Advanced Probability

University of Cambridge,
Part III of the Mathematical Tripos

Michaelmas Term 2006

Grégory Miermont[1]

[1]CNRS & Laboratoire de Mathématique, Equipe Probabilités, Statistique et Modélisation, Bt. 425, Université Paris-Sud, 91405 Orsay, France

Contents

Chapter 1

Conditional expectation

1.1 The discrete case

Let $(\Omega, \mathcal{F}, P)$ be a probability space. If $A, B \in \mathcal{F}$ are two events such that $P(B) > 0$, we define the conditional probability of A given B by the formula

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

We interpret this quantity as the probability of the event A given the fact that B is realized. The fact that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

is called *Bayes' rule*. More generally, if $X \in L^1(\Omega, \mathcal{F}, P)$ is an integrable random variable, we define

$$E[X|B] = \frac{E[X \mathbb{1}_B]}{P(B)},$$

the conditional expectation of X given B.

Example. Toss a fair die (probability space $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $P(\{\omega\}) = 1/6$, for $\omega \in \Omega$) and let $A = \{\text{the result is even}\}, B = \{\text{the result is less than or equal to } 2\}$. Then $P(A|B) = 1/2, P(B|A) = 1/3$. If $X = \omega$ is the result, then $E[X|A] = 4, E[X|B] = 3/2$.

Let $(B_i, i \in I)$ be a countable collection of disjoint events, such that $\Omega = \bigcup_{i \in I} B_i$, and $\mathcal{G} = \sigma\{B_i, i \in I\}$. If $X \in L^1(\Omega, \mathcal{F}, P)$, we define a random variable

$$X' = \sum_{i \in I} E[X|B_i] \mathbb{1}_{B_i},$$

with the convention that $E[X|B_i] = 0$ if $P(B_i) = 0$.

The random variable X' is integrable, since

$$E[|X'|] = \sum_{i \in I} P(B_i)|E[X|B_i]| = \sum_{i \in I} P(B_i) \frac{|E[X \mathbb{1}_{B_i}]|}{P(B_i)} \leq E[|X|].$$

Moreover, it is straightforward to check:

1. X' is $\mathcal{G}$-measurable, and

2. for every $B \in \mathcal{G}$, $E[\mathbb{1}_B X'] = E[\mathbb{1}_B X]$.

Example. If $X \in L^1(\Omega, \mathcal{F}, P)$ and Y is a random variable with values in a countable set E, the above construction gives, by letting $B_y = \{Y = y\}, y \in E$, which partitions $\sigma(Y)$ into measurable events, a random variable

$$E[X|Y] = \sum_{y \in E} E[X|Y = y]\mathbb{1}_{\{Y=y\}}.$$

Notice that the value taken by $E[X|Y = y]$ when $P(Y = y) = 0$, which we have fixed to 0, is actually irrelevant to define $E[X|Y]$, since a random variable is always defined up to a set of zero measure. It is important to keep in mind that conditional expectations are always *a priori* only defined up to a zero-measure set.

1.2 Conditioning with respect to a σ-algebra

We are now going to define the conditional expectation given a sub-σ-algebra of our probability space, by using the properties 1. and 2. of the previous paragraph. The definition is due to Kolmogorov.

Theorem 1.2.1 *Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra, and $X \in L^1(\Omega, \mathcal{F}, P)$. Then there exists a random variable X' with $E[|X'|] < \infty$ such that the following two* characteristic property *are verified:*

1. X' is $\mathcal{G}$-measurable

2. for every $B \in \mathcal{G}$, $E[\mathbb{1}_B X'] = E[\mathbb{1}_B X]$.

Moreover, if X'' is another such random variable, then $X' = X''$ a.s. We denote by $E[X|\mathcal{G}] \in L^1(\Omega, \mathcal{G}, P)$ the class of random variable X'. It is called the conditional expectation of X given $\mathcal{G}$.

Otherwise said, $E[X|\mathcal{G}]$ is the unique element of $L^1(\Omega, \mathcal{G}, P)$ such that $E[\mathbb{1}_B X] = E[\mathbb{1}_B E[X|\mathcal{G}]]$ for every $B \in \mathcal{G}$. Equivalently, an approximation argument allows to replace 2. in the statement by:

2'. For every bounded $\mathcal{G}$-measurable random variable Z, $E[ZX'] = E[ZX]$.

Proof of the uniqueness. Suppose X' and X'' satisfy the two conditions of the statement. Then $B = \{X' > X''\} \in \mathcal{G}$, and therefore

$$0 = E[\mathbb{1}_B(X - X)] = E[\mathbb{1}_B(X' - X'')],$$

which shows $X' \le X''$ a.s., the reverse inequality is obtained by symmetry. □

The existence will need two intermediate steps.

1.2.1 The L^2 case

We first consider L^2 variables. Suppose that $X \in L^2(\Omega, \mathcal{F}, P)$ and let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Notice that $L^2(\Omega, \mathcal{G}, P)$ is a closed vector subspace of the Hilbert space $L^2(\Omega, \mathcal{F}, P)$. Therefore, there exists a unique random variable $X' \in L^2(\Omega, \mathcal{G}, P)$ such that $E[Z(X - X')] = 0$ for every $Z \in L^2(\Omega, \mathcal{G}, P)$, namely, X' is the orthogonal projection of X onto $L^2(\Omega, \mathcal{G}, P)$. This shows the previous theorem in the case $X \in L^2$, and in fact $E[\cdot|\mathcal{G}] : L^2 \to L^2$ is the orthonormal projector onto $L^2(\Omega, \mathcal{G}, P)$, and hence is linear.

It follows from the uniqueness statement that the conditional expectation has the following nice interpretation in the L^2 case: $E[X|\mathcal{G}]$ is the $\mathcal{G}$-measurable random variable that best approximates X. It is useful to keep this intuitive idea even in the general L^1 case, although the word "approximates" becomes more fuzzy.

Notice that $X' := E[X|\mathcal{G}] \geq 0$ a.s. whenever $X \geq 0$ since (notice $\{X' < 0\} \in \mathcal{G}$)

$$E[X \mathbb{1}_{\{X'<0\}}] = E[X' \mathbb{1}_{\{X'<0\}}],$$

and the left-hand side is non-negative while the right hand-side is non-positive, entailing $P(X' < 0) = 0$. Moreover, it holds that $E[E[X|\mathcal{G}]] = E[X]$, because it is the scalar product of X against the constant function $\mathbb{1} \in L^2(\Omega, \mathcal{G}, P)$.

1.2.2 General case

Now let $X \geq 0$ be any non-negative random variable (not necessarily integrable). Then $X \wedge n$ is in L^2 for every $n \in \mathbb{N}$, and $X \wedge n$ increases to X pointwise. Therefore, the sequence $E[X \wedge n|\mathcal{G}]$ is an (a.s.) increasing sequence, because $X \wedge n - X \wedge (n-1) \geq 0)$ and by linearity of $E[\cdot|\mathcal{G}]$ on L^2. It therefore increases a.s. to a limit which we denote by $E[X|\mathcal{G}]$. Notice that $E[E[X \wedge n|\mathcal{G}]] = E[X \wedge n]$ so that by the monotone convergence theorem, $E[E[X|\mathcal{G}]] = E[X]$. In particular, if X is integrable, then so is $E[X|\mathcal{G}]$.

Proof of existence in Theorem 1.2.1. *Existence.* Let $X \in L^1$, and write $X = X^+ - X^-$ (where $X^+ = X \vee 0$, and $X^- = (-X) \vee 0$). Then X^+, X^- are non-negative integrable random variables, so $E[X^+|\mathcal{G}]$ and $E[X^-|\mathcal{G}]$ are finite a.s. and we may define

$$E[X|\mathcal{G}] = E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}].$$

Now, let $B \in \mathcal{G}$. Then $E[(X^+ \wedge n)\mathbb{1}_B] = E[E[X^+ \wedge n|\mathcal{G}]\mathbb{1}_B]$ by definition. The monotone convergence theorem allows to pass to the limit (all integrated random variables are non-negative), and we obtain $E[X^+ \mathbb{1}_B] = E[E[X^+|\mathcal{G}]\mathbb{1}_B]$. The same is easily true for X^-, and by subtracting we see that $E[X|\mathcal{B}]$ indeed satisfies the characteristic properties 1., 2.

The following properties are immediate consequences of the previous theorem and its proof.

Proposition 1.2.1 *Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra and $X, Y \in L^1(\Omega, \mathcal{F}, P)$. Then*

1. $E[E[X|\mathcal{G}]] = E[X]$
2. *If X is $\mathcal{G}$-measurable, then $E[X|\mathcal{G}] = X$.*

3. *If X is independent of $\mathcal{G}$, then $E[X|\mathcal{G}] = E[X]$.*

4. *If $a, b \in \mathbb{R}$ then $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$ (linearity).*

5. *If $X \geq 0$ then $E[X|\mathcal{G}] \geq 0$ (positiveness).*

6. *$|E[X|\mathcal{G}]| \leq E[|X|\,|\mathcal{G}]$, so that $E[|E[X|\mathcal{G}]|] \leq E[|X|]$.*

Important remark. Notice that all statements concerning conditional expectation are about L^1 variables, which are only defined up to a subset of of zero probability, and hence are a.s. statements. This is of crucial importance and reminds the fact that we encountered before, that $E[X|Y = y]$ can be assigned an arbitrary value whenever $P(Y = y) = 0$.

1.2.3 Non-negative case

In the course of proving the last theorem, we actually built an object $E[X|\mathcal{G}]$ as the a.s. increasing limit of $E[X \wedge n|\mathcal{G}]$ for any non-negative random variable X, not necessarily integrable. This random variable enjoys similar properties as the L^1 case, and we state them similarly as in Theorem 1.2.1.

Theorem 1.2.2 *Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra, and $X \geq 0$ a non-negative random variable. Then there exists a random variable $X' \geq 0$ such that*

1. *X' is $\mathcal{G}$-measurable, and*

2. *for every non-negative $\mathcal{G}$-measurable random variable Z, $E[ZX'] = E[ZX]$.*

Moreover, if X'' is another such r.v., $X' = X''$ a.s. We denote by $E[X|\mathcal{G}]$ the class of X' up to a.s. equality.

Proof. Any r.v. in the class of $E[X|\mathcal{G}] = \limsup_n E[X \wedge n|\mathcal{G}]$ trivially satisfies 1. It also satisfies 2. since if Z is a positive $\mathcal{G}$-measurable random variable, we have, by passing to the (increasing) limit in

$$E[(X \wedge n)(Z \wedge n)] = E[E[X \wedge n|\mathcal{G}]Z \wedge n],$$

that $E[XZ] = E[E[X|\mathcal{G}]Z]$.

Uniqueness. If X', X'' are non-negative and satisfy the properties 1. & 2., for any $a < b \in \mathbb{Q}_+$, by letting $B = \{X' \leq a < b \leq X''\} \in \mathcal{G}$, we obtain

$$bP(B) \leq E[X''\mathbb{1}_B] = E[X\mathbb{1}_B] = E[X'\mathbb{1}_B] \leq aP(B),$$

which entails $P(B) = 0$, so that $P(X' < X'') = 0$ by taking the countable union over $a < b \in \mathbb{Q}_+$. Similarly, $P(X' > X'') = 0$. □

The reader is invited to formulate and prove analogs of the properties of Proposition 1.2.1 for positive variables, and in particular, that if $0 \leq X \leq Y$ then $0 \leq E[X|\mathcal{G}] \leq E[Y|\mathcal{G}]$ a.s. The conditional expectation enjoys the following properties, which match those of the classical expectation.

Proposition 1.2.2 *Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra.*

1. *If $(X_n, n \geq 0)$ is an increasing sequence of non-negative random variables with limit X, then (conditional monotone convergence theorem)*

$$E[X_n|\mathcal{G}] \underset{n\to\infty}{\nearrow} E[X|\mathcal{G}] \qquad a.s.$$

2. *If $(X_n, n \geq 0)$ is a sequence of non-negative random variables, then (conditional Fatou theorem)*

$$E[\liminf_{n\to\infty} X_n|\mathcal{G}] \leq \liminf_{n\to\infty} E[X_n|\mathcal{G}] \qquad a.s.$$

3. *If $(X_n, n \geq 0)$ is a sequence of random variables a.s. converging to X, and if there exists $Y \in L^1(\Omega, \mathcal{F}, P)$ such that $\sup_n |X_n| \leq Y$ a.s., then (conditional dominated convergence theorem)*

$$\lim_{n\to\infty} E[X_n|\mathcal{G}] = E[X|\mathcal{G}]\,, \qquad a.s.\ and\ in\ L^1.$$

4. *If $\varphi : \mathbb{R} \to (-\infty, \infty]$ is a convex function and $X \in L^1(\Omega, \mathcal{F}, P)$, and either φ is non-negative or $\varphi(X) \in L^1(\Omega, \mathcal{F}, P)$, then (conditional Jensen inequality)*

$$E[\varphi(X)|\mathcal{G}] \geq \varphi(E[X|\mathcal{G}]) \qquad a.s.$$

5. *If $1 \leq p < \infty$ and $X \in L^p(\Omega, \mathcal{F}, P)$,*

$$\|E[X|\mathcal{G}]\|_p \leq \|X\|_p.$$

   In particular, the linear operator $X \mapsto E[X|\mathcal{G}]$ from $L^p(\Omega, \mathcal{F}, P)$ to $L^p(\Omega, \mathcal{G}, P)$ is continuous.

Proof. 1. Let X' be the increasing limit of $E[X_n|\mathcal{G}]$. Let Z be a positive $\mathcal{G}$-measurable random variable, then $E[ZE[X_n|\mathcal{G}]] = E[ZX_n]$, which by taking an increasing limit gives $E[ZX'] = E[ZX]$, so $X' = E[X|\mathcal{G}]$.

2. We have $E[\inf_{k\geq n} X_k|\mathcal{G}] \leq \inf_{k\geq n} E[X_k|\mathcal{G}]$ for every n by monotonicity of the conditional expectation, and the result is obtained by passing to the limit and using 1.

3. Applying 2. to the nonnegative random variables $Z - X_n, Z + X_n$, we get that $E[Z-X|\mathcal{G}] \leq E[Z|\mathcal{G}] - \limsup E[X_n|\mathcal{G}]$ and that $E[Z+X|\mathcal{G}] \leq E[Z|\mathcal{G}] + \liminf E[X_n|\mathcal{G}]$, giving the a.s. result. The L^1 result is a consequence of the dominated convergence theorem, since $|E[X_n|\mathcal{G}]| \leq E[|X_n|\,|\mathcal{G}] \leq |Z|$ a.s.

4. A convex function φ is the superior envelope of its affine minorants, i.e.

$$\varphi(x) = \sup_{a,b\in\mathbb{R}:\forall y, ay+b\leq\varphi(y)} ax + b = \sup_{a,b\in\mathbb{Q}:\forall y, ay+b\leq\varphi(y)} ax + b.$$

The result is then a consequence of linearity of the conditional expectation and the fact that $\mathbb{Q}$ is countable (this last fact is necessary because of the fact that conditional expectation is defined only a.s.).

5. One deduces from 4. and the previous proposition that $\|E[X|\mathcal{G}]\|_p^p = E[|E[X|\mathcal{G}]|^p] \leq E[E[|X|^p\,|\mathcal{G}]] = E[|X|^p] = \|X\|_p^p$, if $1 \leq p < \infty$ and $X \in L^p(\Omega, \mathcal{F}, P)$. Thus

$$\|E[X|\mathcal{G}]\|_p \leq \|X\|_p.$$

□

1.3 Specific properties of conditional expectation

The "information contained in $\mathcal{G}$" can be factorized out the conditional expectation:

Proposition 1.3.1 *Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra, and let X, Y be real random variables such that either X, Y are non-negative or $X, XY \in L^1(\Omega, \mathcal{F}, P)$. Then, is Y is $\mathcal{G}$-measurable, we have*

$$E[YX|\mathcal{G}] = YE[X|\mathcal{G}].$$

Proof. Let Z be a non-negative $\mathcal{G}$-measurable random variable, then, if X, Y are non-negative, $E[ZYX] = E[ZYE[X|\mathcal{G}]]$ since ZY is non-negative, and the result follows by uniqueness. If X, XY are integrable, the same result follows by letting $X = X^+ - X^-, Y = Y^+ - Y^-$. □

One has the *Tower property* (restricting the information)

Proposition 1.3.2 *Let $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$ be σ-algebras. Then for every random variable X which is positive or integrable,*

$$E[E[X|\mathcal{G}_2]|\mathcal{G}_1] = E[X|\mathcal{G}_1].$$

Proof. For a positive bounded $\mathcal{G}_1$-measurable Z, Z is $\mathcal{G}_2$-measurable as well, so that $E[ZE[E[X|\mathcal{G}_2]|\mathcal{G}_1]] = E[ZE[X|\mathcal{G}_2]] = E[ZX] = E[ZE[X|\mathcal{G}_1]]$, hence the result. □

Proposition 1.3.3 *Let $\mathcal{G}_1, \mathcal{G}_2$ be two sub-σ-algebras of $\mathcal{F}$, and let X be a positive or integrable random variable. Then, if $\mathcal{G}_2$ is independent of $\sigma(X, \mathcal{G}_1)$, $E[X|\mathcal{G}_1 \vee \mathcal{G}_2] = E[X|\mathcal{G}_1]$.*

Proof. Let $A \in \mathcal{G}_1, B \in \mathcal{G}_2$, then

$$\begin{aligned} E[\mathbb{1}_{A\cap B}E[X|\mathcal{G}_1 \vee \mathcal{G}_2]] &= E[\mathbb{1}_A \mathbb{1}_B X] = E[\mathbb{1}_B E[X\mathbb{1}_A|\mathcal{G}_2]] = P(B)E[X\mathbb{1}_A] \\ &= P(B)E[\mathbb{1}_A E[X|\mathcal{G}_1]] = E[\mathbb{1}_{A\cap B}E[X|\mathcal{G}_1]], \end{aligned}$$

where we have used the independence property at the third and last steps. The proof is then done by the monotone class theorem. □

Proposition 1.3.4 *Let X, Y be random variables and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$ such that Y is $\mathcal{G}$-measurable and X is independent of $\mathcal{G}$. Then for any non-negative measurable function f,*

$$E[f(X,Y)|\mathcal{G}] = \int P(X \in \mathrm{d}x) f(x, Y),$$

where $P(X \in \mathrm{d}x)$ is the law of X.

Proof. For any non-negative $\mathcal{G}$-measurable random variable Z, we have that X is independent of (Y, Z), so that the law $P((X, Y, Z) \in \mathrm{d}x\mathrm{d}y\mathrm{d}z)$ is equal to the product $P(X \in \mathrm{d}x)P((Y, Z) \in \mathrm{d}y\mathrm{d}z)$ of the law of X by the law of (Y, Z). Hence,

$$\begin{aligned} E[Zf(X,Y)] &= \int zf(x,y)P(X \in \mathrm{d}x)P((Y,Z) \in \mathrm{d}y\mathrm{d}z) \\ &= \int P(X \in \mathrm{d}x)E[Zf(x,Y)] = E\left[Z\int P(X \in \mathrm{d}x)f(x,Y)\right], \end{aligned}$$

where we used Fubini's theorem in two places. This shows the result.

1.4 Computing a conditional expectation

We give two concrete and important examples of computation of conditional expectations.

1.4.1 Conditional density functions

Suppose X, Y have values in $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, and that the law of (X, Y) has a density: $P((X,Y) \in \mathrm{d}x\mathrm{d}y) = f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$. Let $f_Y(y) = \int_{x\in\mathbb{R}^m} f_{X,Y}(x,y)\mathrm{d}x, y \in \mathbb{R}^n$ be the density of Y. Then for every non-negative measurable $h : \mathbb{R}^m \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}$, we have

$$\begin{aligned} E[h(X)g(Y)] &= \int_{\mathbb{R}^m\times\mathbb{R}^n} h(x)g(y)f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_{\mathbb{R}^n} g(y)f_Y(y)\mathrm{d}y \int_{\mathbb{R}^m} h(x)\frac{f_{X,Y}(x,y)}{f_Y(y)}\mathbb{1}_{\{f_Y(y)>0\}}\mathrm{d}x \\ &= E[\varphi(Y)g(Y)], \end{aligned}$$

so $E[h(X)|Y] = \varphi(Y)$, where

$$\varphi(y) = \frac{1}{f_Y(y)} \int_{\mathbb{R}^m} h(x)f_{X,Y}(x,y)\mathrm{d}x \qquad \text{if } f_Y(y) > 0,$$

and 0 else. We interpret this result by saying that

$$E[h(X)|Y] = \int_{\mathbb{R}^m} h(x)\nu(Y,\mathrm{d}x),$$

where $\nu(y,\mathrm{d}x) = f_Y(y)^{-1}f_{X,Y}(x,y)\mathbb{1}_{\{f_Y(y)>0\}}\mathrm{d}x = f_{X|Y}(x|y)\mathrm{d}x$. The measure $\nu(y,\mathrm{d}x)$ is called *conditional distribution* given $Y = y$, and $f_{X|Y}(x|y)$ is the *conditional density function* of X given $Y = y$. Notice this function of x, y is defined only up to a zero-measure set.

1.4.2 The Gaussian case

Let (X, Y) be a Gaussian vector in $\mathbb{R}^2$. Take $X' = aY + b$ with a, b such that $\mathrm{Cov}\,(X, Y) = a\mathrm{Var}\,Y$ and $aE[Y] + b = E[X]$. In this case, $\mathrm{Cov}\,(Y, X - X') = 0$, hence $X - X'$ is independent of $\sigma(Y)$ by properties of Gaussian vectors. Moreover, $X - X'$ is centered so for every $B \in \sigma(Y)$, one has $E[\mathbb{1}_B X] = E[\mathbb{1}_B X']$, hence $X' = E[X|Y]$.

Chapter 2

Discrete-time martingales

Before we entirely focus on discrete-time martingales, we start with a general discussion on stochastic processes, which includes both discrete and continuous-time processes.

2.1 Basic notions

2.1.1 Stochastic processes, filtrations

Let $(\Omega, \mathcal{F}, P)$ be a probability space. For a measurable states space $(E, \mathcal{E})$ and a subset $I \subset \mathbb{R}$ of "times", or "epochs", an E-valued *stochastic process* indexed by I is a collection $(X_t, t \in I)$ of random variables. Most of the processes we will consider take values in $\mathbb{R}$, $\mathbb{R}^d$, or $\mathbb{C}$, being endowed with their Borel σ-algebras.

A *filtration* is a collection $(\mathcal{F}_t, t \in I)$ of sub-σ-algebras of $\mathcal{F}$ which is increasing ($s \leq t \implies \mathcal{F}_s \subseteq \mathcal{F}_t$). Once a filtration is given, we call $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}, P)$ a *filtered probability space.* A process $(X_t, t \in I)$ is *adapted* to the filtration $(\mathcal{F}_t, t \in I)$ if X_t is $\mathcal{F}_t$-measurable for every t.

The intuitive idea is that $\mathcal{F}_t$ is the quantity of information available up to time t (present). To give an informal example, if we are interested in the evolution of the stock market, we can take $\mathcal{F}_t$ as the past history of the stocks prices (or only some of them) up to time t.

We will let $\mathcal{F}_\infty = \bigvee_{t\in I} \mathcal{F}_t \subseteq \mathcal{F}$ be the information at the end of times.

Example. For every process $(X_t, t \in I)$, one associates the *natural filtration*

$$\mathcal{F}_t^X = \sigma(\{X_s, s \leq t\}), \qquad t \in I.$$

Every process is adapted to its natural filtration, and $\mathcal{F}^X$ is the smallest filtration X is adapted to: $\mathcal{F}_t^X$ contains all the measurable events depending on $(X_s, s \leq t)$.

Last, a real-valued process $(X_t, t \in I)$ is said to be *integrable* if $E[|X_t|] < \infty$ for all $t \in I$.

2.1.2 Martingales

Definition 2.1.1 *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}, P)$ be a filtered probability space. An $\mathbb{R}$-valued adapted integrable process $(X_t, t \in I)$ is:*

- *a* martingale *if for every $s \leq t$, $E[X_t|\mathcal{F}_s] = X_s$.*
- *a* supermartingale *if for every $s \leq t$, $E[X_t|\mathcal{F}_s] \leq X_s$.*
- *a* submartingale *if for every $s \leq t$, $E[X_t|\mathcal{F}_s] \geq X_s$.*

Notice that a (super, sub)martingale remains a martingale with respect to its natural filtration, by the tower property of conditional expectation.

2.1.3 Doob's stopping times

Definition 2.1.2 *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}, P)$ be a filtered probability space. A* stopping time *(with respect to this space) is a random variable $T : \Omega \to I \sqcup \{\infty\}$ such that $\{T \leq t\} \in \mathcal{F}_t$ for every $t \in I$.*

For example, constant times are (trivial) stopping times. If $I = \mathbb{Z}_+$, the random variable $n\mathbb{1}_A + \infty\mathbb{1}_{A^c}$ is a stopping time if $A \in \mathcal{F}_n$ (with the convention $0 \cdot \infty = 0$). The intuitive idea behind this definition is that T is a time when a decision can be taken (given the information we have). For example, for a meteorologist having the weather information up to the present time, the "first day of 2006 when the temperature is above 23°C is a stopping time, but not the "last day of 2006 when the temperature is above 23°C".

Example. If $I \subset \mathbb{Z}_+$, the definition can be replaced by $\{T = n\} \in \mathcal{F}_n$ for all $n \in I$. When I is a subset of the integers, we will denote the time by letters n, m, k rather than t, s, r (so $n \geq 0$ means $n \in \mathbb{Z}_+$). Particularly important instances of stopping times in this case are the *first entrance times*. Let $(X_n, n \geq 0)$ be an adapted process and let $A \in \mathcal{E}$. The first entrance time in A is

$$T_A = \inf\{n \in \mathbb{Z}_+ : X_n \in A\} \in \mathbb{Z}_+ \sqcup \{\infty\}.$$

It is a stopping time since

$$\{T_A \leq n\} = \bigcup_{0\leq m\leq n} X_m^{-1}(A).$$

To the contrary, the *last exit time* before some fixed N,

$$L_A = \sup\{n \in \{0, 1, \dots, N\} : X_n \in A\} \in \mathbb{Z}_+ \sqcup \{\infty\},$$

is **in general** not a stopping time.

As an immediate consequence of the definition, one gets:

Proposition 2.1.1 *Let $S, T, (T_n, n \in \mathbb{N})$ be stopping times (with respect to some filtered probability space). Then $S \wedge T, S \vee T, \inf_n T_n, \sup_n T_n, \liminf_n T_n, \limsup_n T_n$ are stopping times.*

Definition 2.1.3 *Let T be a stopping time with respect to some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}, P)$. We define $\mathcal{F}_T$, the σ-algebra of events before time T*

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq t\} \in \mathcal{F}_t\}.$$

The reader is invited to check that it defines indeed a σ-algebra, which is interpreted as the events that are measurable with respect to the information available at time T: "4 days before the first day (T) in 2005 when the temperature is above 23°C, the temperature was below 10°C" is in $\mathcal{F}_T$. If S, T are stopping times, one checks that

$$S \leq T \implies \mathcal{F}_S \subseteq \mathcal{F}_T. \tag{2.1}$$

Now **suppose that I is countable**. If $(X_t, t \in I)$ is adapted, and T a stopping time, we let $X_T \mathbb{1}_{\{T<\infty\}} = X_{T(\omega)}(\omega)$ if $T(\omega) < \infty$, and 0 else. It is a random variable as the composition of $(\omega, t) \mapsto X_t(\omega)$ and $\omega \mapsto (\omega, T(\omega))$, which are measurable (why?). We also let $X^T = (X_{T\wedge t}, t \in I)$, and we call it the process X *stopped at* T.

Proposition 2.1.2 *Under these hypotheses,*

1. *$X_T \mathbb{1}_{\{T<\infty\}}$ is $\mathcal{F}_T$-measurable,*

2. *the process X^T is adapted,*

3. *if moreover $I = \mathbb{Z}_+$ and X is integrable, then X^T is integrable.*

Proof. 1. Let $A \in \mathcal{E}$. Then $\{X_T \in A\} \cap \{T \leq t\} = \bigcup_{s\in I, s\leq t} \{X_s \in A\} \cap \{T = s\}$. Then notice $\{T = s\} = \{T \leq s\} \setminus \bigcup_{u<s} \{T \leq u\} \in \mathcal{F}_s$.

2. For every $t \in I$, $X_{T\wedge t}$ is $\mathcal{F}_{T\wedge t}$-measurable, hence $\mathcal{F}_t$ measurable since $T \wedge t \leq t$, by (2.1).

3. If $I = \mathbb{Z}_+$ and X is integrable, $E[|X_n^T|] = \sum_{m<n} E[|X_m|\mathbb{1}_{\{T=m\}}] + E[|X_n|\mathbb{1}_{\{T\geq n\}}] \leq n \sup_{0\leq m\leq n} E[|X_m|]$. □

From now on until the end of the section (except in the paragraph on backwards martingales), we will suppose that $E = \mathbb{R}$ and $I = \mathbb{Z}_+$ (discrete-time processes).

2.2 Discrete-time martingales: optional stopping

We consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n), P)$. All the above terminology (stopping times, adapted processes and so on) will be with respect to this space.

We first introduce the so-called 'martingale transform', which is sometimes called the 'discrete stochastic integral' with respect to a (super, sub)martingale X. We say that a process $(C_n, n \geq 1)$ is *previsible* if C_n is $\mathcal{F}_{n-1}$-measurable for every $n \geq 1$. A previsible process can be interpreted as a strategy: one bets at time n only with all the accumulated knowledge up to time $n - 1$.

If $(X_n, n \geq 0)$ is adapted and $(C_n, n \geq 1)$ is previsible, we define an adapted process $C \cdot X$ by

$$(C \cdot X)_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1}).$$

We can interpret this new process as follows: if X_n is a certain amount of money at time n and if C_n is the bet of a player at time n then $(C \cdot X)_n$ is the total winning of the player at time n.

