.S. Varadhan, *Probability theory*, Courant Lecture Notes, 2000.

[Wal75] P. Walters, *Ergodic theory - introductory lectures*, vol. 458, Lecture Notes in Mathematics, 1975.

[Wil91] D. Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, 1991.

Index