Proposition 2.2.1 *In this setting, if X is a martingale, and C is bounded, then $C \cdot X$ is a martingale. If X is a supermartingale (resp. submartingale) and $C_n \geq 0$ for every $n \geq 1$, then $C \cdot X$ is a supermartingale (resp. submartingale).*

Proof. Suppose X is a martingale. Since C is bounded, the process $C \cdot X$ is trivially integrable. Since C_{n+1} is $\mathcal{F}_n$-measurable,

$$E[(C \cdot X)_{n+1} - (C \cdot X)_n | \mathcal{F}_n] = E[C_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n] = C_{n+1}E[X_{n+1} - X_n|\mathcal{F}_n] = 0.$$

The (super-, sub-)martingale cases are similar. □

Theorem 2.2.1 (Optional stopping) *Let $(X_n, n \geq 0)$ be a martingale (resp. super-, sub-martingale).*

(i) *If T is a stopping time, then Then X^T is also a martingale (resp. super-, sub-martingale).*

(ii) *If $S \leq T$ are* **bounded** *stopping times, then $E[X_T|\mathcal{F}_S] = X_S$ (resp. $E[X_T|\mathcal{F}_S] \leq X_S$, $E[X_T|\mathcal{F}_S] \geq X_S$).*

(iii) *If $S \leq T$ are* **bounded** *stopping times, then $E[X_T] = E[X_S]$ (resp. $E[X_T] \leq E[X_S]$, $E[X_T] \geq E[X_S]$).*

Proof. (i) Let $C_n = \mathbb{1}_{\{n \leq T\}}$, then C is a previsible non-negative bounded process, and it is immediate that $C \cdot X = X^T$. The first result follows from Proposition 2.2.1.

(ii) If now S, T are bounded stopping times with $S \leq T$, and $A \in \mathcal{F}_S$, we define $C_n = \mathbb{1}_A \mathbb{1}_{\{S<n\leq T\}}$. Then C is a nonnegative bounded previsible process, since $A \cap \{S < n\} = A \cap \{S \leq n-1\} \in \mathcal{F}_{n-1}$ and $\{n \leq T\} = \{n-1 < T\} \in \mathcal{F}_{n-1}$. Morevoer, X_S, X_T are integrable since S, T are bounded, and $(C \cdot X)_K = \mathbb{1}_A(X_T - X_S)$ as soon as $K \geq T$ a.s. Since $C \cdot X$ is a martingale, $E[(C \cdot X)_K] = E[(C \cdot X)_0] = 0$. Taking expectations entails that $E[X_T|\mathcal{F}_S] = X_S$.

(iii) Follows by taking expectations in (ii). □

Notice that the last two statement is not true in general. For example, if $(Y_n, n \geq 0)$ are independent random variables which take values ± 1 with probability $1/2$, then $X_n = \sum_{1\leq i\leq n} Y_i$ is a martingale. If $T = \inf\{n \geq 0 : X_n = 1\}$ then it is classical that $T < \infty$ a.s., but of course $E[X_T] = 1 > 0 = E[X_0]$. However, for non-negative supermartingales, Fatou's lemma entails:

Proposition 2.2.2 *Suppose X is a non-negative supermartingale. Then for any stopping time which is a.s. finite, we have $E[X_T] \leq E[X_0]$.*

Beware that this $\leq$ sign should *not* in general be turned into a $=$ sign, even if X is a martingale! The very same proposition is actually true without the assumption that $P(T<\infty)=1$, by the martingale convergence theorem 2.3.1 below.

2.3 Discrete-time martingales: the convergence theorem

The martingale convergence theorem is the most important result in this chapter.

Theorem 2.3.1 (Martingale convergence theorem) *If X is a supermartingale which is bounded in $L^1(\Omega,\mathcal{F},P)$, i.e. such that $\sup_n E[|X_n|]<\infty$, then X_n converges a.s. towards an a.s. finite limit X_∞.*

An easy and important corollary for that is

Corollary 2.3.1 *A non-negative supermartingale converges a.s. towards an a.s. finite limit.*

Indeed, for a non-negative supermartingale, $E[|X_n|]=E[X_n]\leq E[X_0]<\infty$.

The proof of Theorem 2.3.1 relies on an estimation of the number of *upcrossings* of a submartingale between to levels $a<b$. If $(x_n,n\geq 0)$ is a real sequence, and $a<b$ are two real numbers, we define two integer-valued sequences $S_k(x),T_k(x),k\geq 1$ recursively as follows. Let $T_0(x)=0$ and for $k\geq 0$, let

$$S_{k+1}(x)=\inf\{n\geq T_k:x_n<a\}\,,\qquad T_{k+1}(x)=\inf\{n\geq S_{k+1}(x):x_m>b\},$$

with the usual convention $\inf\varnothing=\infty$. The number $N_n([a,b],x)=\sup\{k>0:T_k(x)\leq n\}$ is the number of upcrossings of x between a and b before time n, which increases as $n\to\infty$ to the total number of upcrossings $N([a,b],x)=\sup\{k>0:T_k(x)<\infty\}$. The key is the simple following analytic lemma:

Lemma 2.3.1 *A real sequence x converges (in $\overline{\mathbb{R}}$) if and only if $N([a,b],x)<\infty$ for every rationals $a<b$.*

Proof. If there exist $a<b$ rationals such that $N([a,b],x)=\infty$, then $\liminf_n x_n\leq a<b\leq\limsup_n x_n$ so that x does not converge. If x does not converge, then $\liminf_n x_n<\limsup_n x_n$, so by taking two rationals $a<b$ in between, we get the converse statement. □

Theorem 2.3.2 (Doob's upcrossing lemma) *Let X be a supermartingale, and $a<b$ two reals. Then for every $n\geq 0$,*

$$(b-a)E[N_n([a,b],X)]\leq E[(X_n-a)^-].$$

Proof. It is immediate by induction that $S_k = S_k(X), T_k = T_k(X)$ defined as above are stopping times. Define a previsible process C, taking only 0 or 1 values, by

$$C_n = \sum_{k\geq 1} \mathbb{1}_{\{S_k < n \leq T_k\}}.$$

It is indeed previsible since $\{S_k < n \leq T_k\} = \{S_k \leq n-1\} \cap \{T_k \leq n-1\}^c \in \mathcal{F}_{n-1}$. Now, letting $N_n = N([a,b], X)$, we have

$$\begin{aligned}(C \cdot X)_n &= \sum_{i=1}^{N_n} (X_{T_i} - X_{S_i}) + (X_n - X_{S_{N_n+1}}) \mathbb{1}_{\{S_{N_n+1} \leq n\}} \\ &\geq (b-a)N_n + (X_n - a)\mathbb{1}_{\{X_n \leq a\}} \geq (b-a)N_n - (X_n - a)^-.\end{aligned}$$

Since C is a non-negative bounded previsible process, $C \cdot X$ is a supermartingale so finally

$$(b-a)E[N_n] - E[(X_n - a)^-] \leq E[(C \cdot X)_n] \leq 0,$$

hence the result. □

Proof of Theorem 2.3.1. Since $(x+y)^- \leq |x| + |y|$, we get from Theorem 2.3.2 that $E[N_n] \leq (b-a)^{-1} E[|X_n| + a]$, and since N_n increases to $N = N([a,b], X)$ we get by monotone convergence $E[N] \leq (b-a)^{-1}(\sup_n E[|X_n|] + a)$. In particular, we get $N([a,b], X) < \infty$ a.s. for every $a < b \in \mathbb{Q}$, so

$$P\left(\bigcap_{a<b\in\mathbb{Q}} \{N([a,b], X) < \infty\} \right) = 1.$$

Hence the a.s. convergence to some X_∞, possibly infinite.

Now Fatou's lemma gives $E[|X_\infty|] \leq \liminf_n E[|X_n|] < \infty$ by hypothesis, hence $|X_\infty| < \infty$ a.s. □

Exercise. In fact, from Theorem 2.3.2 it is clearly enough that $\sup_n E[X_n^-] < \infty$ is sufficient, prove that this actually implies boundedness is L^1 (provided X is a supermartingale, of course).

2.4 Doob's inequalities and L^p convergence, $p > 1$

2.4.1 A maximal inequality

Proposition 2.4.1 *Let X be a sub-martingale. Then letting $\widetilde{X}_n = \sup_{0\leq k\leq n} X_k$, for every $c > 0$, and $n \geq 0$,*

$$cP(\widetilde{X}_n \geq c) \leq E[X_n \mathbb{1}_{\{X_n^* \geq c\}}] \leq E[X_n^+].$$

Proof. Letting $T = \inf\{k \geq 0 : X_k \geq c\}$, we obtain by optional stopping that

$$E[X_n] \geq E[X_n^T] = E[X_n \mathbb{1}_{\{T>n\}}] + E[X_T \mathbb{1}_{\{T\leq n\}}] \geq E[X_n \mathbb{1}_{\{T>n\}}] + cP(T \leq n).$$

Since $\{T \leq n\} = \{\widetilde{X}_n \geq c\}$, the conclusion follows. □

Theorem 2.4.1 (Doob's L^p inequality) *Let $p > 1$, and X be a martingale, then letting $X_n^* = \sup_{0\leq k\leq n}|X_k|$, we have*

$$\|X_n^*\|_p \leq \frac{p}{p-1}\|X_n\|_p.$$

Proof. Since $x \mapsto |x|$ in convex, the process $(|X_n|, n \geq 0)$ is a non-negative submartingale. Applying Proposition 2.4.1 and Hölder's inequality shows that

$$\begin{aligned} E[(X_n^*)^p] &= \int_0^\infty \mathrm{d}x\, px^{p-1}P(X_n^* \geq x) \\ &\leq \int_0^\infty \mathrm{d}x\, px^{p-2}E[|X_n|\mathbb{1}_{\{X_n^*\geq x\}}] \\ &= pE\left[|X_n|\int_0^{X_n^*}\mathrm{d}x\, x^{p-2}\right] \\ &= \frac{p}{p-1}E[|X_n|(X_n^*)^{p-1}] \leq \frac{p}{p-1}\|X_n\|_p\|X_n^*\|_p^{p-1}, \end{aligned}$$

which yields the result. □

Theorem 2.4.2 *Let X be a martingale and $p > 1$, then the following statements are equivalent:*

1. *X is bounded in $L^p(\Omega, \mathcal{F}, P)$: $\sup_{n\geq 0}\|X_n\|_p < \infty$*
2. *X converges a.s. and in L^p to a random variable X_∞*
3. *There exists some $Z \in L^p(\Omega, \mathcal{F}, P)$ such that*

$$X_n = E[Z|\mathcal{F}_n].$$

Proof. 1. $\Longrightarrow$ 2. Suppose X is bounded in L^p, then in particular, it is bounded in L^1 so it converges a.s. to some finite X_∞ by Theorem 2.3.1. Moreover, $X_\infty \in L^p$ by an easy application of Fatou's theorem. Next, Doob's inequality $\|X_n^*\|_p \leq C\|X_n\|_p < C' < \infty$ entails $\|X_\infty^*\|_p < \infty$ by monotone convergence, where $X_\infty^* = \sup_{n\geq 0}|X_n|$ is the monotone limit of X_n^*. Since $X_\infty^* \geq \sup_{n\in\overline{\mathbb{N}}}|X_n|$, $|X_n - X_\infty| \leq 2X_\infty^* \in L^p$ and dominated convergence entails that X_n converges to X_∞ in L^p.

2. $\Longrightarrow$ 3. Since conditional expectation is continuous as a linear operator on L^p spaces (Proposition 1.2.2), if $X_n \to X_\infty$ in L^p we have for $n \leq m$, $X_n = E[X_m|\mathcal{F}_n] \to_{m\to\infty} E[X_\infty|\mathcal{F}_n]$.

3. $\Longrightarrow$ 1. This is immediate by the conditional Jensen inequality. □

A martingale which has the form in 3. is said to be *closed* (in L^p). Notice that in this case, $X_\infty = E[Z|\mathcal{F}_\infty]$, where $\mathcal{F}_\infty = \bigvee_{n\geq 0}\mathcal{F}_n$. Indeed, $\bigcup_{n\geq 0}\mathcal{F}_n$ is a π-system that spans $\mathcal{F}_\infty$, and moreover if $B \in \mathcal{F}_N$ say is an element of this π-system, $E[\mathbb{1}_B Z] = E[\mathbb{1}_B E[Z|\mathcal{F}_\infty]] = E[\mathbb{1}_B X_N] \to E[\mathbb{1}_B X_\infty]$ as $N \to \infty$. Since $X_\infty = \limsup_n X_n$ is $\mathcal{F}_\infty$-measurable, this gives the result.

Therefore, for $p > 1$, the map $Z \in L^p(\Omega, \mathcal{F}_\infty, P) \mapsto (E[Z|\mathcal{F}_n], n \geq 0)$ is a bijection between $L^p(\Omega, \mathcal{F}_\infty, P)$ and the set of martingales that are bounded in L^p.

2.5 Uniform integrability and convergence in L^1

The case of L^1 convergence is a little different from L^p for $p > 1$, as one needs to suppose uniform integrability rather than a mere boundedness in L^1. Notice that uniform integrability follows from boundedness in L^p.

Theorem 2.5.1 *Let X be a martingale. The following statements are equivalent:*

1. *$(X_n, n \geq 0)$ is uniformly integrable*
2. *X_n converges a.s. and in $L^1(\Omega, \mathcal{F}, P)$ to a limit X_∞*
3. *There exists $Z \in L^1(\Omega, \mathcal{F}, P)$ so that $X_n = E[Z|\mathcal{F}_n], n \geq 0$.*

Proof. 1. $\implies$ 2. Suppose X is uniformly integrable, then it is bounded in L^1 so by Theorem 2.3.1 it converges a.s. By properties of uniform integrability, it then converges in L^1.

2. $\implies$ 3. This follows the same proof as above: $X_\infty = Z$ is a suitable choice.

3. $\implies$ 1. This is a straightforward consequence of the fact that

$$\{E[X|\mathcal{G}] : \mathcal{G} \text{ is a sub-}\sigma\text{-algebra of } \mathcal{F}\}$$

is U.I., see the example sheet 1. □

As above, we then have $E[Z|\mathcal{F}_\infty] = X_\infty$, and this theorem says that there is a one-to-one correspondence between U.I. martingales and $L^1(\Omega, \mathcal{F}_\infty, P)$.

Exercise. Show that if X is a U.I. supermartingale (resp. submartingale), then X_n converges a.s. and in L^1 to a limit X_∞, so that $E[X_\infty|\mathcal{F}_n] \leq X_n$ (resp. $\geq$) for every n.

2.6 Optional stopping in the case of U.I. martingales

We give an improved version of the optional stopping theorem, in which the boundedness condition on the stopping time is lifted, and replaced by a uniform integrability condition on the martingale. Since U.I. martingales have a well defined limit X_∞, we unambiguously let $X_T = X_T \mathbb{1}_{\{T<\infty\}} + X_\infty \mathbb{1}_{\{T=\infty\}}$ for any stopping time T.

Theorem 2.6.1 *Let X be a U.I. martingale, and S, T be two stopping times with $S \leq T$. Then $E[X_T|\mathcal{F}_S] = X_S$.*

Proof. We check that $X_T \in L^1$, indeed, since $|X_n| \leq E[|X_\infty| \,|\mathcal{F}_n]$,

$$E[|X_T|] = \sum_{n=0}^{\infty} E[|X_n| \mathbb{1}_{\{T=n\}}] + E[|X_\infty| \mathbb{1}_{\{T=\infty\}}] \leq \sum_{n \in \overline{\mathbb{N}}} E[|X_\infty| \mathbb{1}_{\{T=n\}}] = E[|X_\infty|].$$

Next, if $B \in \mathcal{F}_T$,

$$E[\mathbb{1}_B X_\infty] = \sum_{n \in \mathbb{Z}_+ \sqcup \{\infty\}} E[\mathbb{1}_B \mathbb{1}_{\{T=n\}} X_\infty] = \sum_{n \in \mathbb{Z}_+ \sqcup \{\infty\}} E[\mathbb{1}_B \mathbb{1}_{\{T=n\}} X_n] = E[\mathbb{1}_B X_T],$$

so that $X_T = E[X_\infty|\mathcal{F}_T]$. Finally, $E[X_T|\mathcal{F}_S] = E[E[X_\infty|\mathcal{F}_T]|\mathcal{F}_S] = X_S$, by the tower property. □

2.7 Backwards martingales

Backwards martingales are martingales whose time-set is $\mathbb{Z}_-$. More precisely, given a filtration $\ldots \subseteq \mathcal{G}_{-2} \subseteq \mathcal{G}_{-1} \subseteq \mathcal{G}_0$, a process $(X_n, n \leq 0)$ is a backward martingale if $E[X_{n+1}|\mathcal{G}_n] = X_n$, as in the usual definition. They are somehow nicer than forward martingales, as they are automatically U.I. since $X_0 \in L^1$, and $E[X_0|\mathcal{G}_n] = X_n$ for every $n \leq 0$. Adapting Doob's upcrossing theorem is a simple exercise: if $N_m([a,b], X)$ is the number of upcrossings of a backwards martingale from a to b between times $-m$ and 0, one has, considering the (forward) supermartingale $(X_{-m+k}, 0 \leq k \leq m)$, that

$$(b-a)E[N_m([a,b],X)] \leq E[(X_0 - a)^-].$$

As $m \to \infty$, $N_m([a,b],X)$ increases to the total number of upcrossings of X from a to b, and this allows to conclude that X_n converges a.s. as $n \to -\infty$ to a $\mathcal{G}_{-\infty}$-measurable random variable $X_{-\infty}$, where $\mathcal{G}_{-\infty} = \bigcap_{n \leq 0} \mathcal{G}_n$. We proved:

Theorem 2.7.1 *Let X be a backwards martingale. Then X_n converges a.s. and in L^1 as $n \to -\infty$ to the random variable $X_{-\infty} = E[X_0|\mathcal{G}_{-\infty}]$.*

Moreover, if $X_0 \in L^p$ for some $p > 1$, then X is bounded in L^p and converges in L^p as $n \to -\infty$.

Chapter 3

Examples of applications of discrete-time martingales

3.1 Kolmogorov's $0-1$ law, law of large numbers

Let $(Y_n, n \geq 1)$ be a sequence of independent random variables.

Theorem 3.1.1 (Kolmogorov's $0-1$ law) *The tail σ-algebra $\mathcal{G}_\infty = \bigcap_{n\geq 0} \mathcal{G}_n$, where $\mathcal{G}_n = \sigma\{X_m, m \geq n\}$, is trivial: every $A \in \mathcal{G}_\infty$ has probability 0 or 1.*

Proof. Let $\mathcal{F}_n = \sigma\{Y_1, \ldots, Y_n\}, n \geq 1$. Let $A \in \mathcal{G}_\infty$. Then $E[\mathbb{1}_A|\mathcal{F}_n] = P(A)$ since $\mathcal{F}_n$ is independent of $\mathcal{G}_{n+1}$, hence of $\mathcal{G}_\infty$. Therefore, the martingale convergence theorem gives $E[\mathbb{1}_A|\mathcal{F}_\infty] = \mathbb{1}_A = P(A)$ a.s., since $\mathcal{G}_\infty \subset \mathcal{F}_\infty$. Hence, $P(A) \in \{0, 1\}$. □

Suppose now that the Y_i are real-valued i.i.d. random variables in L^1. Let $S_n = \sum_{k=1}^n Y_k, n \geq 0$ be the associated random walk.

Theorem 3.1.2 (LLN) *A.s. as $n \to \infty$,*

$$\frac{S_n}{n} \underset{n\to\infty}{\longrightarrow} E[Y_1].$$

Proof. Let $\mathcal{H}_n = \sigma\{S_n, S_{n+1}, \ldots\} = \sigma\{S_n, Y_{n+1}, Y_{n+2}, \ldots\}$. We have $E[S_n|\mathcal{H}_{n+1}] = S_{n+1} - E[X_{n+1}|S_{n+1}]$. Now, by symmetry we have $E[X_{n+1}|S_{n+1}] = E[X_k|S_{n+1}]$ for every $1 \leq k \leq n+1$, so that it equals $(n+1)^{-1}E[S_{n+1}|S_{n+1}] = S_{n+1}/(n+1)$. Finally, $E[S_n/n|\mathcal{H}_{n+1}] = S_{n+1}/(n+1)$, so that $(S_{-n}/(-n), n \leq -1)$ is a backwards martingale with respect to its natural filtration. Therefore, S_n/n converges a.s. and in L^1 to a limit which is a.s. constant by Kolmogorov's $0-1$ law, so it must be equal to its mean value: $E[S_1|\mathcal{H}_\infty] = E[S_1] = E[Y_1]$. □

3.2 Branching processes

Let μ be a probability distribution on $\mathbb{Z}_+$, and consider a Markov process $(Z_n, n \geq 0)$ in $\mathbb{Z}_+$ whose step-transitions are determined by the following rule. Given $Z_n = z$, take z

independent random variables $Y_1, \ldots, Y_z$ with law μ, and let Z_{n+1} have the same distribution as $Y_1 + \ldots + Y_z$. In particular, 0 is an absorbing state for this process. This can be interpreted as follows: Z_n is a number of individuals present in a population, and at each time, each individual dies after giving birth to a μ-distributed number of sons, independently of the others. Notice that $E[Z_{n+1}|\mathcal{F}_n] = E[Z_{n+1}|Z_n] = mZ_n$, where $(\mathcal{F}_n, n \geq 0)$ is the natural filtration, and m is the mean of μ, $m = \sum_z z\mu(\{z\})$. Therefore, supposing $m \in (0, \infty)$,

Proposition 3.2.1 *The process $(m^{-n}Z_n, n \geq 0)$ is a non-negative martingale.*

Notice that the fact that the martingale converges a.s. to a finite value immediately implies that when $m < 1$, there exists some n so that $Z_n = 0$, i.e. the population becomes extinct in finite time. It is also guessed that when $m > 1$, Z_n should be of order m^n so that the population should grow explosively, at least with a positive probability. It is a standard to show that

Exercise Let $\varphi(s) = \sum_{z \in \mathbb{Z}_+} \mu(\{z\})s^z$ be the generating function of μ, we suppose $\mu(\{1\}) < 1$. Show that if $Z_0 = 1$, then the generating function of Z_n is the n-fold composition of φ with itself. Show that the probability of eventual extinction of the population satisfies $\varphi(q) = q$, and that $q > 0 \iff m > 1$. As a hint, φ is a convex function such that $\varphi'(1) = m$.

Notice that, still supposing $Z_0 = 1$, the martingale $(M_n = Z_n/m^n, n \geq 0)$ cannot be U.I. when $m \leq 1$, since it converges to 0 a.s., so $E[M_\infty] < E[M_0]$. This leaves open the question whether $P(M_\infty > 0) > 0$ in the case $m > 1$. We are going to address the problem in a particular case:

Proposition 3.2.2 *Suppose $m > 1$, $Z_0 = 1$ and $\sigma^2 = \mathrm{Var}\,(\mu) < \infty$. Then the martingale M is bounded in L^2, and hence converges a.s. and in L^2 to a variable M_∞ so that $E[M_\infty] = 1$, in particular, $P(M_\infty > 0) > 0$.*

Proof. We compute $E[Z_{n+1}^2|\mathcal{F}_n] = Z_n^2 m^2 + Z_n\sigma^2$. This shows that $E[M_{n+1}^2] = E[M_n^2] + \sigma^2 m^{-n}$, and therefore, since $m^{-n}, n \geq 0$ is summable, M is bounded in L^2 (this statement is actually equivalent to $m > 1$). □

Exercise Show that under these hypotheses, $\{M_\infty > 0\}$ and $\{\lim_n Z_n = \infty\}$ are equal, up to an event of vanishing probability.

3.3 A martingale approach to the Radon-Nikodym theorem

We begin with the following general remark. Let $(\Omega, \mathcal{F}, (\mathcal{F}_n), P)$ be a filtered probability space with $\mathcal{F}_\infty = \mathcal{F}$, and let Q be a finite non-negative measure on $(\Omega, \mathcal{F})$. Let P_n and Q_n denote the restrictions of P and Q to the measurable space $(\Omega, \mathcal{F}_n)$. Suppose that for every n, Q_n has a density M_n with respect to P_n, namely $Q_n(\mathrm{d}\omega) = M_n(\omega)P_n(\mathrm{d}\omega)$, where M_n is an $\mathcal{F}_n$-measurable non-negative function. We also sometimes let $M_n = \mathrm{d}Q_n/\mathrm{d}P_n$.

Then it is immediate that $(M_n, n \geq 0)$ is a martingale with respect to the filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_n), P)$. Indeed, $E[M_n] = Q(\Omega) < \infty$, and for $A \in \mathcal{F}_n$,

$$E^P[M_{n+1}\mathbb{1}_A] = E^{P_{n+1}}[M_{n+1}\mathbb{1}_A] = Q_{n+1}(A) = Q_n(A) = E^{P_n}[M_n\mathbb{1}_A] = E^P[M_n\mathbb{1}_A],$$

where E^P, E^{P_n} denote expectations with respect to the probability measures P, P_n. A natural problem is to wonder whether the identity $Q_n = M_n P_n$ passes to the limit $Q = M_\infty P$ as $n \to \infty$, where M_∞ is the a.s. limit of the non-negative martingale M.

Proposition 3.3.1 *Under these hypotheses, there exists a non-negative random variable* $X := \mathrm{d}Q/\mathrm{d}P$ *such that* $Q = X \cdot P$ *if and only if* $(M_n, n \geq 0)$ *is U.I.*

Proof. If M is U.I., then we can pass to the limit in $E[M_m\mathbb{1}_A] = Q(A)$ for $A \in \mathcal{F}_n$ and $m \to \infty$, to obtain $E[M_\infty\mathbb{1}_A] = Q(A)$ for every $A \in \bigcup_{n\geq 0}\mathcal{F}_n$. Since this last set is a π-system that generates $\mathcal{F}_\infty = \mathcal{F}$, we obtain $M_\infty \cdot P = Q$ by the theorem on uniqueness of measures.

Conversely, if $Q = X \cdot P$, then for $A \in \mathcal{F}_n$, we have $Q(A) = E[M_n\mathbb{1}A] = E[X\mathbb{1}A]$ so that $M_n = E[X|\mathcal{F}_n]$, which shows that M is U.I. $\square$

The Radon-Nikodym theorem (in a particular case) states as follows.

Theorem 3.3.1 (Radon-Nikodym) *Let* $(\Omega, \mathcal{F})$ *be a measurable space such that* $\mathcal{F}$ *is* separable, *i.e. generated by a countable set of events* $F_k, k \geq 1$. *Let* P *be a probability measure on* $(\Omega, \mathcal{F})$ *and* Q *be a finite non-negative measure on* $(\Omega, \mathcal{F})$. *Then the following statements are equivalent.*

(i) Q *is absolutely continuous with respect to* P, *namely*

$$\forall\, A \in \mathcal{F}, P(A) = 0 \implies Q(A) = 0.$$

(ii) $\forall\, \varepsilon > 0, \exists\, \delta > 0, \forall\, A \in \mathcal{F}, P(A) \leq \delta \implies Q(A) \leq \varepsilon.$

(iii) *There exists a non-negative random variable* X *such that* $Q = X \cdot P$.

The separability condition on $\mathcal{F}$ can actually be lifted, see Williams' book for the proof in the general case.

Proof. That (iii) implies (i) is straightforward.

If (ii) is not satisfied then we can find a sequence B_n of events and an $\varepsilon > 0$ such that $P(B_n) < 2^{-n}$ but $Q(B_n) \geq \varepsilon$. But by the Borel-Cantelli lemma, $P(\limsup B_n) = 0$, while $Q(\limsup B_n)$, as the decreasing limit of $Q(\bigcup_{k\geq n} B_k)$ as $n \to \infty$, must be $\geq \limsup_n Q(B_n) \geq \varepsilon$. Hence, (i) does not hold for the set $A = \limsup B_n$. So (i) implies (ii).

Let us now assume (ii). Let $\mathcal{F}_n$ be a filtration such that $\mathcal{F}_n$ is the σ-algebra spanned by events $F_1, \dots, F_n$, Notice that any event of $\mathcal{F}_n$ is a disjoint union of non-empty "atoms" of the form

$$\bigcap_{i\geq 1} G_i,$$

where either $G_i = F_i$ or its complementary set. We let $\mathcal{A}_n$ be the set of atoms of $\mathcal{F}_n$. Let

$$M_n(\omega) = \sum_{A \in \mathcal{A}_n} \frac{Q(A)}{P(A)} \mathbb{1}_A(\omega),$$

with the convention that $0/0 = 0$. Then it is easy to check that M_n is a density for Q_n with respect to P_n, where P_n, Q_n denote restrictions to $\mathcal{F}_n$ as above. Indeed, if $A \in \mathcal{A}_n$,

$$Q_n(A) = \frac{Q(A)}{P(A)} P(A) = E^{P_n}[M_n \mathbb{1}_A].$$

Therefore, $(M_n, n \geq 0)$ is a non-negative $(\mathcal{F}_n, n \geq 0)$-martingale, and $M_n(\omega)$ converges a.s. towards a limit $M_\infty(\omega)$. Moreover, the last proposition tells us that it suffices to shoz that (M_n) is U.I. to conclude the proof.

But note that we have $E[M_n \mathbb{1}_{\{M_n \geq a\}}] = Q(M_n \geq a)$. So for $\varepsilon > 0$ fixed, $P(M_n \geq a) \leq E[M_n]/a = Q(\Omega)/a < \delta$ for all n, as soon as a is large enough, with δ fixed by the claim, and this entails $Q(M_n \geq a) \leq \varepsilon$ for every n. Hence the result. □

Example. Let $\Omega = [0, 1)$ be endowed with its Borel σ-field, which is spanned by $\{I_{k,j} = [j2^{-k}, (j+1)2^{-k}), k \geq 0, 0 \leq j \leq 2^k - 1\}$. The intervals $I_{k,j}, 0 \leq j \leq 2^k - 1$ are called the *dyadic intervals of depth* k, they span a σ-algebra which we call $\mathcal{F}_k$. We let $\lambda(\mathrm{d}\omega)$ be the Lebesgue measure on $[0, 1)$. Let ν be a finite non-negative measure on $[0, 1)$, and

$$M_n(\omega) = 2^n \sum_{j=0}^{2^n - 1} \mathbb{1}_{I_{n,j}}(\omega) \nu(I_{n,j}),$$

then we obtain by the previous theorem that if ν is absolutely continuous with respect to λ, then $\nu = f \cdot \lambda$ for some non-negative measurable f. We then see that a.s., if $I_k(x) = [2^{-k}[2^k x], 2^{-k}([2^k x] + 1))$ denotes the dyadic interval of level k containing x,

$$2^k \int_{I_k(x)} f(x) \lambda(\mathrm{d}x) \underset{k \to \infty}{\to} f(x).$$

This is a particular case of Lebesgue differentiation theorem.

3.4 Product martingales and likelihood ratio tests

Theorem 3.4.1 (Kakutani's theorem) *Let* $(Y_n, n \geq 1)$ *a sequence of independent non-negative random variables, with mean* 1. *Let* $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$. *Then* $X_n = \prod_{1 \leq k \leq n} Y_k, n \geq 0$ *is a* $(\mathcal{F}_n, n \geq 0)$*-martingale, which converges to some* $X_\infty \geq 0$. *Letting* $a_n = E[\sqrt{Y_n}]$, *the following statements are equivalent:*

1. X *is U.I.*

2. $E[X_\infty] = 1$

3. $P(X_\infty > 0) > 0$

4. $\prod_n a_n > 0$.

Proof. The fact that M is a (non-negative) martingale follows from the fact that $E[X_{n+1}|\mathcal{F}_n] = X_n E[Y_{n+1}|\mathcal{F}_n] = X_n E[Y_{n+1}]$. For the same reason, the process

$$M_n = \prod_{k=1}^{n} \frac{\sqrt{Y_n}}{a_n}, \qquad n \geq 0$$

is a non-negative martingale with mean $E[M_n] = 1$, and $E[M_n^2] = \prod_{k=1}^n a_k^{-2}$. Thus, M is bounded in L^2 if and only if $\prod_n a_n > 0$ (notice that $a_n \in (0,1]$ e.g. by the Schwarz inequality $E[1 \cdot \sqrt{Y_n}] \leq \sqrt{E[Y_n]}$).

Now, with the standard notation $X_n^* = \sup_{0\leq k\leq n} X_k$, using Doob's L^2 inequality,

$$E[X_n^*] \leq E[(M_n^*)^2] \leq 4E[M_n^2],$$

which shows that if M is bounded in L^2, then X_∞^* is integrable, hence X is U.I. since it is dominated by X_∞^*. We thus have obtained 4. $\Longrightarrow$ 1. $\Longrightarrow$ 2. $\Longrightarrow$ 3., where the second implication comes from the optional stopping theorem for U.I. martingales, and the implication 2. $\Longrightarrow$ 3. is trivial. On the other hand, if $\prod_n a_n = 0$, since M_n converges a.s. to some $M_\infty \geq 0$, then $\sqrt{X_n} = M_n \prod_{1\leq k\leq n} a_k$ converges to 0, so that 3. does not hold. So 3. $\Longrightarrow$ 4., hence the result. □

Note that, if $Y_n > 0$ a.s. for every n, the event $\{X_\infty = 0\}$ is a tail event, so that 2. above is equivalent to $P(X_\infty > 0) = 1$ by Kolmogorov's $0-1$ law.

As an example of application of this theorem, consider a σ-finite measured space $(E, \mathcal{E}, \lambda)$ and let $\Omega = E^{\mathbb{N}}, \mathcal{F} = \mathcal{E}^{\otimes \mathbb{N}}$ be the product measurable space. We let $X_n(\omega) = \omega_n, n \geq 1$, and $\mathcal{F}_n = \sigma(\{X_1, \ldots, X_n\})$. One says that X is the *canonical (E-valued) process.*

Now suppose given two families of probability measures $(\mu_n, n \geq 1)$ and $(\nu_n, n \geq 1)$ that admit densities $\mathrm{d}\mu_n = f_n \mathrm{d}\lambda, \mathrm{d}\nu_n = g_n \mathrm{d}\lambda$ with respect to λ. We suppose that $f_n(x)g_n(x) > 0$ for every n, x. Let $P = \bigotimes_{n\geq 1} \mu_n$, resp. $Q = \bigotimes_{n\geq 1} \nu_n$ denote the measures on $(\Omega, \mathcal{F})$ under which $(X_n, n \geq 1)$ is a sequence of independent random variables with respective laws μ_n (resp. ν_n). In particular, if $A = \prod_{i=1}^n A_i \times E^{\mathbb{N}}$ is a measurable rectangle in $\mathcal{F}_n$,

$$Q(A) = \int_{E^n} \prod_{i=1}^{n} \frac{g_i(x_i)}{f_i(x_i)} \prod_{i=1}^{n} f_i(x_i)\mathrm{d}x_i = E^P(M_n \mathbb{1}_A),$$

where E^P denotes expectation with respect to P, and

$$M_n = \prod_{i=1}^{n} \frac{g_i(X_i)}{f_i(X_i)}.$$

Since measurable rectangles of $\mathcal{F}_n$ form a π-system that span $\mathcal{F}_n$, the probability $Q|_{\mathcal{F}_n}$ is absolutely continuous with respect to $P|_{\mathcal{F}_n}$, with density M_n, so that $(M_n, n \geq 1)$ is a non-negative martingale with respect to the filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_n, n \geq 0), P)$. Kakutani's theorem then shows that M converges a.s. and in L^1 to its limit M_∞ if and only if

$$\prod_{n\geq 1} \int_E \sqrt{f_n(x)g_n(x)}\lambda(\mathrm{d}x) > 0 \iff \sum_{n\geq 1} \int_E \left(\sqrt{f_n(x)} - \sqrt{g_n(x)}\right)^2 \lambda(\mathrm{d}x).$$

In this case, one has $Q(A) = E^P[M_\infty \mathbb{1}_A]$ for every measurable rectangle of $\mathcal{F}$, and Q is absolutely continuous with respect to P with density M_∞. In the opposite case, $M_\infty = 0$, so Proposition 3.3.1 shows that Q and P are carried by two disjoint measurable sets.

3.4.1 Example: consistency of the likelihood ratio test

In the case where $\mu_n = \mu, \nu_n = \nu$ for every n, we see that $M_\infty = 0$ a.s. if (and only if) $\mu \neq \nu$. This is called the *consistency of the likelihood ratio test* in statistics. Let us recall the background for the application of this test. Suppose given an i.i.d. sample $X_1, X_2, \ldots, X_n$, with an unknown common distribution. Suppose one wants to test the hypothesis (H_0) that this distribution is P against the hypothesis (H_1) that it is Q, where P and Q have everywhere positive densities f, g with respect to some common σ-finite measure λ (for example, a normal distribution and a Cauchy distribution). Letting $M_n = \prod_{1 \leq i \leq n} g(X_i)/f(X_i)$, we use the test $\mathbb{1}_{\{M_n \leq 1\}}$ for acceptance of H_0 against H_1. Then supposing H_0, then $M_\infty = 0$ a.s. so the probability of rejection $P(M_n > 1)$ converges to 0. Similarly, supposing H_1, then $M_\infty = +\infty$ a.s. so the probability of rejection goes to 1.

Chapter 4

Continuous-parameter stochastic processes

In this section, we will consider the case when processes are indexed by a real interval $I \subset \mathbb{R}$, with non-empty interior, in many cases I will be $\mathbb{R}_+$. This makes the whole study more involved, as we now stress here. In all what follows, the states space E is assumed to be a metric space, usually $E = \mathbb{R}$ or $E = \mathbb{R}^d$ endowed with the Euclidean norm.

4.1 Theoretical problems when dealing with continuous time processes

Although the definitions for filtrations, adapted processes, stopping times, martingales, super-, sub-martingales are not changed when compared to the discrete case (see the beginning of Section 2), the use of continuous time induces important measurability problems. Indeed, there is no reason why an adapted process $(\omega, t) \mapsto X_t(\omega)$ should be a measurable map defined on $\Omega \times I$, or even the *sample path* $t \mapsto X_t(\omega)$ for any fixed ω. In particular, stopped processes like $X_T \mathbb{1}_{\{T<\infty\}}$ for a stopping time T have no reason to be random variables.

Even worse, there are in general "very few" stopping times — for example first entrance times $\inf\{t \geq 0 : X_t \in A\}$ for measurable (or even open or closed) subsets of the states space E need not be stopping times.

This is the reason why we add *a priori* requirements on the regularity of random processes under consideration. A quite natural requirement is that they are continuous processes, i.e. that $t \mapsto X_t(\omega)$ is continuous for a.e. ω, because a continuous function is determined by its values on a *countable* dense subset of I. More generally, we will consider processes that are right-continuous and admit left limits everywhere, a.s. — such processes are called *càdlàg*, and are also determined by the values they take on a countable dense subset of I (the notation càdlàg stands for the French 'continu à droite, limité à gauche').

We let $C(I, E), D(I, E)$ denote the spaces of continuous and càdlàg functions from I to E, we consider these sets as measurable spaces by endowing them with the *product* σ-algebra that makes the *projections* $\pi_t : X \mapsto X_t$ measurable for every $t \in I$. Usually, we will consider processes with values in $\mathbb{R}$, or sometimes $\mathbb{R}^d$ for some $d \geq 1$ in the chapter on Brownian motion. The following proposition holds, of which (ii) is an analog of 1., 2. in Proposition 2.1.2.

Proposition 4.1.1 *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t, t \in I), P)$ be a filtered probability space, and let $(X_t, t \in I)$ be an adapted process with values in E.*

(i) *Suppose X is continuous (i.e. $(X_t(\omega), t \in I) \in C(I, E)$ for every ω). If A is a closed set, and $\inf I > -\infty$, then the random time*

$$T_A = \inf\{t \in I : X_t \in A\}$$

is a stopping time.

(ii) *Let T be a stopping time, and suppose X is càdlàg. Then $X_T \mathbb{1}_{\{T<\infty\}} : \omega \mapsto X_{T(\omega)}(\omega)\mathbb{1}_{\{T(\omega)<\infty\}}$ is an $\mathcal{F}_T$-measurable random variable. Moreover, the stopped process $X^T = (X_{T\wedge t}, t \geq 0)$ is adapted.*

Proof. For (i), notice that if A is closed and X is continuous, then for every $t \in I$,

$$\{T_A \leq t\} = \left\{ \inf_{s \in I \cap \mathbb{Q}, s \leq t} d(X_s, A) = 0 \right\},$$

where $d(x, A) = \inf_{y \in A} d(x, y)$ is the distance from x to the set A. Indeed, if $X_s \in A$ for some $s \leq t$, then for q_n converging to s in $\mathbb{Q} \cap I \cap (-\infty, t]$, X_{q_n} converges to X_s, so that $d(X_{q_n}, A)$ converges to 0. Conversely, if there exists $q_n \in \mathbb{Q} \cap I \cap (-\infty, t]$ such that $d(X_{q_n}, A)$ converges to 0, then since $\inf I > -\infty$ we can extract along a subsequence and assume q_n converges to some $s \in I \cap (-\infty, t]$, and this s has to satisfy $d(X_s, A) = 0$ by continuity of X. Since A is closed, this implies $X_s \in A$, so that $\{T_A \leq t\}$.

For (ii), first note that a random variable Z is $\mathcal{F}_T$-measurable if $Z\mathbb{1}_{\{T \leq t\}} \in \mathcal{F}_t$ for every $t \in I$, by approximating Z by a finite sum of the form $\sum \alpha_i \mathbb{1}_{A_i}$, for $A_i \in \mathcal{F}_T$.

Notice also that if T is a stopping time, then, if $\lceil x \rceil$ denotes smallest $n \in \mathbb{Z}_+$ with $n \geq x$, $T_n = 2^{-n}\lceil 2^n T \rceil$ is also a stopping time with $T_n \geq T$, that decreases to T as $n \to \infty$ ($T_n = \infty$ if $T = \infty$). Indeed, $\{T_n \leq t\} = \{T \leq 2^{-n}\lfloor 2^n t \rfloor\} \in \mathcal{F}_t$ (notice $\lceil x \rceil \leq y$ if and only if $x \leq \lfloor y \rfloor$, where $\lfloor y \rfloor$ is the largest $n \in \mathbb{Z}_+$ with $n \leq x$). Moreover, T_n takes values in the set $D_n^* = \{k2^{-n}, k \in \mathbb{Z}_+\} \sqcup \{\infty\}$ of dyadic numbers with level n (or ∞).

Therefore, $X_T \mathbb{1}_{\{T<\infty\}} \mathbb{1}_{\{T \leq t\}} = X_t \mathbb{1}_{\{T=t\}} + X_T \mathbb{1}_{\{T<t\}}$, which by the càdlàg property is equal to

$$X_t \mathbb{1}_{\{T=t\}} + \lim_{n \to \infty} X_{T_n \wedge t} \mathbb{1}_{\{T<t\}}.$$

The variables $X_t \mathbb{1}_{\{T=t\}}$ and $X_{T_n \wedge t} \mathbb{1}_{\{T<t\}}$ are $\mathcal{F}_t$-measurable, because

$$X_{T_n \wedge t} = \sum_{d \in D_n^*, d \leq t} X_d \mathbb{1}_{\{T_n = d\}} + X_t \mathbb{1}_{\{t < T_n\}}.$$

hence the result. For the statement on the stopped process, notice that for every t, $X_{T \wedge t}$ is $\mathcal{F}_{T \wedge t}$, hence $\mathcal{F}_t$-measurable. □

It turns out that (i) does not hold in general for càdlàg processes, although it is a very subtle problem to find counterexamples. See Rogers and Williams' book, Chapters II.74

and II.75. In particular, Lemma 75.1 therein shows that T_A is a stopping time if A is compact and X is an adapted càdlàg process, whenever the filtration $(\mathcal{F}_t, t \in I)$ satisfies the so-called "usual conditions" — see Section 4.3 for the definition of these conditions.

You may check as an exercise that the times T_A for open sets A associated with càdlàg processes, are stopping times with respect to the filtration $(\mathcal{F}_{t+}, t \in I)$, where

$$\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s.$$

Somehow, the filtration $(\mathcal{F}_{t+})$ foresees what will happen 'just after' t.

4.2 Finite marginal distributions, versions

We now discuss the notion of *law* of a process. If $(X_t, t \in I)$ is a stochastic process, we can consider it as a random variable with values in the set E^I of maps $f : I \to E$, where this last space is endowed with the product σ-algebra (the smallest σ-algebra that makes the projections $f \in E^I \mapsto f(t)$ measurable for every $t \in I$). It is then natural to consider the image measure μ of the probability P by the process X as the law of X. However, this measure is uneasy to manipulate, and the quantities that are of true interest are the following simpler objects.

Definition 4.2.1 *Let $(X_t, t \in I)$ be a process. For every finite $J \subset I$, the* finite marginal distribution *of X indexed by J is the law μ_J of the E^J-valued random variable $(X_t, t \in J)$.*

It is a nice fact that the finite marginal distributions $\{\mu_J : J \subset I, \#J < \infty\}$ uniquely characterize the law μ of the process $(X_t, t \in I)$ as defined above. Indeed, by definition, if X and Y are càdlàg processes having the same finite marginal laws, then their distribution agree on the π-system of finite "rectangles" of the form $\prod_{i \in J} A_s \times \prod_{t \in I \setminus J} E$ for finite $J \subset I$, which generate the product σ-algebra, hence the distributions under consideration are equal. Notice that this uniqueness result does not imply the *existence* of a process with given marginal distributions.

The problem with (finite marginal) laws of processes is that they are powerless in dealing with properties of processes that involve more than countably many times, such as continuity or càdlàg properties of the process. For example, if X is a continuous process, there are (many!) non-continuous processes that have the same finite-marginal distributions as X: the finite marginal distributions just do not 'see' the sample path properties of the process. This motivates the following definition.

Definition 4.2.2 *If X and X' are two processes defined on some common probability space $(\Omega, \mathcal{F}, P)$, we say that X' is a* version *of X if for every t, $X_t(\omega) = X'_t(\omega)$ a.s.*

In particular, two versions X and X' of the same process share the same finite-dimensional distribution, however, this does *not* say that there exists an ω so that $X_t(\omega) = X'_t(\omega)$ for every t. This becomes true if both X and X' are *a priori* known to be càdlàg, for instance.

Example. To explain these very abstract notions, suppose we want to find a process $(X_t, 0 \leq t \leq 1)$ whose finite marginal laws are Dirac masses at 0, namely

$$\mu_J(\{(\underbrace{0, 0, \ldots, 0}_{\#J \text{ times}})\}) = P(X_s = 0, s \in J) = 1$$

for every finite $J \subset [0, 1]$. Of course, the process $X_t = 0, 0 \leq t \leq 1$ satisfies this. However, the process $X'_t = \mathbb{1}_{\{U\}}(t), 0 \leq t \leq 1$, where U is a uniform random variable on $[0, 1]$, is a version of X, and therefore has the same law as X. But of course, it is not continuous, and $P(X'_t = 0 \,\forall\, t \in [0, 1]) = 0$. We thus want to consider it as a 'bad' version of the zero process. This example motivates the following way of dealing with processes: when considering a process whose finite marginal distributions are known, we first try to find the most regular version of the process as we can before studying it.

We will discuss two 'regularization theorems' in this course, the martingale regularization theorem and Kolmogorov's continuity criterion, which are instances of situations when there exists a regular (continuous or càdlàg) version of the stochastic process under consideration.

4.3 The martingale regularization theorem

We consider here a martingale $(X_t, t \geq 0)$ on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t, t \geq 0), P)$. We let $\mathcal{N}$ be the set of events in $\mathcal{F}$ with probability 0,

$$\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s, \qquad t \geq 0,$$

and $\widetilde{\mathcal{F}}_t = \mathcal{F}_{t+} \cup \mathcal{N}$.

Theorem 4.3.1 *Let $(X_t, t \geq 0)$ be a martingale. Then there exists a càdlàg process $\widetilde{X}$ which is a martingale with respect to the filtered probability space $(\Omega, \mathcal{F}, (\widetilde{\mathcal{F}}_t, t \geq 0), P)$, so that for every $t \geq 0$, $X_t = E[\widetilde{X}_t | \mathcal{F}_t]$ a.s. If $\widetilde{\mathcal{F}}_t = \mathcal{F}_t$ for every $t \geq 0$, $\widetilde{X}$ is therefore a càdlàg version of X.*

We say that $(\mathcal{F}_t, t \geq 0)$ satisfies the *usual conditions* if $\widetilde{\mathcal{F}}_t = \mathcal{F}_t$ for every t, that is, $\mathcal{N} \subseteq \mathcal{F}_0$ and $\mathcal{F}_{t+} = \mathcal{F}_t$ (a filtration satisfying this last condition for $t \geq 0$ is called *right-continuous*, notice that $(\mathcal{F}_{t+}, t \geq 0)$ is right-continuous for every filtration $(\mathcal{F}_t, t \geq 0)$). As a corollary of Theorem 4.3.1, in the case when the filtration satisfies the usual conditions, a martingale admits a càdlàg version so there is "little to lose" to consider that martingales are càdlàg.

Lemma 4.3.1 *A function $f : \mathbb{Q}_+ \to \mathbb{R}$ admits a left and a right (finite) limit at every $t \in \mathbb{R}_+$ if and only if for every rationals $a < b$ and bounded $I \subset \mathbb{Q}$, f is bounded on I and the number*

$$N([a, b], I, f) = \sup \left\{ n \geq 0 : \begin{array}{c} \exists\, 0 \leq s_1 < t_1 < \ldots < s_n < t_n, \text{ all in } I, \\ f(s_i) < a, f(t_i) > b, 1 \leq i \leq n \end{array} \right\}$$

a upcrossings of f from a to b is finite.

Proof of Theorem 4.3.1. We first show that X is bounded on bounded subsets of $\mathbb{Q}_+$. Indeed, if I is such a subset and $J = \{a_1, \dots, a_k\}$ is a finite subset of I with $a_1 < \dots < a_n$, then $M_l = X_{a_l}, 1 \leq l \leq k$ is a martingale. Doob's maximal inequality applied to the submartingale $|M|$ then shows that

$$cP(M_k^* > c) = cP(\max_{1 \leq l \leq k} |X_{a_l}| > c) \leq E[|X_{a_k}|] \leq E[|X_K|]$$

for any $K > \sup I$. Therefore, taking a monotone limit over finite $J \subset I$ with union I, we have

$$cP(\sup_{t \in I} |X_t| > c) \leq E[|X_K|].$$

This shows that $P(\sup_{t \in I} |X_t| < \infty) = 1$ by letting $c \to \infty$.

Let I still be a bounded subset of $\mathbb{R}_+$, and $a < b \in \mathbb{Q}_+$. By definition, we have $N([a,b], I, X) = \sup_{J \subset I, \text{finite}} N([a,b], J, X)$. So let $J \subset I$ be a finite subset of the form $\{a_1, a_2, \dots, a_k\}$ as above, and again let $M_l = X_{a_l}, 1 \leq l \leq k$. Doob's upcrossing lemma for this martingale gives

$$(b-a)E[N([a,b], J, X)] \leq E[(X_{a_k} - a)^-] \leq E[(X_K - a)^-],$$

for any $K \geq \sup I$, because $((X_t - a)^-, t \geq 0)$ is a submartingale due to the convexity of $x \mapsto (x-a)^-$. Taking the supremum over J shows that $N([a,b], I, X)$ is a.s. bounded, because $E[|X_K|] < \infty$. This shows by letting $K \to \infty$ along integers, that $N([a,b], I, X)$ is finite for every bounded subset I of $\mathbb{Q}_+$, and every $a < b$ rationals, for every ω in an event Ω_0 with probability 1. Therefore, we can define

$$\widetilde{X}_t(\omega) = \lim_{s \in \mathbb{Q}_+, s > t} X_s(\omega), \qquad \omega \in \Omega_0$$

and $\widetilde{X}_t(\omega) = 0$ for every t for $\omega \notin \Omega_0$. The process $\widetilde{X}$ thus obtained then is indeed adapted to the filtration $(\widetilde{\mathcal{F}}_t, t \geq 0)$. It remains to show that $\widetilde{X}$ is an $(\widetilde{\mathcal{F}}_t)$-martingale, satisfies $E[\widetilde{X}_t | \mathcal{F}_t] = X_t$, and is càdlàg.

First, check that if X remains an $(\mathcal{F}_t \vee \mathcal{N}, t \in I)$-martingale, because $E[X | \mathcal{G} \vee \mathcal{N}] = E[X|\mathcal{G}]$ in $L^1(\Omega, \mathcal{G} \vee \mathcal{N}, P)$ for any integrable X and sub-σ-algebra $\mathcal{G} \in \mathcal{F}$. Thus, we may suppose that $\mathcal{N} \subset \mathcal{F}_t$ for every t. Let $s < t \in \mathbb{R}_+$, and $s_n, n \geq 0$ be a (strictly) decreasing sequence of rationals that converges to s, with $s_0 < t$. Then $\widetilde{X}_s = \lim X_{s_n} = \lim E[X_t | \mathcal{F}_{s_n}]$ by definition for $\omega \in \Omega_0$. Now, the process $(M_n = X_{s_{-n}}, n \leq 0)$ is a backwards martingale with respect to the filtration $(\mathcal{G}_n = \mathcal{F}_{s_{-n}}, n \leq 0)$. The backwards martingale convergence theorem thus shows that $\widetilde{X}_s = E[X_t | \mathcal{F}_{s+}]$, and therefore $X_t = E[\widetilde{X}_t | \mathcal{F}_t]$. Moreover, taking a rational sequence (t_n) decreasing to t and using again the backwards martingale convergence theorem, (X_{t_n}) converges to $\widetilde{X}_t$ in L^1, so that $\widetilde{X}_s = E[\widetilde{X}_t | \widetilde{\mathcal{F}}_s]$ for every $s \leq t$.

The only thing that remains to prove is the càdlàg property. If $t \in \mathbb{R}_+$ and if $\widetilde{X}_s(\omega)$ does not converge to $\widetilde{X}_t(\omega)$ as $s \downarrow t$, then $|\widetilde{X}_t - \widetilde{X}_s| > \varepsilon$ for some $\varepsilon > 0$ and for infinitely many $s > t$, so that if $\omega \in \Omega_0$, $|\widetilde{X}_t - X_u| > \varepsilon/2$ for an infinite number of rationals $u > t$, contradicting $\omega \in \Omega_0$. The argument for showing that $\widetilde{X}$ has left limits is similar. □

From now on, when considering martingales in continuous time, we will always take their càdlàg version, provided the underlying filtration satisfies the usual hypotheses.

4.4 Doob's inequalities and convergence theorems for martingales in continuous time

Considering càdlàg martingales makes it straightforward to generalize to the continuous case the inequalities of section 2.4, by density arguments. We leave to the reader to show the following theorems which are analog to the discrete-time case.

Proposition 4.4.1 (A.s. convergence) *Let $(X_t, t \geq 0)$ be a càdlàg martingale which is bounded in L^1. Then X_t converges as $t \to \infty$ a.s. to an (a.s.) finite limit X_∞.*

To prove this, notice that convergence of X_t as $t \to \infty$ to a (possibly infinite) limit is equivalent to the fact that the number of upcrossings of X from below a to above b over the time interval $\mathbb{R}_+$ is finite for every $a < b$ rationals. However, by the càdlàg property, it suffices to restrict our attention to the countable time set $\mathbb{Q}_+$ rather than $\mathbb{R}_+$. Indeed, for each upcrossing of X from a to b between times $s < t$ say, we can find rationals $s' > s, t' > t$ as close to s, t as wanted so that X accomplishes an upcrossing from a to b between times s', t', and this implies that $N(X, \mathbb{R}_+, [a, b]) = N(X, \mathbb{Q}_+, [a, b])$ (possibly infinite). Then, use similar arguments as those used in the first part of the proof of Theorem 4.3.1.

Proposition 4.4.2 (Doob's inequalities) *If $(X_t, t \geq 0)$ is a càdlàg martingale and $X_t^* = \sup_{0 \leq s \leq t} |X_s|$, then for every $c > 0, t \geq 0$,*

$$cP(X_t^* \geq c) \leq E[|X_t|].$$

Moreover, if $p > 1$ then

$$\|X_t^*\|_p \leq \frac{p}{p-1} \|X_t\|_p.$$

To prove this, notice that $X_t^* = \sup_{s \in \{t\} \cup ([0,t] \cap \mathbb{Q})} |X_s|$ by the càdlàg property.

Proposition 4.4.3 (L^p convergence) (i) *If X is a càdlàg martingale and $p > 1$ then $\sup_{t \geq 0} \|X_t\|_p < \infty$ if and only if X converges a.s. and in L^p to its limit X_∞, and this if and only if X is closed in L^p, i.e. there exists $Z \in L^p$ so that $E[Z|\mathcal{F}_t] = X_t$ for every t, a.s. (one can then take $Z = X_\infty$).*

(ii) *If X is a càdlàg martingale then X is U.I. if and only if X converges a.s. and in L^1 to its limit X_∞, and this if and only if X is closed (in L^1).*

Proposition 4.4.4 (Optional stopping) *Let X be a càdlàg U.I. martingale. Then for every stopping times $S \leq T$, one has $E[X_T|\mathcal{F}_S] = X_S$ a.s.*

Proof. Let T_n be the stopping time $2^{-n}\lceil 2^n T \rceil$ as defined in the proof of Proposition 4.1.1. The right-continuity of paths of X shows that X_{T_n} converges to X_T a.s. Moreover, T_n takes values in the countable set D_n^* of dyadic rationals of level n (and ∞), so that

$$E[X_\infty|\mathcal{F}_{T_n}] = \sum_{d \in D_n^*} E[\mathbb{1}_{\{T_n = d\}} X_\infty|\mathcal{F}_{T_n}] = \sum_{d \in D_n^*} \mathbb{1}_{\{T_n = d\}} E[X_\infty|\mathcal{F}_d]$$

(you should check this carefully). Now, since X_t converges to X_∞ in L^1, $X_d = E[X_t|\mathcal{F}_d] = E[X_\infty|\mathcal{F}_d]$ a.s., and $E[X_\infty|\mathcal{F}_{T_n}] = X_{T_n}$. Passing to the limit as $n \to \infty$ and using the backwards martingale convergence theorem, we obtain $E[X_\infty|\mathcal{F}'_T] = X_T$ where $\mathcal{F}'_T = \bigcap_{n\geq 1} \mathcal{F}_{T_n}$, and therefore $E[X_\infty|\mathcal{F}_T] = X_T$ by the tower property, since X_T is $\mathcal{F}_T$-measurable. The theorem then follows as in Theorem 2.6.1. □

4.5 Kolmogorov's continuity criterion

Theorem 4.5.1 (Kolmogorov's continuity criterion) *Let $(X_t, 0 \leq t \leq 1)$ be a stochastic process with real values. Suppose there exist $p > 0, c > 0, \varepsilon > 0$ so that for every $s, t \geq 0$,*

$$E[|X_t - X_s|^p] \leq c|t-s|^{1+\varepsilon}.$$

Then, there exists a modification $\widetilde{X}$ of X which is a.s. continuous (and even α-Hölder continuous for any $\alpha \in (0, \varepsilon/p)$).

Proof. Let $D_n = \{k \cdot 2^{-n}, 0 \leq k \leq 2^n\}$ denote the dyadic numbers of $[0,1]$ with level n, so D_n increases as n increases. Then letting $\alpha \in (0, \varepsilon/p)$, Markov's inequality gives for $0 \leq k < 2^n$,

$$P(|X_{k2^{-n}} - X_{(k+1)2^{-n}}| > 2^{-n\alpha}) \leq 2^{np\alpha} E[|X_{k2^{-n}} - X_{(k+1)2^{-n}}|^p] \leq 2^{np\alpha} 2^{-n-n\varepsilon} \leq 2^{-n} 2^{-(\varepsilon - p\alpha)n}.$$

Summing over D_n we obtain

$$P\left(\sup_{0\leq k<2^n} |X_{k2^{-n}} - X_{(k+1)2^{-n}}| > 2^{-n\alpha}\right) \leq 2^{-n(\varepsilon-p\alpha)},$$

which is summable. Therefore, the Borel-Cantelli lemma shows that for a.a. ω, there exists N_ω so that if $n \geq N_\omega$, the supremum under consideration is $\leq 2^{-n\alpha}$. Otherwise said, a.s.,

$$\sup_{n\geq 0} \sup_{k\in\{0,\ldots,2^n-1\}} \frac{|X_{k2^{-n}} - X_{(k+1)2^{-n}}|}{2^{-n\alpha}} \leq M(\omega) < \infty.$$

We claim that this implies that for every $s, t \in D = \bigcup_{n\geq 0} D_n$, $|X_s - X_t| \leq M'(\omega)|t-s|^\alpha$, for some $M'(\omega) < \infty$ a.s. Indeed, if $s, t \in D$, $s < t$, and if r is the least integer such that $t - s > 2^{-r-1}$ we can write $[s, t)$ as a disjoint unions of intervals of the form $[r, r + 2^{-n})$ with $r \in D_n$ and $n > r$, in such a way that for every $n > r$, at most two of these intervals have length 2^{-n}. This entails that

$$|X_s - X_t| \leq 2 \sum_{n\geq r+1} M(\omega) 2^{-n\alpha} \leq 2(1-2^{-\alpha})^{-1} M(\omega) 2^{-(r+1)\alpha} \leq M'(\omega)|t-s|^\alpha$$

where $M'(\omega) < \infty$ a.s. Therefore, the process $(X_t, t \in D)$ is a.s. uniformly continuous (and even α-Hölder continuous). Since D is an everywhere dense set in $[0,1]$, the latter process a.s. admits a unique continuous extension $\widetilde{X}$ on $[0,1]$, which is also α-Hölder continuous (it is consistently defined by $\widetilde{X}_t = \lim_n X_{t_n}$, where $(t_n, n \geq 0)$ is any D-valued sequence converging to t). On the exceptional set where $(X_d, d \in D)$ is not uniformly continuous, we let $\widetilde{X}_t = 0, 0 \leq t \leq 1$, so $\widetilde{X}$ is continuous. It remains to show that $\widetilde{X}$ is a version of X. To this end, we estimate by Fatou's lemma

$$E[|X_t - \widetilde{X}_t|^p] \leq \liminf_n E[|X_t - X_{t_n}|^p],$$

where $(t_n, n \geq 0)$ is any D-valued sequence converging to t. But since $E[|X_t - X_{t_n}|^p] \leq c|t - t_n|^{1+\varepsilon}$, this converges to 0 as $n \to \infty$. Therefore, $X_t = \widetilde{X}_t$ a.s. for every t. □

The nice thing about this criterion is that is depends only on a control on the two-dimensional marginal distributions of the stochastic process.

In fact, the very same proof can give the following alternative

Corollary 4.5.1 *Let $(X_d, d \in D)$ be a stochastic process indexed by the set D of dyadic numbers in $[0, 1]$. Assume that there exist $c, p, \varepsilon > 0$ so that for every $s, t \in D$,*

$$E[|X_s - X_t|^p] \leq c|s - t|^{1+\varepsilon},$$

thenalmost-surely, the process $(X_d, d \in [0, 1])$ has an extension $(X_t, t \in [0, 1])$ that is continuous, and even Hölder-continuous of any index $\alpha \in (0, \varepsilon/p)$.

Chapter 5

Weak convergence

5.1 Definition and characterizations

Let (M, d) be a metric space, endowed with its Borel σ-algebra. All measures in this chapter will be measures on such a measurable space. Let $(\mu_n, n \geq 0)$ be a sequence of probability measures on M. We say that μ_n converges *weakly* to the non-negative measure μ if for every continuous bounded function $f : M \to \mathbb{R}$, one has $\mu_n(f) \to \mu(f)$. Notice that in this case, μ is automatically a probability measure since $\mu(1) = 1$, and the definition actually still makes sense if we suppose that μ_n (resp. μ) are just finite non-negative measures on M.

Examples. Let $(x_n, n \geq 0)$ be a M-valued sequence that converges to x. Then δ_{x_n} converges weakly to δ_x, where δ_a is the Dirac mass at a. This is just saying that $f(x_n) \to f(x)$ for continuous functions.

Let $M = [0, 1]$ and $\mu_n = n^{-1} \sum_{0 \leq k \leq n-1} \delta_{k/n}$. Then $\mu_n(f)$ is the Riemann sum $n^{-1} \sum_{0 \leq k \leq n-1} f(k/n)$, which converges to $\int_0^1 f(x)\mathrm{d}x$ if f is continuous, which shows that μ_n converges weakly to Lebesgue's measure on $[0, 1]$.

In this two cases, notice that it is *not* true that $\mu_n(A)$ converges to $\mu(A)$ for every Borel set A convergence. This 'pointwise convergence' is stronger, but much more rigid than weak convergence. For example, δ_{x_n} does not converge in that sense to δ_x unless $x_n = x$ eventually. See e.g. Chapter III in Stroock's book for a discussion on the various existing notions of convergence for measures.

Theorem 5.1.1 *Let $(\mu_n, n \geq 0)$ be a sequence of probability distributions. The following assertions are equivalent:*

1. *μ_n converges weakly to μ*
2. *For every open subset G of M, $\liminf_n \mu_n(G) \geq \mu(G)$ ('open sets can lose mass')*
3. *For every closed subset F of M, $\limsup_n \mu_n(F) \leq \mu(F)$ ('closed sets can gain mass')*
4. *For every Borel subset A in M with $\mu(\partial A) = 0$, $\lim_n \mu_n(A) = \mu(A)$. ('mass is lost or gained through the boundary')*

Proof. 1. $\Longrightarrow$ 2. Let G be an open subset with nonempty complement G^c. The distance function $d(x, G^c)$ is continuous and positive if and only if $x \in G$. Let $f_M = 1 \wedge (Md(x, G^c))$. Then f_M increases to $\mathbb{1}_G$ as $M \uparrow \infty$. Now, $\mu_n(f_M) \leq \mu_n(G)$ converges to $\mu(f_M)$, so that $\liminf_n \mu_n(G) \geq \mu(f_M)$ for every M, and by monotone convergence letting $M \uparrow \infty$, one gets the result.

2. $\Longleftrightarrow$ 3. is obvious by taking complementary sets.

2.,3. $\Longrightarrow$ 4. Let A° and $\overline{A}$ respectively denote the interior and the closure of A. Since $\mu(\partial A) = \mu(\overline{A} \setminus A^\circ) = 0$, we obtain $\mu(A^\circ) = \mu(A) = \mu(\overline{A})$.

$$\limsup_n \mu_n(\overline{A}) \leq \mu(A) \leq \liminf_n \mu_n(A^\circ),$$

and since $A^\circ \subset \overline{A}$, this gives the result.

4. $\Longrightarrow$ 1. Let $f : M \to \mathbb{R}_+$ be a continuous bounded non-negative function, then using Fubini's theorem,

$$\int_M f(x)\mu_n(\mathrm{d}x) = \int_M \mu_n(\mathrm{d}x) \int_0^\infty \mathbb{1}_{\{t \leq f(x)\}} \mathrm{d}t = \int_0^K \mu_n(\{f \geq t\})\mathrm{d}t,$$

where K is any upper bound for f. Now $\{f \geq t\} := \{x : f(x) \geq t\}$ is a closed subset of M, whose boundary is included in $\{f = t\}$, because $\{f > t\}$ is open and included in $\{f \geq t\}$, and their difference is $\{f = t\}$. However, there can be at most a countable set of numbers t such that $\mu(\{f = t\}) > 0$, because

$$\{t : \mu(\{f = t\}) > 0\} = \bigcup_{n \geq 1} \{t : \mu(\{f = t\}) \geq n^{-1}\},$$

and the n-th set on the right-hand side has at most n elements. Therefore, for Lebesgue-almost all t, $\mu(\partial\{f \geq t\}) = 0$ and therefore, 4. and dominated convergence over the finite interval $[0, K]$, where the integrated quantities are bounded by 1, show that $\mu_n(f)$ converges to $\int_0^K \mu(\{f \geq t\})\mathrm{d}t = \mu(f)$. The case of functions taking values of both signs is immediate. □

As a consequence, one obtains the following important criterion for weak convergence of measures on $\mathbb{R}$. Recall that the distribution function of a non-negative finite measure μ on $\mathbb{R}$ is the càdlàg function defined by $F_\mu(x) = \mu((-\infty, x]), x \in \mathbb{R}$.

Proposition 5.1.1 *Let $\mu_n, n \geq 0, \mu$ be probability measures on $\mathbb{R}$. Then the following statements are equivalent:*

1. *μ_n converges weakly to μ*
2. *for every $x \in \mathbb{R}$ such that F_μ is continuous at x, $F_{\mu_n}(x)$ converges to $F_\mu(x)$ as $n \to \infty$.*

Proof. The continuity of F_μ at x exactly says that $\mu(\partial A_x) = 0$ where $A_x = (-\infty, x]$, so 1. $\Longrightarrow$ 2. is immediate by Theorem 5.1.1.

Conversely, let G be an open subset of $\mathbb{R}$, which we write as a countable union $\bigcup_k (a_k, b_k)$ of disjoint open intervals. Then

$$\mu_n(G) = \sum_k \mu_n((a_k, b_k)), \tag{5.1}$$

while for every k and $a_k < a' < b' < b_k$. ,

$$\mu_n((a_k, b_k)) = F_{\mu_n}(b_k-) - F_{\mu_n}(a_k) \geq F_{\mu_n}(b') - F_{\mu_n}(a').$$

If we take a', b' to be continuity points of F_μ, we then obtain $\liminf_n \mu_n((a_k, b_k)) \geq F_\mu(b') - F_\mu(a')$. Letting $a' \downarrow a_k$, $b' \uparrow b_k$ along continuity points of F_μ (such points always form a dense set in $\mathbb{R}$) gives $\liminf_n \mu_n((a_k, b_k)) \geq \mu((a_k, b_k))$. On the other hand, applying Fatou's lemma to (5.1) yields $\liminf_n \mu_n(G) \geq \sum_k \liminf_n \mu_n((a_k, b_k))$, whence $\liminf_n \mu_n(G) \geq \mu(G)$. □

5.2 Convergence in distribution for random variables

If $(X_n, n \geq 0)$ is a sequence of random variables with values in a metric space (M, d), and defined on possibly different probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$, we say that X_n converges *in distribution* to a random variable X on $(\Omega, \mathcal{F}, P)$ if the law of X_n converges weakly to that of X. Otherwise said, X_n converges in distribution to X if for every continuous bounded function f, $E[f(X_n)]$ converges to $E[f(X)]$.

The two following examples are the probabilistic counterpart of the examples discussed in the beginning of the previous section.

Examples. If (x_n) is a sequence in M that converges to x, then x_n converges as $n \to \infty$ to x in distribution, if the $x_n, n \geq 0$ and x are considered as random variables!

If U is a uniform random variable on $[0, 1)$ and $U_n = n^{-1}\lfloor nU \rfloor$, we see that U_n has law μ_n and converges in distribution to U.

In the two cases we just discussed, the variables under consideration even converge a.s., which directly entails convergence in distribution, see the example sheets.

The notion of convergence in distribution is related to the other notions of convergence for random variables as follows. See the Example sheet 3 for the proof.

Proposition 5.2.1 *1. If $(X_n, n \geq 1)$ is a sequence of random variables that converges in probability to some random variable X, then X_n converges in distribution to X.*

2. If $(X_n, n \geq 1)$ is a sequence of random variables that converges in distribution to some constant random variable c, then X_n converges to c in probability.

Using Proposition 5.1.1, we can discuss the following

Example: the central limit theorem. The central limit theorem says that if $(X_n, n \geq 1)$ is a sequence of iid random variables in L^2 with $m = E[X_1]$ and $\sigma^2 = \mathrm{Var}\,(X_1)$, then for every $a < b$ in $\mathbb{R}$,

$$P\left(a \leq \frac{S_n - mn}{\sigma\sqrt{n}} \leq b\right) \underset{n\to\infty}{\to} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} \mathrm{d}x,$$

where $S_n = X_1 + \ldots + X_n$. This is exactly saying that $(S_n - mn)/(\sigma\sqrt{n})$ converges in distribution as $n \to \infty$ to a Gaussian $\mathcal{N}(0,1)$ random variable.

5.3 Tightness

Definition 5.3.1 *Let $\{\mu_i, i \in I\}$ be a family of probability measures on M. This family is said to be* tight *if for every $\varepsilon > 0$, there exists a compact subset $K \subset M$ such that*

$$\sup_{i \in I} \mu_i(M \setminus K) \leq \varepsilon,$$

i.e. most of the mass of μ_i is contained in K, uniformly in $i \in I$.

Proposition 5.3.1 (Prokhorov's theorem) *Suppose that the sequence of probability measures $(\mu_n, n \geq 0)$ on M is tight. Then there exists a subsequence $(\mu_{n_k}, k \geq 0)$ along which μ_n converges weakly to some limiting μ.*

The proof is considerably eased when $M = \mathbb{R}$, which we will suppose. For the general case, see Billingsley's book *Convergence of Probability Measures*. Notice that in particular, if $(\mu_n, n \geq 0)$ is a sequence of probability measures on a compact space, then there exists a subsequence μ_{n_k} weakly converging to some μ.

Proof. Let F_n be the distribution function of μ_n. Then it is easy by a diagonal extraction argument to find an extraction $(n_k, k \geq 0)$ and a non-decreasing function $F : \mathbb{Q} \to [0,1]$ such that $F_{n_k}(r) \to F(r)$ as $k \to \infty$ for every rational r. The function F is extended on $\mathbb{R}$ as a càdlàg non-decreasing function by the formula $F(x) = \lim_{r \downarrow x, r \in \mathbb{Q}} F(r)$. It is then elementary by a monotonicity argument to show that $F_{n_k}(x) \to F(x)$ for every x which is a continuity point of F.

To conclude, we must check that F is the distribution function of some measure μ. But the tightness shows that for every $\varepsilon > 0$, there exists $A > 0$ such that $F_n(A) \geq 1 - \varepsilon$ and $F_n(-A) \leq \varepsilon$ for every n. By further choosing A so that F is continuous at A and $-A$, we see that $F(A) \geq 1 - \varepsilon$ and $F(-A) \leq \varepsilon$, whence F has limits 0 and 1 at $-\infty$ and $+\infty$. By a standard corollary of Caratheodory's theorem, there exists a probability measure μ having F as its distribution function. □

Remark. The fact that F_n converges up to extraction to a function F, which need not be a probability distribution function unless the tightness hypothesis is verified, is a particular case of *Helly's theorem*, and says that up to extraction, a family of probability laws μ_n converges *vaguely* to a possibly defective measure μ (i.e. of mass ≤ 1), i.e. $\mu_n(f) \to \mu(f)$

for every f with *compact support*. The problem that could appear is that some of the mass of the μ_n's could 'go to infinity', for example δ_n converges vaguely to the zero measure as $n \to \infty$, and does not converge weakly. This phenomenon of mass 'going away' is exactly what Prokhorov's theorem prevents from happening.

In many situations, showing that a sequence of random variables X_n converges in distribution to a limiting X with law μ is done in two steps. One first shows that the sequence $(\mu_n, n \geq 0)$ of laws of the X_n form a tight sequence. Then, one shows that the limit of μ_n along any subsequence cannot be other than μ. This will be illustrated in the next section.

5.4 Lévy's convergence theorem

In this section, we let d be a positive integer and consider only random variables with values in the states space $\mathbb{R}^d$.

Recall that the *characteristic function* of an $\mathbb{R}^d$-valued random variable X is the function $\Phi_X : \mathbb{R}^d \to \mathbb{C}$ defined by $\Phi_X(\lambda) = E[\exp(\mathrm{i}\langle\lambda, X\rangle)]$. It is a continuous function on $\mathbb{R}^d$, such that $\Phi_X(0) = 1$. Moreover, it induces an injective mapping from (distributions of) random variables to complex-valued functions defined on $\mathbb{R}^d$, in the sense that two random variables with distinct distributions have distinct characteristic functions.

The following theorem is extremely useful in practice.

Theorem 5.4.1 (Lévy's convergence theorem) *Let $(X_n, n \geq 0)$ be a sequence of random variables.*

(i) *If X_n converges in distribution to a random variable X, then $\Phi_{X_n}(\lambda)$ converges to $\Phi_X(\lambda)$ for every $\lambda \in \mathbb{R}^d$.*

(ii) *If $\Phi_{X_n}(\lambda)$ converges to $\Psi(\lambda)$ for every $\lambda \in \mathbb{R}^d$, where Ψ is a function which is continuous at 0, then Ψ is a characteristic function, i.e. there exists a random variable X such that $\Psi = \Phi_X$, and moreover, X_n converges in distribution to X.*

Corollary 5.4.1 *If $(X_n, n \geq 0), X$ are random variables in $\mathbb{R}^d$, then X_n converges in distribution to X as $n \to \infty$ if and only if Φ_{X_n} converges to Φ_X pointwise.*

The proof of (i) in Lévy's theorem is immediate since the function $x \mapsto \exp(\mathrm{i}\lambda \cdot x)$ is continuous and bounded from $\mathbb{R}^d$ to $\mathbb{C}$. For the proof of (ii), we will need to show that the hypotheses imply the tightness of the sequence of laws of $X_n, n \geq 0$. To this end, the following bound is very useful.

Lemma 5.4.1 *Let X be a random variable with values in $\mathbb{R}^d$. Then for any norm $\|\cdot\|$ on $\mathbb{R}^d$ there exists a constant $C > 0$ (depending on d and on the choice of the norm) such that*

$$P(\|X\| \geq K) \leq CK^d \int_{[-K^{-1},K^{-1}]^d} (1 - \Re\Phi_X(u))\mathrm{d}u.$$

Proof. Let μ be the distribution of X. Using Fubini's theorem and a simple recursion, it is easy to check that

$$\frac{1}{\lambda^d}\int_{[-\lambda,\lambda]^d}(1-\Re\Phi_X(u))\mathrm{d}u = 2^d\int_{\mathbb{R}^d}\mu(\mathrm{d}x)\left(1-\prod_{i=1}^d\frac{\sin(\lambda x_i)}{\lambda x_i}\right).$$

Now, the continuous function sinc : $t\in\mathbb{R}\mapsto t^{-1}\sin t$ is such that there exists $0<c<1$ such that $|\mathrm{sinc}\ t|\leq c$ for every $t\geq 1$, so that $f:u\in\mathbb{R}^d\mapsto\prod_{i=1}^d\sin u_i/u_i$ is such that $|f(u)|\leq c$ as soon as $\|u\|_\infty\geq 1$. Therefore, $1-f$ is a non-negative continuous function which is $\geq 1-c$ when $\|u\|_\infty\geq 1$. Letting $C=2^d(1-c)^{-1}$ entails that $C(1-f(u))\geq \mathbb{1}_{\{\|u\|_\infty\geq 1\}}$. Putting things together, one gets the result for the norm $\|\cdot\|_\infty$, and the general result follows from the equivalence of norms in finite-dimensional vector spaces. □

Proof of Lévy's theorem. Suppose Φ_{X_n} converges pointwise to a limit Ψ that is continuous at 0. Then, $|1-\Re\Phi_{X_n}|$ being bounded above by 2, fixing $\varepsilon>0$, the dominated convergence theorem shows that for any $K>0$

$$\lim_n K^d\int_{[-K^{-1},K^{-1}]^d}(1-\Re\Phi_{X_n}(u))\mathrm{d}u = K^d\int_{[-K^{-1},K^{-1}]^d}(1-\Re\Psi(u))\mathrm{d}u.$$

By taking K large enough, we can make this limiting value $<\varepsilon/(2C_d)$, because Ψ is continuous at 0, and it follows by the lemma that for every n large enough, $P(|X_n|\geq K)\leq\varepsilon$. Up to increasing K, this then holds for every n, showing tightness of the family of laws of the X_n. Therefore, up to extracting a subsequence, we see from Prokhorov's theorem that X_n converges in distribution to a limiting X, so that Φ_{X_n} converges pointwise to Φ_X along this subsequence (by part (i)). This is possible only if $\Phi_X=\Psi$, showing that Ψ is a characteristic function. Moreover, this shows that the law of X is the only possible probability measure which is the weak limit of the laws of the X_n along some subsequence, so X_n must converge to X in distribution.

More precisely, if X_n did not converge in distribution to X, we could find a continuous bounded f, some $\varepsilon>0$ and a subsequence X_{n_k} such that for all k,

$$|E[f(X_{n_k})]-E[f(X)]|>\varepsilon \tag{5.2}$$

But since the laws of $(X_{n_k},k\geq 0)$ are tight, we could find a further subsequence along which X_{n_k} converges in distribution to some X', which by (i) would satisfy $\Phi_{X'}=\Psi=\Phi_X$ and thus have same distribution as X, contradicting (5.2). □

Chapter 6

Introduction to Brownian motion

6.1 History up to Wiener's theorem

This chapter is devoted to the construction and some properties of one of probability theory's most fundamental objects. Brownian motion earned its name after R. Brown, who observed around 1827 that tiny particles of pollen in water have an extremely erratic motion. It was observed by Physicists that this was due to a important number of random shocks undertaken by the particles from the (much smaller) water molecules in motion in the liquid. A. Einstein established in 1905 the first mathematical basis for Brownian motion, by showing that it must be an isotropic Gaussian process. The first rigorous mathematical construction of Brownian motion is due to N. Wiener in 1923, using Fourier theory.

In order to motivate the introduction of this object, we first begin by a "microscopical" depiction of Brownian motion. Suppose $(X_n, n \geq 0)$ is a sequence of $\mathbb{R}^d$ valued random variables with mean 0 and covariance matrix $\sigma^2 I_d$, which is the identity matrix in d dimensions, for some $\sigma^2 > 0$. Namely, if $X_1 = (X_1^1, \ldots, X_1^d)$,

$$E[X_1^i] = 0\,, \qquad E[X_1^i X_1^j] = \sigma^2 \delta_{ij}\,, \qquad 1 \leq i, j \leq d.$$

We interpret X_n as the spatial displacement resulting from the shocks due to water molecules during the n-th time interval, and the fact that the covariance matrix is scalar stands for an isotropy assumption (no direction of space is priviledged).

From this, we let $S_n = X_1 + \ldots + X_n$ and we embed this discrete-time process into continuous time by letting

$$B_t^{(n)} = n^{-1/2} S_{[nt]}\,, \qquad t \geq 0.$$

Let $|\cdot|$ be the Euclidean norm on $\mathbb{R}^d$ and for $t > 0$ and $X, y \in \mathbb{R}^d$, define

$$p_t(x) = \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|x|^2}{2t}\right),$$

which is the density of the Gaussian distribution $\mathcal{N}(0, tI_d)$ with mean 0 and covariance matrix tI_d. By convention, the Gaussian law $\mathcal{N}(m, 0)$ is the Dirac mass at m.

Proposition 6.1.1 *Let $0 \leq t_1 < t_2 < \ldots < t_k$. Then the finite marginal distributions of $B^{(n)}$ with respect to times $t_1, \ldots, t_k$ converge weakly as $n \to \infty$. More precisely, if F is a bounded continuous function, and letting $x_0 = 0, t_0 = 0$,*

$$E\left[F(B^{(n)}_{t_1}, \ldots, B^{(n)}_{t_k})\right] \underset{n\to\infty}{\to} \int_{(\mathbb{R}^d)^k} F(x_1, \ldots, x_k) \prod_{1\leq i\leq k} p_{\sigma^2(t_i - t_{i-1})}(x_i - x_{i-1})\mathrm{d}x_i.$$

Otherwise said, $(B^{(n)}_{t_1}, \ldots, B^{(n)}_{t_k})$ converges in distribution to $(G_1, G_2, \ldots, G_k)$, which is a random vector whose law is characterized by the fact that $(G_1, G_2 - G_1, \ldots, G_k - G_{k-1})$ are independent centered Gaussian random variables with respective covariance matrices $\sigma^2(t_i - t_{i-1})I_d$.

Proof. With the notations of the theorem, we first check that $(B^{(n)}_{t_1}, B^{(n)}_{t_2} - B^{(n)}_{t_1}, \ldots, B^{(n)}_{t_k} - B^{(n)}_{t_{k-1}})$ is a sequence of independent random variables. Indeed, one has for $1 \leq i \leq k$,

$$B^{(n)}_{t_i} - B^{(n)}_{t_{i-1}} = \frac{1}{\sqrt{n}} \sum_{j=[nt_{i-1}]+1}^{[nt_i]} X_j,$$

and the independence follows by the fact that $(X_j, j \geq 0)$ is an i.i.d. family. Even better, we have the identity in distribution for the i-th increment

$$B^{(n)}_{t_i} - B^{(n)}_{t_{i-1}} \overset{d}{=} \frac{\sqrt{[nt_i] - [nt_{i-1}]}}{\sqrt{n}} \frac{1}{\sqrt{[nt_i] - [nt_{i-1}]}} \sum_{j=1}^{[nt_i]-[nt_{i-1}]} X_j,$$

and the central limit theorem shows that this converges in distribution to a Gaussian law $\mathcal{N}(0, \sigma^2(t_i - t_{i-1})I_d)$. Summing up our study, and introducing characteristic functions, we have shown that for every $\xi = (\xi_j, 1 \leq j \leq k)$,

$$\begin{aligned} E\left[\exp\left(\mathrm{i}\prod_{j=1}^{k} \xi_j(B^{(n)}_{t_j} - B^{(n)}_{t_{j-1}}\right)\right] &= \prod_{j=1}^{k} E\left[\exp\left(\mathrm{i}\xi_j(B^{(n)}_{t_j} - B^{(n)}_{t_{j-1}}\right)\right] \\ &\underset{n\to\infty}{\to} \prod_{j=1}^{k} E[\exp(\mathrm{i}\xi_j(G_j - G_{j-1})] \\ &= E\left[\exp\left(\mathrm{i}\prod_{j=1}^{k} \xi_i(G_j - G_{j-1})\right)\right], \end{aligned}$$

where $G_1, \ldots, G_k$ is distributed as in the statement of the proposition. By Lévy's convergence theorem we deduce that increments of $B^{(n)}$ between times t_i converge to increments of the sequence G_i, which is easily equivalent to the statement. □

This gives the clue that $B^{(n)}$ should converge to a process B whose increments are independent and Gaussian with covariances dictated by the above formula. This will be set in a rigorous way at the end of this section, with Donsker's invariance theorem.

Definition 6.1.1 *A $\mathbb{R}^d$-valued stochastic process $(B_t, t \geq 0)$ is called a standard Brownian motion if it is a continuous process, that satisfies the following conditions:*

(i) $B_0 = 0$ *a.s.*,

(ii) *for every* $0 = t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_k$, *the increments* $(B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}})$ *are independent, and*

(iii) *for every* $t, s \geq 0$, *the law of* $B_{t+s} - B_t$ *is Gaussian with mean* 0 *and covariance* sI_d.

The term "standard" refers to the fact that B_1 is normalized to have variance I_d, and the choice $B_0 = 0$.

The characteristic properties (i), (ii), (iii) exactly amount to say that the finite-dimensional marginals of a Brownian motion are given by the formula of Proposition 6.1.1. Therefore the law of the Brownian motion is uniquely determined. We now show Wiener's theorem that Brownian motion exists!

Theorem 6.1.1 (Wiener) *There exists a Brownian motion on some probability space.*

Proof. We will first prove the theorem in dimension $d = 1$ and construct a process $(B_t, 0 \leq t \leq 1)$ satisfying the properties of a Brownian motion.

Let $D_0 = \{0, 1\}$, $D_n = \{k2^{-n}, 0 \leq k \leq 2^n\}$ for $n \geq 1$, and $D = \bigcup_{n \geq 0} D_n$ be the set of dyadic rational numbers in $[0, 1]$. On some probability space $(\Omega, \mathcal{F}, P)$, let $(Z_d, d \in D)$ be a collection of independent random variables all having a Gaussian distribution $\mathcal{N}(0, 1)$ with mean 0 and variance 1. We are first going to construct the process $(B_d, d \in D)$ so that B_d is a linear combination of the $Z_{d'}$'s for every d.

It is a well-known and important fact that if random variables $X_1, X_2, \ldots$ are linear combinations of independent centered Gaussian random variables, then $X_1, X_2, \ldots$ are independent if and only if they are pairwise uncorrelated, namely $\mathrm{Cov}\,(X_i, X_j) = E[X_i X_j] = 0$ for every $i \neq j$.

We set $B_0 = 0$ and $B_d = Z_1$. Inductively, given $(B_d, d \in D_{n-1})$, we build $(B_d, d \in D_n)$ in such a way that

- $(B_d, d \in D_n)$ satisfies (i), (ii), (iii) in the definition of the Brownian motion (where the instants under consideration are taken in D_n).
- the random variables $(Z_d, d \in D \setminus D_n)$ are independent of $(B_d, d \in D_n)$.

To this end, take $d \in D_n \setminus D_{n-1}$, and let $d_- = d - 2^{-n}$ and $d_+ = d + 2^{-n}$ so that d_-, d_+ are consecutive dyadic numbers in D_{n-1}. Then write

$$B_d = \frac{B_{d_-} + B_{d_+}}{2} + \frac{Z_d}{2^{(n+1)/2}}\,.$$

Then $B_d - B_{d_-} = (B_{d_+} - B_{d_-})/2 + Z_d/2^{(n+1)/2}$ and $B_{d_+} - B_d = (B_{d_+} - B_{d_-})/2 - Z_d/2^{(n+1)/2}$.

Now notice that $N_d := (B_{d_+} - B_{d_-})/2$ and $N'_d := Z_d/2^{(n+1)/2}$ are by the induction hypothesis two independent centered Gaussian random variables with variance 2^{-n-1}. From this, one deduces $\mathrm{Cov}\,(N_d + N'_d, N_d - N'_d) = \mathrm{Var}\,(N_d) - \mathrm{Var}\,(N'_d) = 0$, so that the increments $B_d - B_{d_-}$ and $B_{d_+} - B_d$ are independent with variance 2^{-n}, as should be. Moreover, these increments are independent of the increments $B_{d'+2^{-n-1}} - B_{d'}$ for $d' \in D_{n-1}, d' \neq d_-$ and of $Z_{d'}, d' \in D_n \setminus D_{n-1}, d' \neq d$ so they are independent of the increments $B_{d''+2^{-n}} - B_{d''}$ for $d'' \in D_n, d'' \notin \{d_-, d\}$. This allows the induction argument to proceed one step further.

Thus, we have a process $(B_d, d \in D)$ satisfying the properties of Brownian motion. Note that $B_t - B - s$ has same

Let $s \leq t \in D$, and notice that for every $p > 0$, since $B_t - B_s$ has same law as $\sqrt{t-s}N$, where N is a standard Gaussian random variable,

$$E[|B_t - B_s|^p] = |t - s|^{p/2} E[|N|^p].$$

Since a Gaussian random variable admits moments of all orders, it follows from Corollary 4.5.1 that $(B_d, d \in D)$ a.s. admits a continuous continuation $(B_t, 0 \leq t \leq 1)$.

Up to modifying B on the exceptional event where such an extension does not exist, replacing it by the 0 function for instance, we see that B can be supposed to be continuous for every ω.

We now check that $(B_t, t \in [0,1])$ thus constructed has the properties of Brownian motion. Let $0 = t_0 < t_1 < \ldots < t_k$, and let $0 = t_0^n < t_1^n < \ldots < t_k^n$ be dyadic numbers such that t_i^n converges to t_i as $n \to \infty$. Then by continuity, $(B_{t_1^n}, \ldots, B_{t_k^n})$ converges a.s. to $(B_{t_1}, \ldots, B_{t_k})$ as $n \to \infty$, while on the other hand, $(B_{t_j^n} - B_{t_{j-1}^n, 1 \leq j \leq k})$ are independent Gaussian random variables with variances $(t_j^n - t_{j-1}^n, 1 \leq j \leq k)$, so it is not difficult using Lvy's theorem to see that this converges in distribution to independent Gaussian random variables with respective variances $t_j - t_{j-1}$, which thus is the distribution of $(B_{t_j} - B_{t_{j-1}}, 1 \leq j \leq k)$, as wanted.

It is now easy to construct a Brownian motion indexed by $\mathbb{R}_+$: simply take independent standard Brownian motions $(B_t^i, 0 \leq t \leq 1), i \geq 0$ as we just constructed, and let

$$B_t = \sum_{i=0}^{\lfloor t \rfloor - 1} B_1^i + B_{t - \lfloor t \rfloor}^{\lfloor t \rfloor}, \qquad t \geq 0.$$

It is easy to check that this has the wanted properties.

Finally, it is straightforward to build a Brownian motion in $\mathbb{R}^d$, by taking d independent copies $B^1, \ldots, B^d$ of B and checking that $((B_t^1, \ldots, B_t^d), t \geq 0)$ is a Brownian motion in $\mathbb{R}^d$. □

Let $\Omega_W = C(\mathbb{R}_+, \mathbb{R}^d)$ be the 'Wiener space' of continuous functions, endowed with the product σ-algebra $\mathcal{W}$ (or the Borel σ-algebra associated with the compact-open topology). Let $X_t(\mathrm{w}) = \mathrm{w}(t), t \geq 0$ denote the *canonical process* $(w \in \Omega_W)$.

Proposition 6.1.2 (Wiener's measure) *There exists a unique measure $W_0(\mathrm{dw})$ on $(\Omega_W, \mathcal{W})$, such that $(X_t, t \geq 0)$ is a standard Brownian motion on $(\Omega_W, \mathcal{W}, W_0(\mathrm{dw}))$.*

Proof. Let $(B_t, t \geq 0)$ be a standard Brownian motion defined on some probability space $(\Omega, \mathcal{F}, P)$ The distribution of B, i.e. the image measure of P by the random variable

$B : \Omega \to \Omega_W$, is a measure $W_0(\mathrm{dw})$ satisfying the conditions of the statement. Uniqueness is obvious because such a measure is determined by the finite-dimensional marginals of Brownian motion. □

For $x \in \mathbb{R}^d$ we also let $W_x(\mathrm{dw})$ to be the image measure of W by $(\mathrm{w}_t, t \geq 0) \mapsto (x + \mathrm{w}_t, t \geq 0)$. A (continuous) process with law $W_x(\mathrm{dw})$ is called a Brownian motion started at x.

We let $(\mathcal{F}_t^B, t \geq 0)$ be the natural filtration of $(B_t, t \geq 0)$.

Notice that Kolmogorov's continuity lemma shows that a standard Brownian motion is also a.s. locally Hölder continuous with any exponent $\alpha < 1/2$, since it is for every $\alpha < 1/2 - 1/p$ for some integer p.

6.2 First properties

The first few following basic (and fundamental) invariance properties of Brownian motion are left as an exercise.

Proposition 6.2.1 *Let B be a standard Brownian motion in $\mathbb{R}^d$.*

1. *If $U \in O(n)$ is an orthogonal matrix, then $UB = (UB_t, t \geq 0)$ is again a Brownian motion. In particular, $-B$ is a Brownian motion.*

2. *If $\lambda > 0$ then $(\lambda^{-1/2} B_{\lambda t}, t \geq 0)$ is a standard Brownian motion (scaling property)*

3. *For every $t \geq 0$, the shifted process $(B_{t+s} - B_t, s \geq 0)$ is a Brownian motion independent of $\mathcal{F}_t^B$ (simple Markov property).*

We now turn to less trivial path properties of Brownian motion. We begin with

Theorem 6.2.1 (Blumenthal's $0-1$ law) *Let B be a standard Brownian motion. The σ-algebra $\mathcal{F}_{0+}^B = \bigcap_{\varepsilon > 0} \mathcal{F}_\varepsilon^B$ is trivial, i.e. constituted of the events of probability 0 or 1.*

Proof. Let $0 < t_1 < t_2 < \ldots < t_k$ and $A \in \mathcal{F}_{0+}^B$. Then if F is continuous bounded function $(\mathbb{R}^d)^k \to \mathbb{R}$, we have by continuity of B and the dominated convergence theorem,

$$E[\mathbb{1}_A F(B_{t_1}, \ldots, B_{t_k})] = \lim_{\varepsilon \downarrow 0} E[\mathbb{1}_A F(B^{(\varepsilon)}_{t_1 - \varepsilon}, \ldots, B^{(\varepsilon)}_{t_k - \varepsilon})],$$

where $B^\varepsilon = (B_{t+\varepsilon - B_\varepsilon}, t \geq 0)$. On the other hand, since A is $\mathcal{F}_\varepsilon^B$-measurable for any $\varepsilon > 0$, the simple Markov property shows that this is equal to

$$P(A) \lim_{\varepsilon \downarrow 0} E[F(B_{t_1 - \varepsilon}, \ldots, B_{t_k - \varepsilon})],$$

which is $P(A)E[F(B_{t_1}, \ldots, B_{t_k})]$, using again dominated convergence and continuity of B and F. This entails that $\mathcal{F}_{0+}$ is independent of $\sigma(B_s, s \geq 0) = \mathcal{F}_\infty^B$. However, $\mathcal{F}_\infty^B$ contains $\mathcal{F}_{0+}^B$, so that the latter σ-algebra is independent of itself, and $P(A) = P(A \cap A) = P(A)^2$, entailing the result. □

Proposition 6.2.2 (i) *For $d = 1$ and $t \geq 0$, let $S_t = \sup_{0\leq s\leq t} B_s$ and $I_t = \inf_{0\leq s\leq t} B_s$ (these are random variables because B is continuous). Then almost-surely, for every $\varepsilon > 0$, one has*

$$S_\varepsilon > 0 \qquad \text{and} \qquad I_\varepsilon < 0.$$

In particular, there exists a zero of B in any interval of the form $(0, \varepsilon)$, $\varepsilon > 0$.

(ii) *A.s.,*

$$\sup_{t\geq 0} B_t = -\inf_{t\geq 0} B_t = +\infty.$$

(iii) *Let C be an open cone in $\mathbb{R}^d$ with non-empty interior and origin at 0, i.e. a set of the form $\{tu : t > 0, u \in A\}$, where A is an non-empty open subset of the unit sphere of $\mathbb{R}^d$. If*

$$H_C = \inf\{t > 0 : B_t \in C\}$$

is the first hitting time *of C, then $H_C = 0$ a.s.*

Proof. (i) The probability that $B_t > 0$ is $1/2$ for every t, so $P(S_t > 0) \geq 1/2$, and therefore if $t_n, n \geq 0$ is any sequence decreasing to 0, $P(\limsup_n\{B_{t_n} > 0\}) \geq \limsup_n P(B_{t_n} > 0) = 1/2$. Since the event $\limsup_n\{B_{t_n} > 0\}$ is in $\mathcal{F}_{0+}$, Blumenthal's law shows that its probability must be 1. The same is true for the infimum by considering the Brownian motion $-B$.

(ii) Let $S_\infty = \sup_{t\geq 0} B_t$. By scaling invariance, for every $\lambda > 0$, $\lambda S_\infty = \sup_{t\geq 0} \lambda B_t$ has same law as $\sup_{t\geq 0} B_{\lambda^2 t} = S_\infty$. This is possible only if either $S_\infty \in \{0, \infty\}$ a.s., however, it cannot be 0 by (i).

(iii) The cone C is invariant by multiplication by a positive scalar, so that $P(B_t \in C)$ is the same as $P(B_1 \in C)$ for every t by the scaling invariance of Brownian motion. Now, if C has nonempty interior, it is straightforward to check that $P(B_1 \in C) > 0$, and one concludes similarly as above. Details are left to the reader. □

6.3 The strong Markov property

We now want to prove an important analog of the simple Markov property, where deterministic times are replaced by stopping times. To begin with, we extend a little the definition of Brownian motion, by allowing it to start from a random location, and by working with filtrations that are larger with the natural filtration of standard Brownian motions.

We say that B is a Brownian motion (started at B_0) if $(B_t - B_0, t \geq 0)$ is a standard Brownian motion which is independent of B_0. Otherwise said, it is the same as the definition as a standard Brownian motion, except that we do not require that $B_0 = 0$. If we want to express this on the Wiener space with the Wiener measure, we have for every measurable functional $F : \Omega_W \to \mathbb{R}_+$,

$$E[F(B_t, t \geq 0)] = E[F(B_t - B_0 + B_0, t \geq 0)],$$

and since $(B_t - B_0)$ has law W_x, this is

$$\int_{\mathbb{R}^d} P(B_0 \in \mathrm{d}x) \int_{\Omega_W} W_0(\mathrm{dw}) F(x + \mathrm{w}(t), t \geq 0) = \int_{\mathbb{R}^d} P(B_0 \in \mathrm{d}x) W_x(F) = E[W_{B_0}(F)],$$

where as above, W_x is the image of W_0 by the translation $\mathrm{w} \mapsto x + \mathrm{w}$, and $W_{B_0}(F)$ is the random variable $\omega \mapsto W_{B_0(\omega)}(F)$. Using Proposition 1.3.4 actually shows that $E[F(B)|B_0] = W_{B_0}(F)$.

Let $(\mathcal{F}_t, t \geq 0)$ be a filtration. We say that a Brownian motion B is an $(\mathcal{F}_t)$-Brownian motion if B is adapted to $(\mathcal{F}_t)$, and if $B^{(t)} = (B_{t+s} - B_t, s \geq 0)$ is independent of $\mathcal{F}_t$ for every $t \geq 0$. For instance, if $(\mathcal{F}_t)$ is the natural filtration of a 2-dimensional Brownian motion $(B^1_t, B^2_t, t \geq 0)$, then $(B^1_t, t \geq 0)$ is an $(\mathcal{F}_t)$-Brownian motion. If B' is a standard Brownian motion and X is a random variable independent of B', then $B = (X + B'_t, t \geq 0)$ is a Brownian motion (started at $B_0 = X$), and it is an $(\mathcal{F}^B_t) = (\sigma(X) \vee \mathcal{F}^{B'}_t)$-Brownian motion. A Brownian motion is always an $(\mathcal{F}^B_t)$-Brownian motion. If B is a standard Brownian motion, then the completed filtration $\mathcal{F}_t = \mathcal{F}^B_t \vee \mathcal{N}$ ($\mathcal{N}$ being the set of events of probability 0) can be shown to be right-continuous, i.e. $\mathcal{F}_{t+} = \mathcal{F}_t$ for every $t \geq 0$, and B is an $(\mathcal{F}_t)$-Brownian motion.

Let $(B_t, t \geq 0)$ be an $(\mathcal{F}_t)$-Brownian motion in $\mathbb{R}^d$ and T be an $(\mathcal{F}_t)$-stopping time. We let $B^{(T)}_t = B_{T+t} - B_T$ for every $t \geq 0$ on the event $\{T < \infty\}$, and 0 otherwise. Then

Theorem 6.3.1 (Strong Markov property) *Conditionally on $\{T < \infty\}$, the process $B^{(T)}$ is a standard Brownian motion, which is independent of $\mathcal{F}_T$. Otherwise said, conditionally given $\mathcal{F}_T$ and $\{T < \infty\}$, the process $(B_{T+t}, t \geq 0)$ is an $(\mathcal{F}_{T+t})$-Brownian motion started at B_T.*

Proof. Suppose first that $T < \infty$ a.s. Let $A \in \mathcal{F}_T$, and consider times $t_1 < t_2 < \ldots < t_k$. We want to show that for every bounded continuous function F on $(\mathbb{R}^d)^k$,

$$E[\mathbb{1}_A F(B^{(T)}_{t_1}, \ldots, B^{(T)}_{t_k})] = P(A) E[F(B_{t_1}, \ldots, B_{t_k})]. \tag{6.1}$$

Indeed, taking $A = \Omega$ entails that $B^{(T)}$ is a Brownian motion, while letting A vary in $\mathcal{F}_T$ entails the independence of $(B^{(T)}_{t_1}, \ldots, B^{(T)}_{t_k})$ and $\mathcal{F}_T$ for every $t_1, \ldots, t_k$, hence of $B^{(T)}$ and $\mathcal{F}_T$.

Now, suppose first that T takes its values in a countable subset E of $\mathbb{R}_+$. Then

$$\begin{aligned} E[\mathbb{1}_A F(B^{(T)}_{t_1}, \ldots, B^{(T)}_{t_k})] &= \sum_{s \in E} E[\mathbb{1}_{A \cap \{T=s\}} F(B^{(s)}_{t_1}, \ldots, B^{(s)}_{t_k})] \\ &= \sum_{s \in E} P(A \cap \{T = s\}) E[F(B_{t_1}, \ldots, B_{t_k})], \end{aligned}$$

where we used the simple Markov property and the fact that $A \cap \{T = s\} \in \mathcal{F}_s$ by definition. Back to the general case, we can apply this result to the stopping time $T_n = 2^{-n}\lceil 2^n T \rceil$. Since $T_n \geq T$, it holds that $\mathcal{F}_T \subset \mathcal{F}_{T_n}$ so that we obtain for $A \in \mathcal{F}_T$

$$E[\mathbb{1}_A F(B^{(T_n)}_{t_1}, \ldots, B^{(T_n)}_{t_k})] = P(A) E[F(B_{t_1}, \ldots, B_{t_n})]. \tag{6.2}$$

Now, by a.s. continuity of B, it holds that $B^{(T_n)}_t$ converges a.s. to $B^{(T)}_t$ as $n \to \infty$, for every $t \geq 0$. Since F is bounded, the dominated convergence theorem allows to pass to the limit in (6.2), obtaining (6.1).

Finally, if $P(T = \infty) > 0$, check that (6.1) remains true when replacing A by $A \cap \{T < \infty\}$, and divide by $P(\{T < \infty\})$. □

An important example of application of the strong Markov property is the so-called reflection principle. Recall that $S_t = \sup_{0\leq s\leq t} B_s$.

Theorem 6.3.2 (Reflection principle) *Let $(B_t, t \geq 0)$ be an $(\mathcal{F}_t)$-Brownian motion started at* 0*, and T be an $(\mathcal{F}_t)$-stopping time. Then, the process*

$$\widetilde{B}_t = B_t \mathbb{1}_{\{t\leq T\}} + (2B_T - B_t)\mathbb{1}_{\{t>T\}}\,, \qquad t \geq 0$$

is also an $(\mathcal{F}_t)$-Brownian motion started at 0.

Proof. By the strong Markov property, the processes $(B_t, 0 \leq t \leq T)$ and $B^{(T)}$ are independent. Moreover, $B^{(T)}$ is a standard Brownian motion, and hence has same law as $-B^{(T)}$. Therefore, the pair $((B_t, 0 \leq t \leq T), B^{(T)})$ has same law as $((B_t, 0 \leq t \leq T), -B^{(T)})$. On the other hand, the trajectory B is a measurable $G((B_t, 0 \leq t \leq T), B^{(T)})$, where $G(X, Y)$ is the concatenation of the paths X, Y. The conclusion follows from the fact that $G((B_t, 0 \leq t \leq T), -B^{(T)}) = \widetilde{B}$. □

Corollary 6.3.1 (Sometimes also called the reflection principle) *Let $0 < b$ and $a \leq b$, then for every $t \geq 0$,*

$$P(S_t \geq b, B_t \leq a) = P(B_t \geq 2b - a).$$

Proof. Let $T_x = \inf\{t \geq 0 : B_t \geq x\}$ be the first entrance time of B_t in $[x, \infty)$ for $x > 0$. Then T_x is an $(\mathcal{F}^B_t)$-stopping time for every x by (i), Proposition 4.1.1. Notice that $T_x < \infty$ a.s. since $S_\infty = \infty$ a.s., where $S_\infty = \lim_{t\to\infty} S_t$.

Now by continuity of B, $B_{T_x} = x$ for every x. By the reflection principle applied to $T = T_b$, we obtain (with the definition $\widetilde{B}$ given in the statement of the reflection principle)

$$P(S_t \geq b, B_t \leq a) = P(T_b \leq t, 2b - B_t \geq 2b - a) = P(T_b \leq t, \widetilde{B}_t \geq 2b - a),$$

since $2b - B_t = \widetilde{B}_t$ as soon as $t \geq T_b$. On the other hand, the event $\{\widetilde{B}_t \geq 2b - a\}$ is contained in $\{T_b \leq t\}$ since $2b - a \geq b$. Therefore, we obtain $P(S_t \geq b, B_t \leq a) = P(\widetilde{B}_t \geq 2b - a)$, and the result follows since $\widetilde{B}$ is a Brownian motion. □

Notice also that the probability under consideration is equal to $P(S_t > b, B_t < a) = P(B_t > 2b - a)$, i.e. the inequalities can be strict or not. Indeed, for the right-hand side, this is due to the fact that the distribution of B_t is non-atomic, and for the left-hand side, this boils down to showing that for every x,

$$T_x = \inf\{t \geq 0 : B_t > x\}\,, \text{ a.s.},$$

which is a straightforward consequence of the strong Markov property at time T_x, combined with Proposition 6.2.2.

Corollary 6.3.2 *The random variable S_t has the same law as $|B_t|$, for every fixed $t \geq 0$. Moreover, for every $x > 0$, the random time T_x has same law as $(x/B_1)^2$.*

Proof. As $a \uparrow b$, the probability $P(S_t \geq b, B_t \leq a)$ converges to $P(S_t \geq b, B_t \leq b)$, and this is equal to $P(B_t \geq b)$ by Corollary 6.3.1. Therefore,

$$P(S_t \geq b) = P(S_t \geq b, B_t \leq b) + P(B_t \geq b) = 2P(B_t \geq b) = P(|B_t| \geq b),$$

because $\{B_t \geq b\} \subset \{S_t \geq b\}$, and this gives the result. We leave the computation of the distribution of T_x as an exercise. □

6.4 Some martingales associated to Brownian motion

One of the nice features of Brownian motion is that there are a tremendous amount of martingales that are associated with it.

Proposition 6.4.1 *Let $(B_t, t \geq 0)$ be an $(\mathcal{F}_t)$-Brownian motion.*

(i) *If $d = 1$ and $B_0 \in L^1$, the process $(B_t, t \geq 0)$ is a $(\mathcal{F}_t)$-martingale.*

(ii) *If $d = 1$ and $B_0 \in L^2$, the process $(B_t^2 - t, t \geq 0)$ is a $(\mathcal{F}_t)$-martingale.*

(iii) *In any dimension, let $u = (u_1, \ldots, u_d) \in \mathbb{C}^d$. If $E[\exp(\langle u, B_0 \rangle)] < \infty$, the process $M = (\exp(\langle u, B_t \rangle - tu^2/2), t \geq 0)$ is also a $(\mathcal{F}_t)$-martingale for every $u \in \mathbb{C}^d$, where u^2 is a notation for $\sum_{i=1}^d u_i^2$.*

Notice that in (iii), we are dealing with $\mathbb{C}$-valued processes. The definition of $E[X|\mathcal{G}]$ the conditional expectation for a random variable $X \in L^1(\mathbb{C})$ is $E[\Re X|\mathcal{G}] + \mathrm{i}E[\Im X|\mathcal{G}]$, and we say that an integrable process $(X_t, t \geq 0)$ with values in $\mathbb{C}$, and adapted to a filtration $(\mathcal{F}_t)$, is a martingale if its real and imaginary parts are. Notice that the hypothesis on B_0 in (iii) is automatically satisfied whenever $u = \mathrm{i}v$ with $v \in \mathbb{R}$ is purely imaginary.

Proof. (i) If $s \leq t$, $E[B_t - B_s|\mathcal{F}_s] = E[B_{t-s}^{(s)}] = 0$, where $B_u^{(s)} = B_{u+s} - B_s$ has mean 0 and is independent of $\mathcal{F}_s$, by the simple Markov property. The integrability of the process is obvious by hypothesis on B_0.

(ii) Integrability is an easy exercise using that $B_t - B_0$ is independent of B_0. We have, for $s \leq t$, $B_t^2 = (B_t - B_s)^2 + 2B_s(B_t - B_s) + B_s^2$. Taking conditional expectation given $\mathcal{F}_s$ and using the simple Markov property gives that $E[B_t^2] = (t - s) + B_s^2$, hence the result.

(iii) Integrability comes from the fact that $E[\exp(\lambda B_t)] = \exp(t\lambda^2/2)$ whenever B is a standard Brownian motion, and the fact that

$$E[\exp(\langle u, B_t \rangle)] = E[\exp(\langle u, (B_t - B_0 + B_0) \rangle)] = E[\exp(\langle u, B_t - B_0 \rangle)]E[\exp(\langle u, B_0 \rangle)] < \infty.$$

For $s \leq t$, $M_t = \exp(\mathrm{i}\langle u, (B_t - B_s) \rangle + \mathrm{i}\langle u, B_s \rangle + t|u|^2/2)$. We use the Markov property again, and the fact that $E[\exp(\mathrm{i}\langle u, B_t - B_s \rangle)] = \exp(-(t - s)|u|^2/2)$, which is the characteristic function of a Gaussian law with mean 0 and variance $|u|^2$. $\square$

From this, one can show that

Proposition 6.4.2 *Let $(B_t, t \geq 0)$ be a standard Brownian motion and $T_x = \inf\{t \geq 0 : B_t = x\}$. Then for $x, y > 0$, one has*

$$P(T_{-y} < T_x) = \frac{x}{x + y}, \quad E[T_x \wedge T_{-y}] = xy.$$

Proposition 6.4.3 *Let $(B_t, t \geq 0)$ be a $(\mathcal{F}_t)$-Brownian motion. Let $f(t, x) : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{C}$ be continuously differentiable in the variable t and twice continuously differentiable in x, and suppose that f and its derivatives of all order are bounded. Then,*

$$M_t = f(t, B_t) - f(0, B_0) - \int_0^t \mathrm{d}s \left(\frac{\partial}{\partial t} + \frac{1}{2}\Delta \right) f(s, B_s), \quad t \geq 0$$

is a $(\mathcal{F}_t)$-martingale, where $\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}$ is the Laplacian operator acting on the spatial coordinate of f.

This is the first symptom of the famous Itō formula, which says what this martingale actually *is*.

Proof. Integrability is trivial from the boundedness of f, as well as adptedness, since M_t is a function of $(B_s, 0 \le s \le t)$. Let $s, t \ge 0$. We estimate

$$E[M_{t+s}|\mathcal{F}_t] = M_t + E\left[f(t+s, B_{t+s}) - f(t, B_t) - \int_t^{t+s} \mathrm{d}u \left(\frac{\partial}{\partial t} + \frac{\Delta}{2} \right) f(u, B_u) \middle| \mathcal{F}_t \right].$$

On the one hand, $E[f(t+s, B_{t+s}) - f(t, B_t)|\mathcal{F}_t] = E[f(t+s, B_{t+s} - B_t + B_t)|\mathcal{F}_t] - f(t, B_t)$, and since $B_{t+s} - B_t$ is independent of $\mathcal{F}_t$ with law $\mathcal{N}(0, s)$, using Proposition 1.3.4, this is equal to

$$\int_{\mathbb{R}} f(t+s, B_t + x) p(s, x) \mathrm{d}x - f(t, B_t), \tag{6.3}$$

where $p(s, x) = (2\pi s)^{-d/2} \exp(-|x|^2/(2s))$ is the probability density function for $\mathcal{N}(0, s)$. On the other hand, if we let $L = \partial/\partial t + \Delta/2$,

$$E\left[\int_t^{t+s} \mathrm{d}u \, Lf(u, B_u) \middle| \mathcal{F}_t \right] \quad = \quad E\left[\int_0^{s} \mathrm{d}u \, Lf(u+t, B_{t+s} - B_t + B_t) \middle| \mathcal{F}_t \right].$$

This expression is of the form $E[F((B^{(t)}, B_t))|\mathcal{F}_t]$, where F is measurable and $B^{(t)}_s = B_{t+s} - B_t, s \ge 0$ is independent of B_t by the simple Markov property, and has law $W_0(\mathrm{dw})$, the Wiener measure. If $(X_t, t \ge 0)$ is the canonical process $X_t(\mathrm{w}) = \mathrm{w}_t$, then this last expression rewrites, by Proposition 1.3.4,

$$\begin{aligned} E\left[\int_t^{t+s} Lf(u, B_u) \mathrm{d}u \middle| \mathcal{F}_t \right] &= \int_{\Omega_W} W_0(\mathrm{dw}) \int_0^s \mathrm{d}u \, Lf(u+t, X_s + B_t) \\ &= \int_0^s \mathrm{d}u \int_{\Omega_W} W_0(\mathrm{dw}) L\widetilde{f}(u, X_s + B_t) \\ &= \int_0^s \mathrm{d}u \int_{\mathbb{R}^d} \mathrm{d}x \, p(s, x) L\widetilde{f}(u, x + B_t), \end{aligned}$$

where $\widetilde{f}(s, x) = f(s+t, x)$, and we made a use of Fubini's theorem. Next, the boundedness of $L\widetilde{f}$ entails that this is equal to

$$\lim_{\varepsilon \downarrow 0} \int_\varepsilon^s \mathrm{d}u \int_{\mathbb{R}^d} \mathrm{d}x \, p(s, x) L\widetilde{f}(u, x + B_t).$$

From the expression for L, we can split this into two parts. Using integration by parts,

$$\begin{aligned} &\int_{\mathbb{R}} \mathrm{d}x \int_\varepsilon^s \mathrm{d}u \, p(u, x) \frac{\partial \widetilde{f}}{\partial t}(u, x + B_t) \\ &= \int_{\mathbb{R}} \mathrm{d}x \, p(s, x) \widetilde{f}(s, x + B_t) - \int_{\mathbb{R}} \mathrm{d}x \, p(\varepsilon, x) \widetilde{f}(\varepsilon, x + B_t) \\ &\quad - \int_{\mathbb{R}^d} \mathrm{d}x \int_\varepsilon^s \mathrm{d}u \, \frac{\partial p}{\partial t}(u, x) \, f(t+u, x + B_t). \end{aligned}$$

Similarly, integrating by parts twice yields

$$\int_\varepsilon^s \mathrm{d}u \int_{\mathbb{R}} \mathrm{d}x\, p(u,x)\, \frac{\Delta}{2} f(u+t, x+B_t) = \int_\varepsilon^s \mathrm{d}u \int_{\mathbb{R}} \mathrm{d}x\, \frac{\Delta}{2} p(u,x)\, f(u+t,x+B_t).$$

Now, $p(t,x)$ satisfies the heat equation $(\partial_t - \Delta/2)p = 0$. Therefore, the integral terms cancel each other, and it remains

$$E\left[\int_t^{t+s} \mathrm{d}u\, Lf(u,B_u)\middle|\mathcal{F}_t\right] = \int_{\mathbb{R}^d} \mathrm{d}x\, p(s,x)\widetilde{f}(s,x+B_t) - \lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}^d} \mathrm{d}x\, p(\varepsilon,x)\widetilde{f}(\varepsilon,x+B_t),$$

which by dominated convergence is exactly (6.3). This shows that $E[M_{t+s} - M_t|\mathcal{F}_t] = 0$. □

6.5 Recurrence and transience properties of Brownian motion

From this section on, we are going to introduce a bit of extra notation. We will suppose that the reference measurable space $(\Omega, \mathcal{F})$ on which $(B_t, t \geq 0)$ is defined endowed with probability measures $P_x, x \in \mathbb{R}^d$ such that under P_x, $(B_t - x, t \geq 0)$ is a standard Brownian motion. A possibility is to choose the Wiener space and endow it with the measures W_x, so that the canonical process $(X_t, t \geq 0)$ is a Brownian motion started at x under W_x. We let E_x be the expectation associated with P_x. In the sequel, $B(x,r), \overline{B}(x,r)$ will respectively denote the open and closed Euclidean balls with center x and radius r, in $\mathbb{R}^d$ for some $d \geq 1$.

Theorem 6.5.1 (i) *If $d = 1$, Brownian motion is point-recurrent in the sense that under P_0 (or any $P_y, y \in \mathbb{R}$),*

$$a.s., \ \{t \geq 0 : B_t = x\} \text{ is unbounded for every } x \in \mathbb{R}.$$

(ii) *If $d = 2$, Brownian motion is neighborhood-recurrent, in the sense that for every x, under P_x,*

$$a.s., \ \{t \geq 0 : |B_t| \leq \varepsilon\} \text{ is unbounded for every } x \in \mathbb{R}^d, \varepsilon > 0.$$

However, points are polar *in the sense that for every $x \in \mathbb{R}^d$,*

$$P_0(H_x = \infty) = 1\,, \ \text{where } H_x = \inf\{t > 0 : B_t = x\}$$

is the hitting time *of x.*

(iii) *If $d \geq 3$, Brownian motion is transient, in the sense that a.s. under P_0, $|B_t| \to \infty$ as $t \to \infty$.*

Proof. (i) is a consequence of (ii) in Proposition 6.2.2.

For (ii), let $0 < \varepsilon < R$ be real numbers and f be a $\mathcal{C}^\infty$ function, which is bounded with all its derivatives, and that coincides with $x \mapsto \log|x|$ on $D_{\varepsilon,R} = \{x \in \mathbb{R}^2 : \varepsilon \leq |x| \leq R\}$.

Then one can check that $\Delta f = 0$ on the interior of $D_{\varepsilon,R}$, and therefore, if we let $S = \inf\{t \geq 0 : |B_t| = \varepsilon\}$ and $T = \inf\{t \geq 0 : |B_t| = R\}$, then $S, T, H = S \wedge T$ are stopping times, and from Proposition 6.4.3, the stopped process $(\log |B_{t\wedge H}|, t \geq 0)$ is a (bounded) martingale. If $\varepsilon < |x| < R$, we thus obtain that $E_x[\log |B_H|] = \log |x|$. Since $H \leq T < \infty$ a.s. (Brownian motion is unbounded a.s.), and since $|B_S| = \varepsilon, |B_T| = R$ on the event that $S < \infty, T < \infty$, the left-hand side is $(\log \varepsilon)P_x(S < T) + (\log R)P_x(S > T)$. Therefore,

$$P_x(S < T) = \frac{\log R - \log |x|}{\log R - \log \varepsilon}. \tag{6.4}$$

Letting $\varepsilon \to 0$ shows that the probability of hitting 0 before hitting the boundary of the ball with radius R is 0, and therefore, letting $R \to \infty$, the probability of hitting 0 (starting from $x \neq 0$) is 0. The announced result (for $x \neq 0$) is then obtained by translation. We thus have that $P_0(H_x < \infty) = 0$ for every $x \neq 0$. Next, we have

$$P_0(\exists t \geq a : B_t = 0) = P_0(\exists s \geq 0 : B_{s+a} - B_a + B_a = 0),$$

and the Markov property at time a shows that this is

$$\begin{aligned} P_0(\exists t \geq a : B_t = 0) &= \int_{\mathbb{R}^2} P_0(B_a \in \mathrm{d}y) P_0(\exists s \geq 0 : B_s + y = 0) \\ &= \int_{\mathbb{R}^2} P_0(B_a \in \mathrm{d}y) P_y(\exists s \geq 0 : B_s = 0) = 0, \end{aligned}$$

because the law of B_a under P_0 is a Gaussian law that does not charge the point 0 (we have been using the notation $P(X \in \mathrm{d}x)$ for the law of the random variable X).

On the other hand, letting $R \to \infty$ first in (6.4), we get that the probability of hitting the ball with center 0 and radius ε is 1 for every ε, starting from any point: $P_x(\exists t \geq 0 : |B_t| \geq \varepsilon) = 1$. Thus, for every $n \in \mathbb{Z}_+$, a similar application of the Markov property at time n gives

$$P_x(\exists t \geq n : |B_t| \leq \varepsilon) = \int_{\mathbb{R}^2} P_x(B_n \in \mathrm{d}y) P_y(\exists t \geq 0 : |B_t| \leq \varepsilon) = 1.$$

Hence the result.

For (iii) Since the three first components of a Brownian motion in $\mathbb{R}^d$ form a Brownian motion, it is clearly sufficient to treat the case $d = 3$. So assume $d = 3$. Let f be a $\mathcal{C}^\infty$ function with all derivatives that are bounded, and that coincides with $x \mapsto 1/|x|$ on $D_{\varepsilon,R}$, which is defined as previously but for $d = 3$. Then $\Delta f = 0$ on the interior of $D_{\varepsilon,R}$, and the same argument as above shows that for $x \in D_{\varepsilon,R}$, defining S, T as above,

$$P_x(S < T) = \frac{|x|^{-1} - R^{-1}}{\varepsilon^{-1} - R^{-1}}.$$

This converges to $\varepsilon/|x|$ as $R \to \infty$, which is thus the probability of ever visiting $B(0, \varepsilon)$ when starting from x (with $|x| \geq \varepsilon$). Define two sequences of stopping times $S_k, T_k, k \geq 1$ by $S_1 = \inf\{t \geq 0 : |B_t| \leq \varepsilon\}$, and

$$T_k = \inf\{t \geq S_k : |B_t| \geq 2\varepsilon\}, \qquad S_{k+1} = \inf\{t \geq S_k : |B_t| \geq 2\varepsilon\}.$$

If S_k is finite, we get that T_k is also finite, because Brownian motion is an a.s. unbounded process, so $\{S_k < \infty\} = \{T_k < \infty\}$ up to a zero-probability event. The strong Markov property at time T_k gives

$$\begin{aligned} P_x(S_{k+1} < \infty | S_k < \infty) &= P_x(S_{k+1} < \infty | T_k < \infty) \\ &= P_x(\exists s \geq T_k : |B_s - B_{T_k} + B_{T_k}| \leq \varepsilon \,|\, T_k < \infty) \\ &= \int_{\mathbb{R}^3} P_x(B_{T_k} \in \mathrm{d}y | T_k < \infty) P_y(\exists s \geq 0 : |B_s| \leq \varepsilon), \end{aligned}$$

where $P_x(B_{T_k} \in \mathrm{d}y | T_k < \infty)$ is the law of B_{T_k} under the probability measure $P_x(A|T_k < \infty), A \in \mathcal{F}$. Since $|B_{T_k}| = 2\varepsilon$ on the event $\{T_k < \infty\}$, we have that the last probability is $\varepsilon/|y| = 1/2$. Finally, we obtain by induction that $P_x(S_k < \infty) \leq P_x(S_1 < \infty)2^{-k+1}$, and the Borel-Cantelli lemma entails that a.s., $S_k = \infty$ for some k. Therefore, Brownian motion in dimension 3 a.s. eventually leaves the ball of radius ε for good, and letting $\varepsilon = n \to \infty$ along $\mathbb{Z}_+$ gives the result. □

Remark. If $B(x, \varepsilon)$ is the Euclidean ball of center x and radius ε, notice that the property of (ii) implies the fact that $\{t \geq 0 : B_t \in B(x, \varepsilon)\}$ is unbounded for every $x \in \mathbb{R}^2$ and every $\varepsilon > 0$, almost surely (indeed, one can cover $\mathbb{R}^2$ by a countable union of balls of a fixed radius). In particular, the trajectory of a 2-dimensional Brownian motion is everywhere dense. On the other hand, it will a.s. never hit a fixed countable family of points (except maybe at time 0), like the points with rational coordinates!

6.6 Brownian motion and the Dirichlet problem

Let D be a connected open subset of $\mathbb{R}^d$ for some $d \geq 2$. We will say that D is a *domain*. Let ∂D be the boundary of D. We denote by Δ the Laplacian on $\mathbb{R}^d$. Suppose given a measurable function $g : \partial D \to \mathbb{R}$. A solution of the *Dirichlet problem* with boundary condition g on D is a function $u : \overline{D} \to \mathbb{R}$ of class $\mathcal{C}^2(D) \cap \mathcal{C}(\overline{D})$, such that

$$\begin{cases} \Delta u = 0 & \text{on } D \\ u|_{\partial D} = g. \end{cases} \tag{6.5}$$

A solution of the Dirichlet problem is the mathematical counterpart of the following physical problem: given an object made of homogeneous material, such that the temperature $g(y)$ is imposed at point y of its boundary, the solution $u(x)$ of the Dirichlet problem gives the temperature at the point x in the object when equilibium is attained.

As we will see, it is possible to give a probabilistic resolution of the Dirichlet problem with the help of Brownian motion. This is essentially due to Kakutani. We let E_x be the law of the Brownian motion in R^d started at x. In the remaining of the section, let $T = \inf\{t \geq 0 : B_t \notin D\}$ be the first exit time from D. It is a stopping time, as it is the first entrance time in the closed set D^c. We will assume that the domain D is such that $P(T < \infty) = 1$ to avoid complications. Hence B_T is a well-defined random variable.

In the sequel, $|\cdot|$ is the euclidean norm on $\mathbb{R}^d$. The goal of this section is to prove the following

Theorem 6.6.1 *Suppose that $g \in \mathcal{C}(\partial D, \mathbb{R})$ is bounded, and assume that D satisfies a local exterior cone condition (l.e.c.c.), i.e. for every $y \in \partial D$, there exists a nonempty open convex cone with origin at y such that $C \cap B(y,r) \subset D^c$ for some $r > 0$. Then the function*

$$u : x \mapsto E_x\left[g(B_T)\right]$$

is the unique bounded solution of the Dirichlet problem (6.5).

In particular, if D is bounded and satisfies the l.e.c.c., then u is the unique solution of the Dirichlet problem.

We start with a uniqueness statement.

Proposition 6.6.1 *Let g be a bounded function in $\mathcal{C}(\partial D, \mathbb{R})$. Set*

$$u(x) = E_x\left[g(B_T)\right].$$

If v is a bounded solution of the Dirichlet problem, then $v = u$.

In particular, we obtain uniqueness when D is bounded. Notice that we do not make any assumption on the regularity of D here besides the fact that $T < \infty$ a.s.

Proof. Let v be a bounded solution of the Dirichlet problem. For every $N \geq 1$, introduce the reduced set $D_N = \{x \in D : |x| < N \text{ and } d(x, \partial D) > 1/N\}$. Notice it is an open set, but which need not be connected. We let T_N be the first exit time of D_N. By Proposition 6.4.3, the process

$$M_t = \widetilde{v}_N(B_t) - \widetilde{v}_N(B_0) + \int_0^t -\frac{1}{2}\Delta\widetilde{v}_N(B_s)\mathrm{d}s\,, \qquad t \geq 0$$

is a martingale, where $\widetilde{v}_N$ is a $\mathcal{C}^2$ function that coincides with v on D_N, and which is bounded with all its partial derivatives (this may look innocent, but the fact that such a function exists is highly non-trivial, the use of such a function could be avoided by a stopped analog of Proposition 6.4.3). Moreover, the martingale stopped at T_N is $M_{t\wedge T_N} = v(B_{t\wedge T_N}) - v(B_0)$, because $\Delta v = 0$ inside D, and it is bounded (because v is bounded on D_N), hence uniformly integrable. By optional stopping at T_N, we get that for every $x \in D_N$,

$$0 = E_x[M_{T_N}] = E_x[v(B_{T_N})] - v(x) \tag{6.6}$$

Now, as $N \to \infty$, B_{T_N} converges to B_T a.s. by continuity of paths and the fact that $T < \infty$ a.s. Since v is bounded, we can use dominated convergence as $N \to \infty$, and get that for every $x \in D$,

$$v(x) = E_x[v(B_T)] = E_x[g(B_T)],$$

hence the result. □

For every $x \in \mathbb{R}^d$ and $r > 0$, let $\sigma_{x,r}$ be the uniform probability measure on the sphere $\mathbb{S}_{x,r} = \{y \in \mathbb{R}^d : |y - x| = r\}$. It is the unique probability measure on $\mathbb{S}_{x,r}$ that is invariant under isometries of $\mathbb{S}_{x,r}$. We say that a locally bounded measurable function $h : D \to \mathbb{R}$ is *harmonic* on D if for every $x \in D$ and every $r > 0$ such that the closed ball $\overline{B}(x,r)$ with center x and radius r is contained in D,

$$h(x) = \int_{\mathbb{S}_{x,r}} h(y)\sigma_{x,r}(\mathrm{d}y).$$

Proposition 6.6.2 *Let h be harmonic on a domain D. Then $h \in \mathcal{C}^\infty(D, \mathbb{R})$, and $\Delta h = 0$ on D.*

Proof. Let $x \in D$ and $\varepsilon > 0$ such that $\overline{B}(x, \varepsilon) \subset D$. Then let $\varphi \in \mathcal{C}^\infty(D, \mathbb{R})$ be non-negative with non-empty compact support in $[0, \varepsilon[$. We have, for $0 < r < \varepsilon$,

$$h(x) = \int_{\mathbb{S}(0,r)} h(x + y)\sigma_{0,r}(\mathrm{d}y).$$

Multiplying by $\varphi(r)r^{d-1}$ and integrating in r gives

$$ch(x) = \int_{\mathbb{R}^d} \varphi(|z|)h(x + z)\mathrm{d}z,$$

where $c > 0$ is some constant, where we have used the fact that

$$\int_{\mathbb{R}^d} f(x)\mathrm{d}x = C \int_{\mathbb{R}_+} r^{d-1}\mathrm{d}r \int_{\mathbb{S}(0,r)} f(ry)\sigma_{0,r}(\mathrm{d}y)$$

for some $C > 0$. Therefore, $ch(x) = \int_{\mathbb{R}^d} \varphi(|z - x|)h(z)\mathrm{d}z$ and by derivation under the $\int$ sign, we easily get that h is $\mathcal{C}^\infty$.

Next, by translation we may suppose that $0 \in D$ and show only that $\Delta h(0) = 0$. we may apply Taylor's formula to h, obtaining, as $x \to 0$,

$$h(x) = h(0) + \langle \nabla h(0), x\rangle + \sum_{i=1}^{d} x_i^2 \frac{\partial^2 h}{\partial x_i^2}(0) + \sum_{i \neq j} x_i x_j \frac{\partial^2 h}{\partial x_i \partial x_j}(0) + o(|x|^2).$$

Now, integration over $\mathbb{S}_{0,r}$ for r small enough yields

$$\int_{\mathbb{S}_{x,r}} h(x)\sigma_{0,r}(\mathrm{d}x) = h(0) + C_r \Delta h(0) + o(|r|^2),$$

where $C_r = \int_{\mathbb{S}_{0,r}} x_1^2 \sigma_{0,r}(\mathrm{d}x)$, as the reader may check that all the other integrals up to the second order are 0, by symmetry. since the left-hand side is $h(0)$, we obtain $\Delta h(0) = 0$. $\square$

Therefore, harmonic functions are solutions of certain Dirichlet problems.

Proposition 6.6.3 *Let g be a bounded measurable function on ∂D, and let $T = \inf\{t \geq 0 : B_t \notin D\}$. Then the function $h : x \in D \mapsto E_x[g(B_T)]$ is harmonic on D, and hence $\Delta h = 0$ on D.*

Proof. For every Borel subsets $A_1, \ldots, A_k$ of $\mathbb{R}^d$ and times $t_1 < \ldots < t_k$, the map

$$x \mapsto P_x(B_{t_1} \in A_1, \ldots, B_{t_n} \in A_n)$$

is measurable by Fubini's theorem, once one has written the explicit formula for this probability. Therefore, by the monotone class theorem, $x \mapsto E_x[F]$ is measurable for every integrable random variable F, which is measurable with respect to the product σ-algebra on $\mathcal{C}(\mathbb{R},\mathbb{R}^d)$. Moreover, h is bounded by assumption.

Now, let $S = \inf\{t \geq 0 : |B_t - x| \geq r\}$ the first exit time of B form the ball of center x and radius r. Then by (ii), Proposition 6.2.2, $S < \infty$ a.s. By the strong Markov property, $\widetilde{B} = (B_{S+t}, t \geq 0)$ is an $(\mathcal{F}_{S+t})$ Brownian motion started at B_S. Moreover, the first hitting time of ∂D for $\widetilde{B}$ is $\widetilde{T} = T - S$. Moreover, $\widetilde{B}_{\widetilde{T}} = B_T$, so that

$$E_x[g(B_T)] = E_x[g(\widetilde{B}_{\widetilde{T}})] = \int_{\mathbb{R}^d} P_x(B_S \in \mathrm{d}y) E_y[g(B_T)\mathbb{1}_{\{T<\infty\}}],$$

and we recognize $\int P_x(B_S \in \mathrm{d}y)h(y)$ in the last expression.

Since B starts from x under P_x, the rotation invariance of Brownian motion shows that $B_S - x$ has a distribution on the sphere of center 0 and radius r which is invariant under the orthogonal group, so we conclude that the distribution of B_S is the uniform measure on the sphere of center x and radius r, and therefore that h is harmonic on D. $\square$

It remains to understand whether the function u of Theorem 6.6.1 is actually a solution of the Dirichlet problem. Indeed, is not the case in general that $u(x)$ has limit $g(y)$ as $x \in D, x \to y$, and the reason is that some points of ∂D may be 'invisible' to Brownian motion. The reader can convince himself, for example, that if $D = B(0,1)\backslash\{0\}$ is the open ball of $\mathbb{R}^2$ with center 0 and radius 1, whose origin has been removed, and if $g = \mathbb{1}_{\{0\}}$, then no solution of the Dirichlet problem with boundary constraint g exists. The probabilistic reason for that is that Brownian motion does not see the boundary point 0. This is the reason why we have to make regularity assumptions on D in the following theorem.

Proof of Theorem 6.6.1.

It remains to prove that under the l.e.c.c., u is continuous on $\overline{D}$, i.e. $u(x)$ converges to $g(y)$ as $x \in D$ converges to $y \in \partial D$. In order to do that, we need a preliminary lemma. Recall that T is the first exit time of D for the Brownian path.

Lemma 6.6.1 *Let D be a domain satisfying the l.e.c.c., and let $y \in \partial D$. Then for every $\eta > 0$, $P_x(T < \eta) \to 1$ as $x \in D \to y$.*

Proof. Let $C_y = y + C$ be a nonempty open convex cone with origin at y such that for some $\eta > 0$, $C_y \subset D^c$ (we leave as an exercise the case when only a neighborhood of this cone around y is contained in D^c). Then it is an elementary geometrical fact that for every $\eta > 0$ small enough, there exist $\delta > 0$ and a nonempty open convex cone C' with origin at 0, such that $x + (C' \setminus \overline{B}(0,\eta)) \subseteq C_y$ for every $x \in B(y,\delta)$. Now by (iii) in Proposition 6.2.2, if $H^\varepsilon_{C'} = \inf\{t > 0 : B_t \in C' \setminus \overline{B}(0,\varepsilon)\}$, then $P_0(H^\varepsilon_{C'} < \eta) \nearrow P_0(H_{C'} < \eta) = 1$ as $\eta \downarrow 0$.

Since hitting $x + (C' \setminus \overline{B}(0,\eta))$ implies hitting C_y and therefore leaving D, we obtain, after translating by x, that for every $\eta, \varepsilon' > 0$, $P_x(T < \eta)$ can be made $\geq 1 - \varepsilon'$ for x belonging to a sufficiently small δ-neighborhood of y in D. $\square$

We can now finish the proof of Theorem 6.6.1. Let $y \in \partial D$. We want to estimate the quantity $E_x[g(B_T)] - g(y)$ for some $y \in \partial D$. For $\eta, \delta > 0$, let

$$A_{\eta,\delta} = \left\{ \sup_{0 \leq t \leq \eta} |B_t - x| \geq \delta/2 \right\}.$$

This event decreases to $\varnothing$ as $\eta \downarrow 0$ because B has continuous paths. Now, for any $\delta, \eta > 0$,

$$\begin{aligned} E_x[|g(B_T) - g(y)|] &= E_x[|g(B_T) - g(y)| \,;\, \{T \le \eta\} \cap A^c_{\delta,\eta}] \\ &\quad + E_x[|g(B_T) - g(y)| \,;\, \{T \le \eta\} \cap A_{\delta,\eta}] \\ &\quad + E_x[|g(B_T) - g(y)| \,;\, \{T \ge \eta\}] \end{aligned}$$

Fix $\varepsilon > 0$. We are going to show that each of these three quantities can be made $< \varepsilon/3$ for x close enough to y. Since g is continuous at y, for some $\delta > 0$, $|y - z| < \delta$ with $y, z \in \partial D$ implies $|g(y) - g(z)| < \varepsilon/3$. Moreover, on the event $\{T \le \eta\} \cap A^c_{\delta,\eta}$, we know that $|B_T - x| < \delta/2$, and thus $|B_T - y| \le \delta$ as soon as $|x - y| \le \delta/2$. Therefore, for every $\eta > 0$, the first quantity is less than $\varepsilon/3$ for $x \in \overline{B}(y, \delta/2)$.

Next, if M is an upper bound for $|g|$, the second quantity is bounded by $2MP(A_{\delta,\eta})$. Hence, by now choosing η small enough, this is $< \varepsilon/3$.

Finally, with δ, η fixed as above, the third quantity is bounded by $2MP_x(T \ge \eta)$. By the previous lemma, this is $< \varepsilon/3$ as soon as $x \in B(y, \alpha) \cap D$ for some $\alpha > 0$. Therefore, for any $x \in B(y, \alpha \wedge \delta/2) \cap D$, $|u(x) - g(y)| < \varepsilon$. This entails the result. □

Corollary 6.6.1 *A function $u : D \to \mathbb{R}$ is harmonic in D if and only if it is in $\mathcal{C}^2(D, \mathbb{R})$, and satisfies $\Delta u = 0$.*

Proof. Let u be of class $\mathcal{C}^2(D)$ be of zero Laplacian, and let $x \in D$. Let ε be such that $B(x, \varepsilon) \subseteq D$, and notice that $u|_{\overline{B}(x,\varepsilon)}$ is a solution of the Dirichlet problem on $\overline{B}(x, \varepsilon)$ with boundary values $u|_{\partial B(x,\varepsilon)}$. Then $B(x, \varepsilon)$ satisfies the l.e.c.c., so that $u|_{B(x,\varepsilon)}$ is the unique such solution, which is also given by the harmonic function of Theorem 6.6.1. Therefore, u is harmonic on D. □

6.7 Donsker's invariance principle

The following theorem completes the description of Brownian motion as a 'limit' of centered random walks as depicted in the beginning of the chapter, and strengthen the convergence of finite-dimensional marginals to that convergence in distribution. We endow $\mathcal{C}([0, 1], \mathbb{R})$ with the supremum norm, and recall (see the exercises on continuous-time processes) that the product σ-algebra associated with it coincides with the Borel σ-algebra associated with this norm. We say that a function $F : \mathcal{C}([0, 1]) \to \mathbb{R}$ is continuous if it is continuous with respect to this norm.

Theorem 6.7.1 (Donsker's invariance principle) *Let $(X_n, n \ge 1)$ be a sequence of $\mathbb{R}$-valued integrable independent random variables with common law μ, such that*

$$\int x\mu(\mathrm{d}x) = 0 \ \ \textit{and} \ \int x^2\mu(\mathrm{d}x) = \sigma^2 \in (0, \infty).$$

Let $S_0 = 0$ and $S_n = X_1 + \ldots + X_n$, and define a continuous process that interpolates linearly between values of S, namely

$$S_t = (1 - \{t\})S_{[t]} + \{t\}S_{[t]+1} \qquad t \ge 0,$$

where $[t]$ *denotes the integer part of* t *and* $\{t\} = t - [t]$*. Then* $S^{[N]} := ((\sigma^2 N)^{-1/2} S_{Nt}, 0 \leq t \leq 1)$ *converges in distribution to a standard Brownian motion between times* 0 *and* 1*, i.e. for every bounded continuous function* $F : \mathcal{C}([0,1]) \to \mathbb{R}$*,*

$$E\left[F(S^{[N]})\right] \underset{n\to\infty}{\to} E_0[F(B)].$$

Notice that this is much stronger than what Proposition 6.1.1 says. Despite the slight difference of framework between these two results (one uses càdlàg continuous-time version of the random walk, and the other uses an interpolated continuous version), Donsker's invariance principle is stronger. For instance, one can infer from this theorem that the random variable $N^{-1/2} \sup_{0\leq n\leq N} S_n$ converges to $\sup_{0\leq t\leq 1} B_t$ in distribution, because $f \mapsto \sup f$ is a continuous operation on $\mathcal{C}([0,1],\mathbb{R})$. Proposition 6.1.1 would be powerless to address this issue.

The proof we give here is an elegant demonstration that makes use of a coupling of the random walk with the Brownian motion, called the Skorokhod embedding theorem. It is however specific to dimension $d = 1$. Suppose we are given a Brownian motion $(B_t, t \geq 0)$ on some probability space $(\Omega, \mathcal{F}, P)$.

Let $\mu_+(\mathrm{d}x) = P(X_1 \in \mathrm{d}x)\mathbb{1}_{\{x\geq 0\}}, \mu_-(\mathrm{d}y) = P(-X_1 \in \mathrm{d}y)\mathbb{1}_{\{y\geq 0\}}$ define two non-negative measures. Assume that $(\Omega, \mathcal{F}, P)$ is a rich enough probability space so that we can further define on it, independently of $(B_t, t \geq 0)$, a sequence of independent identically distributed $\mathbb{R}^2$-valued random variables $((Y_n, Z_n), n \geq 1)$ with distribution

$$P((Y_n, Z_n) \in \mathrm{d}x\mathrm{d}y) = C(x + y)\mu_+(\mathrm{d}x)\mu_-(\mathrm{d}y),$$

where $C > 0$ is the appropriate normalizing constant that makes this expression a probability measure.

Next, consider let $\mathcal{F}_0 = \sigma\{(Y_n, Z_n), n \geq 1\}$ and $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{F}^B_t$, so that $(\mathcal{F}_t, t \geq 0)$ is a filtration such that B is an $(\mathcal{F}_t)$-Brownian motion. We define a sequence of random times, by $T_0 = 0, T_1 = \inf\{t \geq 0 : B_t \in \{Y_1, -Z_1\}\}$, and recursively,

$$T_n = \inf\{t \geq T_{n-1} : B_t - B_{T_{n-1}} \in \{Y_n, -Z_n\}\}.$$

By (ii) in Proposition 6.2.2, these times are a.s. finite, and they are stopping times with respect to the filtration $(\mathcal{F}_t)$. We claim that

Lemma 6.7.1 *The sequence* $(B_{T_n}, n \geq 0)$ *has the same law as* $(S_n, n \geq 0)$*. Moreover, the intertimes* $(T_n - T_{n-1}, n \geq 0)$ *form an independent sequence of random variables with same distribution, and expectation* $E[T_1] = \sigma^2$*.*

Proof. By repeated application of the Markov property at times $T_n, n \geq 1$, and the fact that the $(Y_n, Z_n), n \geq 1$ are independent with same distribution, we obtain that the processes $(B_{t+T_{n-1}} - B_{T_{n-1}}, 0 \leq t \leq T_n - T_{n-1})$ are independent with the same distribution. The fact that the differences $B_{T_n} - B_{T_{n-1}}, n \geq 1$ and $T_n - T_{n-1}, n \geq 0$ form sequences of independent and identicaly distributed random variables follows from this observation.

It therefore remains to check that B_{T_1} has same law as X_1 and $E[T_1] = \sigma^2$. Remember from Proposition 6.4.2 that given Y_1, Z_1, the probability that $B_{T_1} = Y_1$ is $Z_1/(Y_1 + Z_1)$, as follows from the optional stopping theorem. Therefore, for every non-negative measurable function f, by first conditioning on (Y_1, Z_1), we get

$$\begin{aligned} E[f(B_{T_1})] &= E\left[f(Y_1)\frac{Z_1}{Y_1+Z_1} + f(-Z_1)\frac{Y_1}{Y_1+Z_1}\right] \\ &= \int_{\mathbb{R}_+\times\mathbb{R}_+} C(x+y)\mu_+(\mathrm{d}x)\mu_-(\mathrm{d}y)\left(f(x)\frac{y}{x+y} + f(-y)\frac{x}{x+y}\right) \\ &= C'\int_{\mathbb{R}_+} (f(x)\mu_+(\mathrm{d}x) + f(-x)\mu_-(\mathrm{d}x)) = C'E[f(X_1)], \end{aligned}$$

for $C' = C\int x\mu_+(\mathrm{d}x)$ which can only be $= 1$ (by taking $f = 1$). Here, we have used the fact that $\int x\mu_+(\mathrm{d}x) = \int x\mu_-(\mathrm{d}x)$, which amounts to say that X_1 is centered.

For $E[T_1]$, recall from Propostition 6.4.2 that $E[\inf\{t \geq 0 : B_t \in \{x, -y\}\}] = xy$, so by a similar conditioning argument as above,

$$E[T_1] = \int_{\mathbb{R}_+\times\mathbb{R}_+} C(x+y)xy\mu_+(\mathrm{d}x)\mu_-(\mathrm{d}y) = \sigma^2,$$

where we again used that $C\int x\mu_+(\mathrm{d}x) = 1$. □

Proof of Donsker's invariance principle. We suppose given a Brownian motion B. For $N \geq 1$, define $B^{(N)}_t = N^{1/2}B_{N^{-1}t}, t \geq 0$, which is a Brownian motion by scaling invariance. Perform the Skorokhod embedding construction on $B^{(N)}$ to obtain variables $T^{(N)}_n, n \geq 0$. Then, let $S^{(N)}_n = B^{(N)}_{T^{(N)}_n}$. Then by Lemma 6.7.1, $S^{(N)}_n, n \geq 0$ is a random walk with same law as $S_n, n \geq 0$. We interpolate linearly between integers to obtain a continuous process $S^{(N)}_t, t \geq 0$. Finally, let $\widetilde{S}^{(N)}_t = (\sigma^2 N)^{-1/2}S^{(N)}_{Nt}, t \geq 0$ and $\widetilde{T}^{(N)}_n = N^{-1}T^{(N)}_n$.

We are going to show that the supremum norm of $B_t - \widetilde{S}^{(N)}_t, 0 \leq t \leq 1$ converges to 0 in probability.

By the law of large numbers, T_n/n converges a.s. to σ^2 as $n \to \infty$. Thus, by a monotonicity argument, $N^{-1}\sup_{0\leq n\leq N}|T_n - \sigma^2 n|$ converges to 0 a.s. as $N \to \infty$. As a consequence, this supremum converges to 0 in probability, meaning that for every $\delta > 0$,

$$P\left(\sup_{0\leq n\leq N}|\widetilde{T}^{(N)}_n - n/N| \geq \delta\right) \underset{N\to\infty}{\to} 0.$$

On the other hand, for every $t \in [n/N, (n+1)/N]$, there exists some $u \in [\widetilde{T}^{(N)}_n, \widetilde{T}^{(N)}_{n+1}]$ with $B_u = \widetilde{S}^{(N)}_t$, because $\widetilde{S}^{(N)}_{n/N} = B_{\widetilde{T}^{(N)}_n}$ for every n and by the intermediate values theorem, $\widetilde{S}^{(N)}$ and B being continuous. Therefore, the event $\{\sup_{0\leq t\leq 1}|\widetilde{S}^{(N)}_t - B_t| > \varepsilon\}$ is contained in the union $K^N \cup L^N$, where

$$K^N = \left\{\sup_{0\leq n\leq N}|\widetilde{T}^{(N)}_n - n/N| > \delta\right\}$$

and

$$L^N = \{\exists\, t \in [0,1],\ \exists\, u \in [t-\delta-1/N, t+\delta+1/N] : |B_t - B_u| > \varepsilon\}.$$

We already know that $P(K^N) \to 0$ as $N \to \infty$. For L^N, since B is a.s. uniformly continuous on $[0,1]$, by taking δ small enough and then N large enough, we can make $P(L^N)$ as small as wanted. Therefore, we have showed that

$$P\left(\|\widetilde{S}^{(N)} - B\|_\infty > \varepsilon\right) \underset{n\to\infty}{\to} 0.$$

Therefore, $(\widetilde{S}^{(N)}, 0 \le t \le 1)$ converges in probability for the uniform norm to $(B_t, 0 \le t \le 1)$, which entails convergence in distribution by Proposition 5.2.1. This concludes the proof. □

Chapter 7

Poisson random measures and Poisson processes

7.1 Poisson random measures

Let $(E, \mathcal{E})$ be a measurable space, and let μ be a non-negative σ-finite measure on $(E, \mathcal{E})$. We denote by E^* the set of σ-finite atomic measures on E, i.e. the set of σ-finite measures taking values in $\mathbb{Z}_+ \sqcup \{\infty\}$ (in fact, we will only consider measures that can be put in the form $\sum_{i \in I} \delta_{x_i}$ with I countable and $x_i \in E, i \in I$). The set E^* is endowed with the product σ-algebra $\mathcal{E}^* = \sigma(X_A, A \in \mathcal{E})$, where $X_A(m) = m(A)$ for $m \in E^*$, and $A \in \mathcal{E}$. Otherwise said, for every $A \in \mathcal{E}$, the mapping $m \mapsto m(A)$ from E^* to $\mathbb{Z}_+ \sqcup \{\infty\}$ is measurable with respect to $\mathcal{E}^*$. For $\lambda > 0$ we denote by $\mathcal{P}(\lambda)$ the Poisson distribution with parameter λ, which assigns mass $e^{-\lambda}\lambda^n/n!$ to the integer n.

Definition 7.1.1 *A Poisson measure on $(E, \mathcal{E})$ with intensity μ is a random variable M with values in E^*, such that if $(A_k, k \geq 1)$ is a sequence of disjoint sets in $\mathcal{E}$, with $\mu(A_k) < \infty$ for every k,*

(i) *the random variables $M(A_k), k \geq 1$ are independent, and*

(ii) *the law of $M(A_k)$ is $\mathcal{P}(\mu(A_k))$ for $k \geq 1$.*

Notice that properties (i) and (ii) completely characterize the law of the random variable M. Indeed, notice that events which are either empty or of the form

$$\{m \in E^* : m(A_1) = i_1, \ldots, m(A_k) = i_k\},$$

with pairwise disjoint $A_1, \ldots, A_k \in \mathcal{E}, \mu(A_j) < \infty, 1 \leq j \leq k$ and where $(i_1, \ldots, i_k)$ are integers, form a π-system that generates $\mathcal{E}^*$. If now M is a Poisson random measure with intensity μ, on some probability space $(\Omega, \mathcal{F}, P)$, then

$$P(M(A_1) = i_1, \ldots, M(A_k) = i_k) = \prod_{j=1}^{k} e^{-\mu(A_j)} \frac{\mu(A_j)^{i_j}}{n_j!}.$$

Hence the uniqueness of the law of a random measure satisfying (i), (ii). Existence is stated in the next

Proposition 7.1.1 *For every σ-finite non-negative measure μ on $(E,\mathcal{E})$, there exists a Poisson random measure on $(E,\mathcal{E})$ with intensity μ.*

Proof. Suppose first that $\lambda = \mu(E) < \infty$. We let N be a Poisson random variable with parameter λ, and $X_1, \dots$ be independent random variables, independent of N, with law $\mu/\mu(E)$. Finally, we let $M_\omega = \sum_{i=1}^{N(\omega)} \delta_{X_i(\omega)}$.

Now, if N is Poisson with parameter λ and $(Y_i, i \geq 1)$ are independent and independent of N, with $P(Y_i = j) = p_j, 1 \leq j \leq k$, it holds that $\sum_{i=1}^N \mathbb{1}_{\{Y_i=j\}}, 1 \leq j \leq k$ are independent with respective laws $\mathcal{P}(p_j\lambda), 1 \leq j \leq k$. It follows that M is a Poisson measure with intensity μ: for disjoint $A_1, \dots, A_k$ in $\mathcal{E}$ with finite μ-measures, we let $Y_i = j$ whenever $X_i \in A_j$, defining independent random variables in $\{1, \dots, k\}$ with $P(Y_i = j) = \mu(A_j)/\mu(E)$, so that $M(A_j), 1 \leq j \leq k$ are independent $\mathcal{P}(\mu(E)\mu(A_j)/\mu(E)), 1 \leq j \leq k$ random variables.

In the general case, since μ is σ-finite, there is a partition of E into measurable sets $E_k, k \geq 1$ that are disjoint and have finite μ-measure. We can construct independent Poisson measures M_k on E_k with intensity $\mu(\cdot \cap E_k)$, for $k \geq 1$. We claim that

$$M(A) = \sum_{k \geq 1} M_k(A \cap E_k), \qquad A \in \mathcal{E},$$

defines a Poisson random measure with intensity μ. This is an easy consequence of the property that if $Z_1, Z_2, \dots$ are independent Poisson variables with respective parameters $\lambda_1, \lambda_2, \dots$, then the sum $Z_1 + Z_2 + \dots$ is Poisson with parameter $\lambda_1 + \lambda_2 + \dots$ (with the convention that $\mathcal{P}(\infty)$ is a Dirac mass at ∞). □

From the construction, we obtain the following important property of Poisson random measures:

Proposition 7.1.2 *Let M be a Poisson random measure on E with intensity μ, and let $A \in \mathcal{E}$ be such that $\mu(A) < \infty$. Then $M(A)$ has law $\mathcal{P}(\mu(A))$, and given $M(A) = k$, the restriction $M|_A$ has same law as $\sum_{i=1}^k \delta_{X_i}$, where $(X_1, X_2, \dots, X_k)$ are independent with law $\mu(\cdot \cap A)/\mu(A)$. Moreover, if $A, B \in \mathcal{E}$ are disjoint, then the restrictions $M|_A, M|_B$ are independent. Last, any Poisson random measure can be written in the form $M(\mathrm{d}x) = \sum_{i \in I} \delta_{x_i}(\mathrm{d}x)$ where I is a countable index-set and the $x_i, i \in I$ are random variables.*

7.2 Integrals with respect to a Poisson measure

Proposition 7.2.1 *Let M be a Poisson random measure on E, with intensity μ. Then for every measurable $f : E \to \mathbb{R}_+$, the quantity*

$$M(f) := \int_E f(x) M(\mathrm{d}x)$$

defines a random variable, and

$$E[\exp(-M(f))] = \exp\left(-\int_E \mu(\mathrm{d}x)(1 - \exp(-f(x)))\right).$$

Moreover, if $f : E \to \mathbb{R}$ is measurable and in $L^1(\mu)$, then $f \in L^1(M)$ a.s., $\int_E f(x)M(\mathrm{d}x)$ defines a random variable, and

$$E[\exp(\mathrm{i}M(f))] = \exp\left(\int_E \mu(\mathrm{d}x)(\exp(\mathrm{i}f(x)) - 1)\right).$$

The first formula is sometimes called the *Laplace functional* formula, or *Campbell* formula. Notice that by replacing f by af, differentiating the formula with respect to a and letting $a \downarrow 0$, one gets the *first moment* formula

$$E[M(f)] = \int_E f(x)\mu(\mathrm{d}x),$$

whenever $f \geq 0$, or f is integrable w.r.t. μ (in this case, consider first f^+, f^-). Similarly,

$$\mathrm{Var}\, M(f) = \int_E f(x)^2 \mu(\mathrm{d}x)$$

(for this, first notice that the restrictions of M to $\{f \geq 0\}$ and $\{f < 0\}$ are independent).

Proof. Let $E_n, n \geq 0$ be a measurable partition of E into sets with finite μ-measure. First assume that $f = \mathbb{1}_A$ for $A \in \mathcal{E}$, $\mu(A) < \infty$. Then $M(A)$ is a random variable by definition of M, and this extends to any $A \in \mathcal{E}$ by considering $A \cap E_n, n \geq 0$ and summation. Since any measurable non-negative function is the increasing limit of finite linear combinations of such indicator functions, we obtain that $M(f)$ is a random variable as a limit of random variables. Moreover, a similar argument shows that $M(f\mathbb{1}_{E_n}), n \geq 0$ are independent random variables.

Next, assume $f \geq 0$. The number N_n of atoms of M that fall in E_n has law $\mathcal{P}(\mu(E_n))$ and given $N_n = k$, the atoms can be supposed to be independent random variables with law $\mu(\cdot \cap E_n)/\mu(E_n)$. Therefore,

$$\begin{aligned} E[\exp(-M(f\mathbb{1}_{E_n}))] &= \sum_{k=0}^{\infty} e^{-\mu(E_n)} \frac{\mu(E_n)^k}{k!} \left(\int_{E_n} \frac{\mu(\mathrm{d}x)}{\mu(E_n)} e^{-f(x)}\right)^k \\ &= \exp\left(-\int_{E_n} \mu(\mathrm{d}x)(1 - \exp(-f(x)))\right) \end{aligned}$$

From the independence of the variables $M(f\mathbb{1}_{E_n})$, we can then take products over $n \geq 0$ (i.e. apply monotone convergence) and obtain the wanted formula.

From this, we obtain the first moment formula for functions $f \geq 0$. If f is a measurable function from $E \to \mathbb{R}$, applying the result to $|f|$ shows that if $f \in L^1(\mu)$, then $M(|f|) < \infty$ a.s. so $M(f)$ is well-defined for almost every ω, and defines a random variable as it is equal to $M(f^+) - M(f^-)$.

To establish the last formula of the theorem, in the case where $f \in L^1(\mu)$, follows by the same kind of arguments: first, we establish the formula for $f\mathbb{1}_{E_n}$ in place of f. Then, to obtain the result, we must show that $\int_{A_n} \mu(\mathrm{d}x)(e^{\mathrm{i}f(x)} - 1)$ converges to $\int_E \mu(\mathrm{d}x)(e^{\mathrm{i}f(x)} - 1)$, where $A_n = E_0 \cup \cdots \cup E_n$. But $|e^{\mathrm{i}f(x)} - 1| \leq |f(x)|$, whence the function under consideration is integrable with respect to μ, giving the result ($|\int_{E \setminus A_n} g(x)\mu(\mathrm{d}x)| \leq \int_{E\setminus A_n} |g(x)|\mu(\mathrm{d}x)$ decreases to 0 whenever g is integrable). □

7.3 Poisson point processes

We now show how Poisson random measures can be used to define certain stochastic processes. Let $(E, \mathcal{E})$ be a measurable space, and consider a σ-finite measure G on $(E, \mathcal{E})$. Let μ be the product measure $\mathrm{d}t \otimes G(\mathrm{d}x)$ on $\mathbb{R}_+ \times E$, where $\mathrm{d}t$ is the Lebesgue measure on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. Otherwise said, μ is the unique measure such that $\mu([0,t] \times A) = tG(A)$ for $t \geq 0$ and $A \in \mathcal{E}$.

A *Lévy process* $(X_t, t \geq 0)$ (with values in $\mathbb{R}$) is a process with independent and stationary increments, i.e. such that for every $0 = t_0 \leq t_1 \leq \ldots \leq t_k$, the random variables $(X_{t_i} - X_{t_{i-1}}, 1 \leq i \leq k)$ are independent with respective laws that of $X_{t_i - t_{i-1}}, 1 \leq i \leq k$. Equivalently, X is a Lévy process if and only if $X^{(t)} = (X_{t+s} - X_t, s \geq 0)$ has same law as X and is independent of $\mathcal{F}_t^X = \sigma(X_s, 0 \leq s \leq t)$ for every $t \geq 0$ (simple Markov property).

Proposition 7.3.1 *A Poisson random measure M whose intensity μ is of the above form is called a* Poisson point process*. If f be a measurable G-integrable function on E, then the process*

$$N_t^f = \int_{[0,t]\times E} f(x) M(\mathrm{d}s, \mathrm{d}x)\,, \quad t \geq 0\,,$$

is a Lévy process. Moreover, the process

$$M_t^f = \int_{[0,t]\times E} f(x) M(\mathrm{d}s, \mathrm{d}x) - t \int_E f(x) G(\mathrm{d}x)\,, \quad t \geq 0,$$

is a martingale with respect to the filtration $\mathcal{F}_t = \sigma(M([0,s] \times A), s \leq t, A \in \mathcal{E}), t \geq 0$. If moreover $f \in L^2(\mu)$, the process

$$(M_t^f)^2 - t \int_E f(x)^2 G(\mathrm{d}x),$$

is an $(\mathcal{F}_t)$-martingale.

Proof. For $s \leq t$, we have $N_t^f - N_s^f = \int_{(s,t]\times E} f(x) M(\mathrm{d}u, \mathrm{d}x)$. Moreover, it is easy to check that $M(\mathrm{d}u, \mathrm{d}x)\mathbb{1}_{\{u \in (s,t]\}}$ has same law as the image of $M(\mathrm{d}u, \mathrm{d}x)\mathbb{1}_{\{u \in (0,t-s]\}}$ under $(u, x) \mapsto (s + u, x)$ from $\mathbb{R}_+ \times E$ to itself, and is independent of $M(\mathrm{d}u, \mathrm{d}x)\mathbb{1}_{\{u \in [0,s]\}}$. We obtain that N^f has stationary and independent increments. The fact that M^f is a martingale is a straightforward consequence of the first moment formula and the simple Markov property. The last statement comes from writing $(M_t^f)^2 = (M_t^f - M_s^f + M_s^f)^2$ and expanding, then using the variance formula and the simple Markov property. □

7.3.1 Example: the Poisson process

Let $X_1, X_2, \ldots$ be a sequence of independent exponential random variables with parameter θ, and define $0 = T_0 \leq T_1 \leq \ldots$ by $T_n = X_1 + \ldots + X_n$. We let

$$N_t^\theta = \sum_{n=1}^{\infty} \mathbb{1}_{\{T_n \leq t\}}\,, \qquad t \geq 0,$$

be the càdlàg process that counts the number of times T_n that are $\leq t$. The process $(N^\theta_t, t \geq 0)$ is called the (homogeneous) Poisson process with intensity θ. This is the so-called Markovian description of the Poisson process, which is a jump-hold Markov process. The following alternative description makes use of Poisson random measures. We give the statement without proof, which can be found in textbooks, or make a good exercise (first notice that with both definitions, N^θ is a process with stationary and independent increments).

Proposition 7.3.2 *Let $\theta > 0$, and let M be a Poisson random measure with intensity $\theta \mathrm{d}t$ on $\mathbb{R}_+$. Then the process*

$$N^\theta_t = M([0,t]) \,, \qquad t \geq 0$$

is a Poisson process with intensity θ.

The set of atoms of the measure M itself is sometimes also called a Poisson (point) process with intensity θ.

7.3.2 Example: compound Poisson processes

A compound Poisson process with intensity ν is a process of the form

$$N^\nu_t = \int_{[0,t]\times\mathbb{R}} x M(\mathrm{d}s, \mathrm{d}x) \,, \qquad t \geq 0,$$

where M is a Poisson random measure with intensity $\mathrm{d}t \otimes \nu(\mathrm{d}x)$ and ν is a finite measure on $\mathbb{R}$. Alternatively, if we write M in the form $\sum_{i\in I} \delta_{(t_i,x_i)}$, for every $t \geq 0$ we can write $X_t = x_i$ whenever $t = t_i$ and $X_t = 0$ otherwise. As an exercise, one can prove that this is a.s. well-defined, i.e. that a.s., for every $t \geq 0$, the set $\{i \in I : t_i = t\}$ has at most one element. With this notation, we can write

$$N^\nu_t = \sum_{0\leq s\leq t} X_s \,, \qquad t \geq 0$$

(notice that there is a.s. a finite set of times $s \in [0,t]$ such that $X_s \neq 0$, so the sum is meaningful).

There is a Markov jump-hold description for these processes as well: if N^θ is a Poisson process with parameter $\theta = \nu(\mathbb{R})$ and jump times $0 < T_1 < T_2 < \ldots$, and if $Y_1, Y_2, \ldots$ is a sequence of i.i.d. random variables, independent of N^θ and with law ν/θ, then the process

$$\sum_{n\geq 1} Y_n \mathbb{1}_{\{T_n\leq t\}} \,, \qquad t \geq 0,$$

is a compound Poisson process with intensity ν. This comes from the following *marking* property of Poisson measures: suppose we have a description of a Poisson random measure $M(\mathrm{d}x)$ with intensity μ as $\sum_{i\in I} \delta_{X_i}(\mathrm{d}x)$, where $(X_i, i \in I)$ is a countable family of random variables. If $(Y_i, i \in I)$ is a family of i.i.d. random variables with law ν, and independent of M, then $M' = \sum_{i\in I} \delta_{(X_i,Y_i)}$ is a Poisson random measure with intensity the product measure $\mu \otimes \nu$.

We let $\mathrm{CP}(\nu)$ be the law of N_1^ν, it is called the compound Poisson distribution with intensity ν. It can be written in the form

$$\mathrm{CP}(\nu) = \sum_{n\geq 0} e^{-\nu(\mathbb{R})} \frac{\nu^{*n}}{n!} ,$$

where ν^{*n} is the n-fold convolution of the measure ν. Recall that the convolution of two finite measures μ, ν on $\mathbb{R}$ is the unique measure $\mu * \nu$ which is characterized by

$$\mu * \nu(A) = \iint \mathbb{1}_A(x+y)\mu(\mathrm{d}x)\nu(\mathrm{d}y) , \qquad A \in \mathcal{B}_{\mathbb{R}},$$

and that if μ, ν are probability measures, then $\mu*\nu$ is the law of the sum of two independent random variables with respective laws μ, ν). The characteristic function of $\mathrm{CP}(\nu)$ is given by

$$\Phi_{\mathrm{CP}(\nu)}(u) = \exp(-\nu(\mathbb{R})(1 - \Phi_{\nu/\nu(\mathbb{R})}(u))),$$

where $\Phi_{\nu/\nu(\mathbb{R})}$ is the characteristic function of $\nu/\nu(\mathbb{R})$.

Chapter 8

Infinitely divisible laws and Lévy processes

In this chapter, we consider only random variables and processes with values in $\mathbb{R}$.

8.1 Infinitely divisible laws and Lévy-Khintchine formula

Definition 8.1.1 *Let μ be a probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$. We say that μ is infinitely divisible (ID) if for every $n \geq 1$, there exists a probability distribution μ_n such that if $X_1, \ldots, X_n$ are independent with law μ_n, then their sum $X_1 + \ldots + X_n$ has law μ.*

Otherwise said, for every n, there exists μ_n such that $\mu_n^{*n} = \mu$, where $*$ stands for the convolution operation for measures. Yet otherwise said, the characteristic function Φ of μ is such that for every $n \geq 1$, there exists another characteristic function Φ_n with $\Phi_n^n = \Phi$. We stress that it is not the existence of a function whose n-th power is Φ which is problematic, but really that this function is a characteristic function.

To start with, let us mention examples of ID laws. Constant random variables are ID. The Gaussian $\mathcal{N}(m, \sigma^2)$ is the convolution of n laws $\mathcal{N}(m/n, \sigma^2/n)$, so it is ID. The Poisson law $\mathcal{P}(\lambda)$ is also ID as the convolution of n laws $\mathcal{P}(\lambda/n)$. More generally, a compound Poisson law $\mathrm{CP}(\nu)$ is ID, as the n-th power of $\mathrm{CP}(\nu/n)$.

It is a bit harder to see, but however true, that exponential and geometric distributions are ID. However, the uniform distribution on $[0, 1]$, or the Bernoulli distribution with parameter $p \in (0, 1)$, are not ID. Suppose indeed that an ID law μ has a support which is bounded above and below by $M > 0$. Then the support of μ_n is bounded by M/n, but then its variance is $\leq M^2/n^2$, which shows that the variance of μ is $\leq M^2/n$ for every n, hence μ is a Dirac mass.

The main goal of this chapter is to give a structural theorem for ID laws, the Lévy-Khintchine formula. Say that a triple (a, q, Π) is a *Lévy triple* if

- $a \in \mathbb{R}$,
- $q \geq 0$,

- Π is a σ-finite measure on $\mathbb{R}$ such that $\Pi(\{0\}) = 0$ and $\int (x^2 \wedge 1)\, \Pi(\mathrm{d}x) < \infty$.

In particular, $\Pi(\mathbb{1}_{\{|x|>\varepsilon\}}) < \infty$ for every $\varepsilon > 0$.

Theorem 8.1.1 (Lévy-Khintchine formula) *Let μ be an ID law. Then there exist a unique Lévy triple (a, q, Π) such that if Φ is the characteristic function of μ, $\Phi(u) = e^{\psi(u)}$, where ψ is the* characteristic exponent *given by*

$$\psi(u) = \mathrm{i}au - \frac{q}{2}u^2 + \int_{\mathbb{R}} (e^{\mathrm{i}ux} - 1 - \mathrm{i}ux\mathbb{1}_{\{|x|<1\}})\Pi(\mathrm{d}x).$$

We reobtain the constants for $a = q = 0$, the normal laws for $\Pi = 0$, and the compound Poisson laws for $a = q = 0$ and $\Pi(\mathrm{d}x) = \nu(\mathrm{d}x)$, a finite measure.

Lemma 8.1.1 *The characteristic function Φ of an ID law never vanishes, and therefore, the characteristic exponent ψ with $\psi(0) = 0$ is well-defined and unique.*

Proof. If μ is ID, then $\Phi = \Phi_n^n$ for all n, where Φ_n is the characteristic exponent of some law μ_n. Therefore, $|\Phi| = |\Phi_n|^n$, and taking logarithms, as $n \to \infty$, we see that Φ_n converges pointwise to $\mathbb{1}_{\{\Phi \neq 0\}}$. However, Φ is continuous and takes the value 1 at 0, so it is non-zero in a neighborhood of 0, and $\mathbb{1}_{\{\Phi \neq 0\}}$ equals 1 (hence is continuous) in a neighborhood of 0. By Lévy's convergence theorem, this shows that μ_n weakly converges to some distribution, which has no choice but to be δ_0. In particular, Φ never vanishes.

To conclude, it is a standard topology exercise that a continuous function $f : \mathbb{R} \to \mathbb{C}$ that never vanishes and such that $f(0) = 1$ can be uniquely lifted into a continuous function $g : \mathbb{R} \to \mathbb{C}$ with $g(0) = 0$, so that $e^g = f$. □

As a corollary, notice that Φ_n, the n-th 'root' of Φ, can itself be written in the form e^{ψ_n} for a unique continuous ψ_n satisfying $\psi_n(0) = 0$, so that $\psi_n = \psi/n$. It also entails the uniqueness of μ_n such that $\mu_n^{*n} = \mu$.

Lemma 8.1.2 *An ID law is the weak limit of compound Poisson laws.*

Proof. Let Φ_n be the characteristic function of μ_n, as defined above. Then since $(1-(1-\Phi_n))^n = \Phi$, and $\Phi_n \to 1$ pointwise, we obtain that $-n(1-\Phi_n) \to \psi$ pointwise, taking the complex logarithm in a neighborhood of 1. In fact, this convergence even holds uniformly on compact neighborhood of 0, a fact that we will need later on. Exponentiating gives $\exp(-n(1-\Phi_n)) \to \Phi$. However, on the left-hand size we can recognize the characterisitc function of a compound Poisson law with intensity $n\mu_n$. □

Proof of the Lévy-Khintchine formula. We must now prove that the limit ψ of $-n(1-\Phi_n)$ has the form given in the statement of the theorem. First of all, we make a technical modification of the statement, replacing the $\mathbb{1}_{\{|x|<1\}}$ in the statement by a continuous function h such that $\mathbb{1}_{\{|x|<1\}} \leq h \leq \mathbb{1}_{\{|x|\leq 2\}}$. This will just modify the value of a in the statement.

Let $\eta_n(\mathrm{d}x) = (1 \wedge x^2) n\mu_n(\mathrm{d}x)$, which is a sequence of measures with finite total mass. Suppose we know that the sequence $(\eta_n, n \geq 1)$ is tight and $(\eta_n(\mathbb{R}), n \geq 1)$ is bounded, and let η be the limit of η_n along some subsequence n_k. Then

$$\begin{aligned} \int_{\mathbb{R}} (e^{\mathrm{i}ux} - 1) n\mu_n(\mathrm{d}x) &= \int_{\mathbb{R}} (e^{\mathrm{i}ux} - 1) \frac{\eta_n(\mathrm{d}x)}{x^2 \wedge 1} \\ &= \int_{\mathbb{R}} \frac{(e^{\mathrm{i}ux} - 1 - \mathrm{i}uxh(x))}{x^2 \wedge 1} \eta_n(\mathrm{d}x) + \mathrm{i}u \int_{\mathbb{R}} \frac{xh(x)}{x^2 \wedge 1} \eta_n(\mathrm{d}x) \\ &= \int_{\mathbb{R}} \Theta(u, x) \eta_n(\mathrm{d}x) + \mathrm{i}ua_n \end{aligned} \tag{8.1}$$

where

$$\Theta_n(u, x) = \begin{cases} (e^{\mathrm{i}ux} - 1 - \mathrm{i}uxh(x))/(x^2 \wedge 1) & \text{if } x \neq 0 \\ -u^2/2 & \text{if } x = 0, \end{cases}$$

and $a_n = \int_{\mathbb{R}} \frac{xh(x)}{x^2 \wedge 1} \eta_n(\mathrm{d}x)$. Now, for each fixed u, $\Theta(\cdot, u)$ is a continuous bounded function, and therefore, along the subsequence n_k, $\int_{\mathbb{R}} \Theta(u, x) \eta_n(\mathrm{d}x)$ converges to $\int_{\mathbb{R}} \Theta(u, x) \eta(\mathrm{d}x)$. Since the left-hand side in (8.1) converges to $\psi(u)$, this implies that a_{n_k} converges to some $a \in \mathbb{R}$. Therefore, if $q = \eta(\{0\})$, we obtain that

$$\psi(u) = \mathrm{i}ua - \frac{q}{2} u^2 + \int_{\mathbb{R}} (e^{\mathrm{i}ux} - 1 - \mathrm{i}uxh(x)) \Pi(\mathrm{d}x),$$

where $\Pi(dx) = \mathbb{1}_{\{x \neq 0\}} (x^2 \wedge 1)^{-1} \eta(\mathrm{d}x)$ is a measure that is σ-finite, integrates $x^2 \wedge 1$, and does not charge 0. Hence the result.

So, let us prove that $(\eta_n, n \geq 1)$ is tight and that the total masses are bounded. First, $x^2 \mathbb{1}_{\{|x| \leq 1\}} \leq C(1 - \cos x)$ for some $C > 0$, so

$$\eta_n(|x| \leq 1) = \int_{\mathbb{R}} x^2 \mathbb{1}_{\{|x| \leq 1\}} n\mu_n(\mathrm{d}y) \leq C \int_{\mathbb{R}} (1 - \cos x) n\mu_n(\mathrm{d}x),$$

which converges to $-C\Re\psi(1)$ as $n \to \infty$. Second, adapting Lemma 5.4.1, since $\eta_n \mathbb{1}_{\{|x| \geq 1\}} = n\mu_n \mathbb{1}_{\{|x| \geq 1\}}$, for some $C > 0$, and every $K \geq 1$,

$$\eta_n(|x| \geq K) \leq CK \int_{\{|x| \leq K^{-1}\}} n(1 - \Re\Phi_n(x)) \mathrm{d}x \underset{n \to \infty}{\to} -CK \int_{-K^{-1}}^{K^{-1}} \Re\psi(x) \mathrm{d}x,$$

where the limit can be taken because the convergence of the integrand is uniform on compact neighborhoods of 0, as stressed in the proof of Lemma 8.1.2. Now the limit can be made as small as wanted for K large enough, because ψ is continuous and $\psi(0) = 0$. This entails the result.

The uniqueness statement will be proved in the next section. □

8.2 Lévy processes

In this section, all the Lévy processes under consideration start at $X_0 = 0$. Lévy processes are closely related to ID laws, indeed, if X is a Lévy process, then the random variable

X_1 can be written as a sum of i.i.d. variables

$$X_1 = \sum_{k=1}^{n} (X_{k/n} - X_{(k-1)/n}),$$

hence is ID. In fact, (laws of) càdlàg Lévy processes are in one-to-one correspondence with ID laws, as we show in this section. The first problem we address is that the mapping $X \mapsto X_1$ is injective from the set of (laws of) càdlàg Lévy processes to the set of ID laws.

Proposition 8.2.1 *Let μ be an ID law. Then there exists at most one càdlàg Lévy process $(X_t, t \geq 0)$ such that X_1 has law μ. Moreover, if μ has a Lévy triple (a, q, Π) with associated characteristic exponent*

$$\psi(u) = \mathrm{i}au - \frac{q}{2}u^2 + \int_{\mathbb{R}} (e^{\mathrm{i}ux} - 1 - \mathrm{i}ux \mathbb{1}_{\{|x|<1\}}) \Pi(\mathrm{d}x),$$

then the law of such a process X is entirely characterized by the formula.

$$E[\exp(\mathrm{i}uX_t)] = \exp(t\psi(u)).$$

Proof. If X is as in the statement, then for $n \geq 1$, $X_{1/n}$ must have ψ/n as characteristic exponent by uniqueness of the characteristic exponent of the n-th root of an ID law. From this we deduce easily that $E[\exp(\mathrm{i}uX_t)] = \exp(t\psi(u))$ for every $t \in \mathbb{Q}_+$ and $u \in \mathbb{R}$. Since X is càdlàg we deduce the result for every $t \in \mathbb{R}_+$ by approximating t by $2^{-n}\lceil 2^n t \rceil$. Therefore, the one-dimensional marginal distributions of X are uniquely determined by μ. It is then easy to check that the finite-marginal distributions of Lévy processes are in turn determined by their one-dimensional marginal distributions, because the increments $(X_{t_j} - X_{t_{j-1}}, 1 \leq j \leq k)$, for any $0 = t_0 \leq t_1 \leq \ldots \leq t_k$, are independent with respective laws $(X_{t_j - t_{j-1}}, 1 \leq j \leq k)$. Hence the result. $\square$

The next theorem is a kind of converse to this theorem, and gives an explicit construction of 'the' càdlàg Lévy process whose law at time 1 is a given ID law μ. Let (a, q, Π) be a Lévy triple associated to an ID law μ. Consider a Poisson random measure M on $\mathbb{R}_+ \times \mathbb{R}$ with intensity $\mathrm{d}t \otimes \Pi(\mathrm{d}x)$, and let $(t, \Delta_t) = (t, x)$ if M has an atom of the form (t, x), and $\Delta_t = 0$ otherwise. For any $n \geq 1$, consider the martingale

$$Y_t^n = \int_{[0,t]} \int_{\mathbb{R}} \mathbb{1}_{\{n^{-1} \leq |y| < 1\}} y M(\mathrm{d}s, \mathrm{d}y) - t \int y \Pi(\mathrm{d}y) \mathbb{1}_{\{n^{-1} \leq |y| < 1\}}, \qquad t \geq 0$$

associated by Proposition 7.3.1 with the Poisson measure $M(\mathrm{d}t, \mathrm{d}x)\mathbb{1}_{\{n^{-1} \leq |x| < 1\}}$. Notice also that this last measure has always a finite number of atoms, because $\Pi(\mathrm{d}x)\mathbb{1}_{\{|x| > n^{-1}\}}$ is a finite measure by assumption on Π, so that

$$Y_t^n = \sum_{0 \leq s \leq t} \Delta_t \mathbb{1}_{\{n^{-1} \leq |\Delta_t| < 1\}} - t \int y \Pi(\mathrm{d}y) \mathbb{1}_{\{n^{-1} \leq |y| < 1\}}, \qquad t \geq 0. \tag{8.2}$$

Independently of M, let B_t be a standard Brownian motion. Finally notice that

$$Y_t^0 = \sum_{0 \leq s \leq t} \Delta_t \mathbb{1}_{\{|\Delta_t| \geq 1\}}, \qquad t \geq 0 \tag{8.3}$$

is a compound Poisson process with intensity $\Pi(\mathrm{d}x)\mathbb{1}_{\{|x|>1\}}$. We let $\mathcal{F}_t$ be the σ-algebra generated by $\{B_s, Y_s^0, Y_s^n, n \geq 1; 0 \leq s \leq t\}$.

Theorem 8.2.1 (Lévy-Itō's theorem) *Let μ be an ID law, with Lévy triple (a, q, Π), and let $B, Y^0, Y^n, n \geq 1$ denote the processes associated with this triple as explained above. Then there exists a càdlàg square integrable $(\mathcal{F}_t)$-martingale Y^∞ such that for every $t \geq 0$,*

$$E\left[\sup_{0\leq s\leq t} |Y_s^n - Y_s^\infty|^2\right] \underset{n\to\infty}{\longrightarrow} 0.$$

Moreover, the process

$$X_t = at + \sqrt{q}B_t + Y_t^0 + Y_t^\infty\,, \qquad t \geq 0$$

is a Lévy process such that X_1 has distribution μ.

This theorem, which is extremely useful in practice, is an explicit construction of any càdlàg Lévy process, out of four independent ingredients: a deterministic drift, a Brownian motion, and a jump part made of a compound Poisson process, and a *compensated* L^2 càdlàg martingale. The compensation by a drift in the formula defining Y^n is crucial, because the identity function is in general not in $L^1(\Pi)$, so that $\int_{[0,t]\times[0,1]} xM(\mathrm{d}s, \mathrm{d}x)$ is in general ill-defined.

Proof. For every $n > m > 0$, the process $Y^n - Y^m$ is a càdlàg martingale, and Doob's L^2 inequality gives

$$\begin{aligned} E\left[\sup_{0\leq s\leq t} |Y_s^n - Y_s^m|^2\right] &\leq 4E[|Y_t^n - Y_t^m|^2] = 4t\int_{\mathbb{R}} y^2\Pi(\mathrm{d}y)\mathbb{1}_{\{n^{-1}\leq|x|<m^{-1}\}} \\ &\leq 4t\int_{\mathbb{R}} y^2\Pi(\mathrm{d}y)\mathbb{1}_{\{0<|x|<m^{-1}\}}\,, \end{aligned}$$

where we used the last statement of Proposition 7.3.1 for the second equality. Since $\int y^2\Pi(\mathrm{d}y)\mathbb{1}_{\{0<x<1\}} < \infty$, this can be made as small as wanted for m large enough. In particular, for every t, Y_t^n is a Cauchy sequence in L^2, and thus converges to a limit Y_t^∞ in L^2. The process $(Y_t^\infty, t \geq 0)$ then defines a martingale, as is checked by passing to the limit as $n \to \infty$ in $E[Y_t^n|\mathcal{F}_s] = Y_s^n$. Moreover, by passing to the limit as $n \to \infty$, we obtain that $\sup_{0\leq s\leq t} |Y_s^\infty - Y_s^m|^2$ converges in L^2 to 0 as $m \to \infty$, for every $t \geq 0$. By extracting along a subsequence, we may assume that the convergence is almost-sure, so that Y^∞ is the a.s. uniform limit over compacts of càdlàg processes, hence is also a càdlàg process (in fact, admits a càdlàg version).

Therefore, the process X defined in the statement is indeed a càdlàg process, and it is easy to show that it is a Lévy process, being a pointwise L^2 limit of Lévy processes. The last thing that remains to be proved is that X_1 has law μ. But from the independence of the components used to build X, we obtain that if $X_t^n = at + \sqrt{q}B_t + Y_t^0 + Y_t^n$,

$$\begin{aligned} E[\exp(\mathrm{i}uX_1^n)] &= \exp\left(\mathrm{i}ua - \frac{q}{2}u^2 + \int_{\mathbb{R}} (e^{\mathrm{i}uy} - 1)\Pi(\mathrm{d}y)\mathbb{1}_{\{|x|\geq 1\}}\right. \\ &\qquad \left. + \int_{\mathbb{R}} (e^{\mathrm{i}uy} - 1 - \mathrm{i}uy)\Pi(\mathrm{d}y)\mathbb{1}_{\{n^{-1}\leq|y|<1\}}\right). \end{aligned}$$

By passing to the limit as $n \to \infty$, we obtain that X_1 has the characteristic function associated to μ. $\square$

Proof of the uniqueness in Theorem 8.1.1. Let μ be an ID law with Lévy triple (a, q, Π), and let X be the unique càdlàg Lévy process such that X_1 has law μ, given by Proposition 8.2.1. Then Theorem 8.2.1 shows that the jumps of X between times 0 and t, i.e. the process $(\Delta_s, 0 \leq s \leq t)$ defined by $\Delta_s = X_s - X_{s-}, s \geq 0$, are the atoms of a Poisson random measure M with intensity $t\Pi$ on $\mathbb{R}$. This intensity is determined by the law of M through the first moment formula $t\Pi(A) = E[M(A)]$ of Proposition 7.2.1. Then, by defining Y^0 and Y^n by the formulas (8.2), (8.3) and letting $Y^\infty = \lim_n Y^n$ along a subsequence according to which the limit is almost-sure uniformly on compacts, we obtain that $X - Y^0 - Y^\infty$ is a (scaled) Brownian motion with drift, with same law as $\widetilde{B} = (at + \sqrt{q}B_t, t \geq 0)$. We can recover a as the expectation of $\widetilde{B}_1$, and $\sqrt{q}$ as its variance. Finally, we see that μ uniquely determines its Lévy triple. $\square$

Chapter 9

Exercises

Warmup

The exercises of this section are designed to help remind you of basic concepts of probability theory (random variables, expectation, classical probability distributions, Borel-Cantelli lemmas). The last one is a longer exercise that contains the basic results on uniform integrability that are needed in this course.

Exercise 9.0.1
Remind yourself what the following classical discrete distributions are : Bernoulli with parameter $p \in [0, 1]$, binomial with parameters $(n, p) \in \mathbb{N} \times [0, 1]$, geometric with parameter $p \in [0, 1]$, Poisson with parameter $\lambda \geq 0$.

Do so with the following classical distributions on $\mathbb{R}$: uniform on $[a, b]$, exponential with mean θ^{-1}, gamma with (positive) parameters (a, θ) (mean a/θ, variance a/θ^2), beta with (positive) parameters (a, b), Gaussian with mean m and variance σ^2, Cauchy with parameter a.

Exercise 9.0.2
Compute the distribution of $1/N^2$, where N is a standard Gaussian $\mathcal{N}(0, 1)$ random variable. What is the distribution of N/N', where N, N' are two independent such random variables ?

Exercise 9.0.3
Show that for any countable set I and I-indexed family $(X_i, i \in I)$ of non-negative random variables, then $\sup_{i \in I} E[X_i] \leq E[\sup_{i \in I} X_i]$. Show that these two quantities are equal if for every $i, j \in I$ there exists some $k \in I$ such that $X_i \vee X_j \leq X_k$.

Exercise 9.0.4
Fix $\alpha > 0$, and let $(Z_n, n \geq 0)$ be a sequence of independent random variables with values in $\{0, 1\}$, whose laws are characterized by

$$P(Z_n = 1) = \frac{1}{n^\alpha} = 1 - P(Z_n = 0).$$

Show that Z_n converges to 0 in L^1. Show that $\limsup_n Z_n$ is 0 a.s. if $\alpha > 1$ and 1 a.s. if $\alpha \leq 1$.

Exercise 9.0.5
Let $(X_n, n \geq 1)$ be a sequence of independent exponential random variables with mean 1. Show that $\limsup_n (\log n)^{-1} X_n = 1$ a.s.

Exercise 9.0.6
Let N be a random $\mathcal{N}(0,1)$ random variable. Show that

$$P(N > x) \leq \frac{1}{x\sqrt{2\pi}} \exp(-x^2/2).$$

Show in fact that as $x \to \infty$,

$$P(N > x) = \frac{1}{x\sqrt{2\pi}} \exp(-x^2/2)(1 + o(1)).$$

Let $(Y_n, n \geq 1)$ be a sequence of independent such Gaussian variables. Show that $\limsup_n (2\log n)^{-1/2} Y_n = 1$ a.s.

Exercise 9.0.7 The basics of uniform integrability
Let $(E, \mathcal{A}, \mu)$ be a measured space, with $\mu(E) < \infty$. If f is a measurable non-negative function, we let $\mu(f)$ be a shorthand for $\int_E f \mathrm{d}\mu$.

A family of $\mathbb{R}$-valued functions $(f_i, i \in I)$ in $L^1(E, \mathcal{A}, \mu)$ is said to be *uniformly integrable* (U.I. in short) if the following holds :

$$\sup_{i \in I} \mu(|f_i| \mathbb{1}_{\{|f_i| > a\}}) \underset{a \to \infty}{\to} 0.$$

You may think of $(E, \mathcal{A}, \mu)$ and the f_i as being a probability space and random variables.

1. Show that a U.I. family is bounded in $L^1(E, \mathcal{A}, \mu)$. Show that the converse is not true.

2. Show that a finite family of integrable functions is U.I.

3. Let $G : \mathbb{R}_+ \to \mathbb{R}_+$ be a measurable function such that $\lim_{x \to \infty} x^{-1} G(x) = +\infty$. Show that for every $C > 0$, the family

$$\{f \in L^1(E, \mathcal{A}, \mu) : \mu(G(|f|)) \leq C\}$$

is U.I. Deduce that a family of measurable functions that is bounded in $L^p(E, \mathcal{A}, \mu)$ for some $p > 1$ is U.I.

4. (Harder) Show that the converse is true : if $(f_i, i \in I)$ is a U.I. family, then there exists a function G as in 3. so that $(f_i, i \in I)$ is included in a set of the form of previous displayed expression. (Hint : consider an increasing positive sequence $(a_n, n \geq 0)$ such that $\sup_{i \in I} \mu(|f_i| \mathbb{1}_{\{f_i \geq a_n\}}) \leq 2^{-n}$ for every n)

5. Let $(f_i, i \in I)$ be a family that is bounded in $L^1(E, \mathcal{A}, \mu)$. Show that (i) and (ii) below are equivalent :

(i) $(f_i, i \in I)$ is U.I.

(ii) $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $\forall A \in \mathcal{A}, \mu(A) < \delta \implies \sup_{i \in I} \mu(|f_i| \mathbb{1}_A) < \varepsilon$.

6. Show that if $(f_i, i \in I)$ and $(g_j, j \in J)$ are two U.I. families, then $(f_i + g_j, i \in I, j \in J)$ is also U.I.

7. Let $(f_n, n \geq 0)$ be a sequence of L^1 functions that converges in measure to a measurable function f, i.e. for every $\varepsilon > 0$,

$$\mu(\{|f - f_n| > \varepsilon\}) \underset{n\to\infty}{\to} 0.$$

Show that $(f_n, n \geq 0)$ converges in L^1 to f if and only if $(f_n, n \geq 0)$ is U.I. Hint : For the necessary condition, you might find useful to consider sets such as $\{|f - f_n| > 1\}, \{\varepsilon < |f - f_n| \leq 1\}$ and $\{|f - f_n| \leq \varepsilon\}$.

Remark. This shows that a sequence of random variables converging in probability (or a.s.) to some other random variable has an 'upgraded' L^1 convergence if and only if it is uniformly integrable.

9.1 Conditional expectation

Exercise 9.1.1
Let X, Y be two random variables in L^1 so that

$$E[X|Y] = Y \qquad \text{and} \qquad E[Y|X] = X.$$

Show that $X = Y$ a.s. As a hint, you may want to consider quantities like $E[(X - Y)\mathbb{1}_{\{X>c,Y\leq c\}}] + E[(X - Y)\mathbb{1}_{\{X\leq c,Y\leq c\}}]$.

Exercise 9.1.2
Let X, Y be two independent Bernoulli random variables with parameter $p \in (0, 1)$. Let $Z = \mathbb{1}_{\{X+Y=0\}}$. Compute $E[X|Z], E[Y|Z]$.

Exercise 9.1.3
Let $X \geq 0$ be a random variable on a probability space $(\Omega, \mathcal{F}, P)$, and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. Show that $X > 0$ implies that $E[X|\mathcal{G}] > 0$, up to an event of zero probability. Show that $\{E[X|\mathcal{G}] > 0\}$ is actually the smallest $\mathcal{G}$-measurable event that contains the event $\{X > 0\}$, up to zero probability events.

Exercise 9.1.4
Check that the sum Z of two independent exponential random variables X, Y with parameter $\theta > 0$ (mean $1/\theta$) is a gamma distribution with parameter $(2, \theta)$, whose density with respect to Lebesgue measure is $\theta^2 x \exp(-\theta x)\mathbb{1}_{\{x\geq 0\}}$. Show that for every non-negative measurable h,

$$E[h(X)|Z] = \frac{1}{Z}\int_0^Z h(u)\mathrm{d}u.$$

Conversely, let Z be a random variable with a $\Gamma(2, \theta)$ distribution, and suppose X is a random variable whose conditional distribution given Z is uniform on $[0, Z]$. Namely, for every Borel non-negative function h, $E[h(X)|Z] = Z^{-1}\int_0^Z h(x)\mathrm{d}x$ a.s. Show that X and $Z - X$ are independent, with exponential law.

Exercise 9.1.5
Suppose given $a, b > 0$, and let X, Y be two random variables with values in $\mathbb{Z}_+$ and $\mathbb{R}_+$ respectively, whose distribution is characterized by the formula

$$P(X = n, Y \leq t) = b \int_0^t \frac{(ay)^n}{n!} \exp(-(a+b)y)\mathrm{d}y.$$

Let $n \in \mathbb{Z}_+$ and $h : \mathbb{R}_+ \to \mathbb{R}_+$ be a measurable function, compute $E[h(Y)|X = n]$. Then compute $E[Y/(X+1)]$, $E[\mathbb{1}_{\{X=n\}}|Y]$ and $E[X|Y]$.

Exercise 9.1.6
Let $(X, Y_1, \ldots, Y_n)$ be a random vector with components in L^2. Show that the best approximation of X in the L^2 norm by an affine combination of the $(Y_i, 1 \leq i \leq n)$, say of the form $\lambda_0 + \sum_{i=1}^n \lambda_i(Y_i - E[Y_i])$, is given by $\lambda_0 = E[X]$ and any solution $(\lambda_1, \ldots, \lambda_n)$ of the linear system

$$\mathrm{Cov}\,(X, Y_j) = \sum_{i=1}^n \lambda_i \mathrm{Cov}\,(Y_i, Y_j)\,, \qquad 1 \leq j \leq n.$$

This affine combination is called the *linear regression* of X with respect to $(Y_1, \ldots, Y_n)$.

If $(X, Y_1, \ldots, Y_n)$ is a Gaussian random vector, show that $E[X|Y_1, \ldots, Y_n]$ equals the linear regression of X with respect to $(Y_1, \ldots, Y_n)$.

Exercise 9.1.7
Let $X \in L^1(\Omega, \mathcal{F}, P)$. Show that the family

$$\{E[X|\mathcal{G}] : \mathcal{G} \text{ is a sub-}\sigma\text{-algebra of } \mathcal{F}\}$$

is uniformly integrable.

Exercise 9.1.8 Conditional independence
Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. Two random variables X, Y are said to be independent conditionally on $\mathcal{G}$ if for every non-negative measurable f, g,

$$E[f(X)g(Y)|\mathcal{G}] = E[f(X)|\mathcal{G}]\,E[g(Y)|\mathcal{G}].$$

What are two random variables independent conditionally on $\{\varnothing, \Omega\}$? On $\mathcal{F}$?

1. Show that X, Y are independent conditionally on $\mathcal{G}$ if and only if for every non-negative $\mathcal{G}$-measurable random variable Z, and every f, g non-negative measurable functions,

$$E[f(X)g(Y)Z] = E[f(X)ZE[g(Y)|\mathcal{G}]],$$

and this if and only if for every measurable non-negative g,

$$E[g(Y)|\mathcal{G} \vee \sigma(X)] = E[g(Y)|\mathcal{G}].$$

Comment the case $\mathcal{G} = \{\varnothing, \Omega\}$.

2. Suppose given three random variables X, Y, Z with a positive density $p(x, y, z)$. Suppose X, Y are independent conditionally on $\sigma(Z)$. Show that there exist measurable positive functions r, s so that $p(x, y, z) = q(z)r(x, z)s(y, z)$ where q is the density of Z, and conversely.

9.2 Discrete-time martingales

Exercise 9.2.1
Let $(X_n, n \geq 0)$ be an integrable process with values in a countable subset $E \subset \mathbb{R}$. Show that X is a martingale with respect to its natural filtration if and only if for every n and every $i_0, \ldots, i_n \in E$,

$$E[X_{n+1}|X_0 = i_0, \ldots, X_n = i_n] = i_n.$$

Exercise 9.2.2
Let $(X_n, n \geq 1)$ be a sequence of independent random variables with respective laws given by

$$P(X_n = -n^2) = \frac{1}{n^2}, \qquad P\left(X_n = \frac{n^2}{n^2-1}\right) = 1 - \frac{1}{n^2}.$$

Let $S_n = X_1 + \ldots + X_n$. Show that $S_n/n \to 1$ a.s. as $n \to \infty$, and deduce that $(S_n, n \geq 0)$ is a martingale which converges to $+\infty$.

Exercise 9.2.3
Let $(\Omega, \mathcal{F}, (\mathcal{F}_n), P)$ be filtered probability space. Let $A \in \mathcal{F}_n$ for some n, and let $m, m' \geq n$. Show that $m\mathbb{1}_A + m'\mathbb{1}_{A^c}$ is a stopping time.

Show that an adapted process $(X_n, n \geq 0)$ with respect to some filtered probability space is a martingale if and only if it is integrable, and for every bounded stopping time T, $E[X_T] = E[X_0]$.

Exercise 9.2.4
Let X be a martingale (resp. supermartingale) on some filtered probability space, and let T be an a.s. finite stopping time. Prove that $E[X_T] = E[X_0]$ (resp. $E[X_T] \leq E[X_0]$) if either one of the following conditions holds:

1. X is bounded ($\exists M > 0 : \forall n \geq 0, |X_n| \leq M$ a.s.).
2. X has bounded increments ($\exists M > 0 : \forall n \geq 0, |X_{n+1} - X_n| \leq M$ a.s.) and $E[T] < \infty$.

Exercise 9.2.5
Let $(X_n, n \geq 0)$ be a non-negative supermartingale. Show the maximal inequality for $a > 0$:

$$aP\left(\max_{0 \leq k \leq n} X_k \geq a\right) \leq E[X_0].$$

Exercise 9.2.6
Let T be an $(\mathcal{F}_n, n \geq 0)$-stopping time such that for some integer $N > 0$ and $\varepsilon > 0$,

$$P(T \leq N + n|\mathcal{F}_n) \geq \varepsilon, \qquad \text{for every } n \geq 0.$$

Show that $E[T] < \infty$. Hint: Find bounds for $P(T > kN)$.

Exercise 9.2.7
Your winnings per unit stake on game n are ε_n, where $(\varepsilon_n, n \geq 0)$ is a sequence of independent random variables with

$$P(\varepsilon_n = 1) = p\,, \qquad P(\varepsilon_n = -1) = 1 - p = q,$$

where $p \in (1/2, 1)$. Your stake C_n on game n must lie between 0 and Z_{n-1}, where Z_{n-1} is your fortune at time $n-1$. Your object is to maximize the expected 'interest rate' $E[\log(Z_N/Z_0)]$ where N is a given integer representing the length of the game, and Z_0, your fortune at time 0, is a given constant. Let $\mathcal{F}_n = \sigma\{\varepsilon_1, \dots, \varepsilon_n\}$. Show that if C is any previsible strategy, that is if C_n is $\mathcal{F}_{n-1}$-measurable for all n, then $\log Z_n - n\alpha$ is a supermartingale, where α denotes the *entropy*

$$\alpha = p \log p + q \log q + \log 2,$$

so that $E[\log(Z_N/Z_0)] \leq N\alpha$, but that, for a certain strategy, $\log Z_n - n\alpha$ is a martingale. What is the best strategy?

Exercise 9.2.8 Pólya's urn
Consider an urn that initially contains two balls, one black, one white. One picks at random one of the balls with equal probability, checks the color, replaces the ball in the urn and adds another ball of the same color. Then resume the procedure. After step n, $n+2$ balls are in the urn, of which $B_n + 1$ are black and $n + 1 - B_n$ are white.

1. Show that $((n+2)^{-1}(B_n+1), n \geq 0)$ is a martingale with respect to a certain filtration you should indicate. Show that it converges a.s. and in L^p for all $p \geq 1$ to a $[0,1]$-valued random variable X_∞.

2. Show that for every k, the process

$$\frac{(B_n+1)(B_n+2)\dots(B_n+k)}{(n+2)(n+3)\dots(n+k+1)}\,, \qquad n \geq 1$$

is a martingale. Deduce the value of $E[X_\infty^k]$, and finally the law of X_∞.

3. Re-obtain this result by directly showing that $P(B_n = k) = (n+1)^{-1}$ for every $n \geq 1, 1 \leq k \leq n$. As a hint, let Y_i be the indicator that the i-th picked ball is black, and compute $P(Y_i = a_i, 1 \leq i \leq n)$ for any $(a_i, 1 \leq i \leq n) \in \{0,1\}^n$.

4. Show that for $0 < \theta < 1$, $(N_n^\theta, n \geq 0)$ is a martingale, where

$$N_n^\theta = \frac{(n+1)!}{B_n!(n-B_n)!}\theta^{B_n}(1-\theta)^{n-B_n}$$

Exercise 9.2.9 Bayes' urn
Let U be a uniform random variable on $[0,1]$, and conditionally on U, let $X_1, X_2, \dots$ be independent Bernoulli random variables with parameter U. Let $B_n = \sum_{i=1}^n X_i$. Show that for every n, $(B_1, \dots, B_n)$ has the same law as the sequence $(B_1, \dots, B_n)$ in the previous exercise. Show that N_n^θ is a conditional density function of U given $B_1, \dots, B_n$.

Exercise 9.2.10 Monkey typing ABRACADABRA
A monkey types a text at random on a keyboard, so that each new letter is picked

uniformly at random among the 26 letters of the roman alphabet. Let X_n be the n-th letter of the monkey's masterpiece, and let T be the first time when the monkey has typed the exact word ABRACADABRA

$$T = \inf\{n \geq 0 : (X_{n-10}, X_{n-9}, \ldots, X_n) = (\mathrm{A, B, R, A, C, A, D, A, B, R, A})\}.$$

Show that $E[T] < \infty$. The goal is to give the exact value of $E[T]$. For this, suppose that just before *each* time n, a player P_n comes and bets 1 gold coin (GC) that X_n will be A. If he loses, he leaves the game, and if he wins, he earns 26GC, which he entirely plays on X_{n+1} being B. If he loses, he leaves, else he earns 26^2GC which he bets on X_{n+2} being R, and so on. Show that

$$E[T] = 26^{11} + 26^4 + 26.$$

(Hint: Use exercise 9.2.4) Why is that larger than the average first time the monkey has typed ABRACADABRI?

Exercise 9.2.11

Let $(X_n, n \geq 0)$ be a sequence of $[0,1]$-valued random variables, which satisfy the following property. First, $X_0 = a$ a.s. for some $a \in (0,1)$, and for $n \geq 0$,

$$P\left(X_{n+1} = \frac{X_n}{2} \middle| \mathcal{F}_n\right) = 1 - X_n\,, \qquad P\left(X_{n+1} = \frac{1+X_n}{2} \middle| \mathcal{F}_n\right) = X_n,$$

where $\mathcal{F}_n = \sigma\{X_k, 0 \leq k \leq n\}$. Here, we have denoted $P(A|\mathcal{G}) = E[\mathbb{1}_A|\mathcal{G}]$.

1. Prove that $(X_n, n \geq 0)$ is a martingale that converges in L^p for every $p \geq 1$.

2. Check that $E[(X_{n+1} - X_n)^2] = E[X_n(1-X_n)]/4$. Then determine $E[X_\infty(1-X_\infty)]$ and deduce the law of X_∞.

Exercise 9.2.12

Let $(X_n, n \geq 0)$ be a martingale in L^2. Show that its increments $(X_{n+1} - X_n, n \geq 0)$ are pairwise orthogonal. Conclude that X is bounded in L^2 if and only if

$$\sum_{n \geq 0} E[(X_{n+1} - X_n)^2] < \infty,$$

and that X_n converges in L^2 in this case, without using the L^2 convergence theorem for martingales.

Exercise 9.2.13 Wald's identity

Let $(X_n, n \geq 0)$ be a sequence of independent and identically distributed real integrable random variables, which are not a.s. 0. We let $S_n = X_1 + \ldots + X_n$ be the associated random walk, and recall that $(S_n - nE[X_1], n \geq 0)$ is a martingale. Let T be a $(\mathcal{F}_n)$-stopping time.

1. Show that

$$E[|S_{T\wedge n} - S_T|] \leq \sum_{k=n+1}^{\infty} E[|X_k|\mathbb{1}_{\{T \geq k\}}] \leq E[|X_1|]E[T\mathbb{1}_{\{T \geq n+1\}}].$$

Deduce that if $E[T] < \infty$, then $S_{T\wedge n}$ converges to S_T in L^1. Deduce that if $E[T] < \infty$, then $E[S_T] = E[X_1]E[T]$.

2. Suppose $E[X_1] = 0$ and $T_a = \inf\{n \geq 0 : S_n > a\}$ for some $a > 0$. Show that $E[T_a] = \infty$.

3. Let now $a < 0 < b$ and $T_{a,b} = \inf\{n \geq 0 : S_n < a \text{ or } S_n > b\}$. Assume that $E[X_1] \neq 0$. By discussing separately the cases where X_1 is bounded or not, prove that $E[T_{a,b}] < \infty$ and that $E[S_{T_{a,b}}] = E[X_1]E[T_{a,b}]$.

4. Assume that $E[X_1] = 0$. Show that $E[T_{a,b}] < \infty$. Hint: consider again separately the cases when X_1 is bounded and unbounded. In the bounded case, think how far $(S_n^2, n \geq 0)$ is from being a martingale.

Exercise 9.2.14 The gambler's ruin

Let $0 < K < N$ be integers. Consider a sequence of independent random variables $(X_n, n \geq 1)$ with $P(X_n = 1) = p = 1 - P(X_n = -1)$, where $p \in (0, 1/2) \cup (1/2, 1)$. Let $S_n = X_1 + \ldots + X_n$ and define

$$T_0 = \inf\{n \geq 1 : S_n = 0\}, \qquad T_N = \inf\{n \geq 1 : S_n = N\}.$$

Show that $T := T_0 \wedge T_N$ is a.s. finite (and in fact has finite expectation). Then show that, letting $q = 1 - p$,

$$M_n = \left(\frac{q}{p}\right)^{S_n}, \qquad N_n = S_n - (p - q)n, \quad n \geq 0,$$

defines two martingales with respect to the natural filtration of $(S_n, n \geq 1)$. Compute $P(T_0 < T_N)$ and $E[S_T]$, $E[T]$.

What happens to this exercise if $p = 1/2$?

Exercise 9.2.15 Azuma-Hœffding inequality

1. Let Y be a random variable taking values in $[-c, c]$ for some $c > 0$, and such that $E[Y] = 0$. Show that for every $\theta \in \mathbb{R}$,

$$E[e^{\theta Y}] \leq \cosh \theta c \leq \exp\left(\frac{\theta^2 c^2}{2}\right).$$

As a hint, the convexity of $z \mapsto e^{z\theta}$ entails that

$$e^{y\theta} \leq \frac{y + c}{2c} e^{c\theta} + \frac{c - y}{2c} e^{-c\theta}.$$

Also, state and prove a conditional version of this fact.

2. Let M be a martingale with $M_0 = 0$, and such that there exists a sequence $(c_n, n \geq 0)$ of positive real numbers such that $|M_n - M_{n-1}| \leq c_n$ for every n. Show that for $x \geq 0$,

$$P\left(\sup_{0 \leq k \leq n} M_k \geq x\right) \leq \exp\left(-\frac{x^2}{2\sum_{k=1}^{n} c_k^2}\right).$$

As a hint, notice that $(e^{\theta M_n}, n \geq 0)$ is a submartingale, and optimize over θ.

Exercise 9.2.16 A discrete Girsanov theorem
Let Ω be the space of real-valued sequences $(\omega_n, n \geq 0)$ such that $\limsup_n \omega_n = +\infty$ and $\liminf_n \omega_n = -\infty$. We say that such sequences *oscillate*. Let $\mathcal{F}_n = \sigma\{X_k, 0 \leq k \leq n\}$ where $X_k(\omega) = \omega_k$ is the k-th projection, and $\mathcal{F} = \mathcal{F}_\infty$. Show that $p = 1/2$ is the only real in $(0, 1)$ such that there exists a probability measure P_p on $(\Omega, \mathcal{F})$ that makes $(X_n, n \geq 0)$ a simple random walk with step distributions

$$P_p(X_1 = 1) = p = 1 - P(X_1 = -1).$$

Let $P_{p,n}$ be the unique probability measure on $(\Omega, \mathcal{F}_n)$ that makes $(X_k, 0 \leq k \leq n)$ a simple random walk with these step distributions. If $p \in (0, 1) \setminus \{1/2\}$, identify the martingale

$$M_n = \frac{\mathrm{d}P_{p,n}}{\mathrm{d}P_{1/2,n}}.$$

Find a finite stopping time T such that $E_{1/2}[M_T] < 1$.

Exercise 9.2.17
Let $f : [0, 1] \to \mathbb{R}$ be a Lipschitz function, i.e. $|f(x) - f(y)| \leq K|x - y|$ for some $K > 0$ and every x, y. Let f_n be the function obtained by interpolating linearly between the values of f taken at numbers of the form $k2^{-n}, 0 \leq k \leq 2^n$, and let $M_n = f'_n$.

1. Show that M_n is a martingale in some filtration.

2. Deduce that there exists an integrable function $g : [0, 1] \to \mathbb{R}$ such that $f(x) = f(0) + \int_0^x g(y)\mathrm{d}y$ for almost every $0 \leq x \leq 1$.

Exercise 9.2.18 Doob's decomposition of submartingales
Let $(X_n, n \geq 0)$ be a submartingale.

1. Show that there exists a unique martingale M_n and a unique previsible process $(A_n, n \geq 0)$ such that $A_0 = 0$, A is increasing and $X = M + A$.

2. Show that M, A are bounded in L^1 if and only if X is, and that $A_\infty < \infty$ a.s. in this case (and even that $E[A_\infty] < \infty$), where A_∞ is the increasing limit of A_n as $n \to \infty$.

Exercise 9.2.19
Let $(X_n, n \geq 0)$ be a U.I. submartingale.

1. Show that if $X = M + A$ is the Doob decomposition of X, then M is U.I.

2. Show that for every pair of stopping times S, T, with $S \leq T$,

$$E[X_T|\mathcal{F}_S] \geq X_S$$

Exercise 9.2.20 Quadratic variation
Let $(X_n, n \geq 0)$ be a square-integrable martingale.

1. Show that there exists a unique increasing previsible process starting at 0, which we denote by $(\langle X\rangle_n, n \geq 0)$, so that $(X_n^2 - \langle X\rangle_n, n \geq 0)$ is a martingale.

2. Let C be a bounded previsible process. Compute $\langle C \cdot X\rangle$.

3. Let T be a stopping time, show that $\langle X^T\rangle = \langle X\rangle^T$.

4. (Harder) Show that $\langle X\rangle_\infty < \infty$ implies that X_n converges as $n \to \infty$, up to a zero probability event. Is the converse true? Show that it is when $\sup_{n\geq 0} |X_{n+1} - X_n| \leq K$ a.s. for some $K > 0$.

9.3 Continuous-time processes

Exercise 9.3.1 Gaussian processes
A real-valued process $(X_t, t \geq 0)$ is called a *Gaussian process* if for every $t_1 < t_2 < \ldots < t_k$, the random vector $(X_{t_1}, \ldots, X_{t_k})$ is a Gaussian random vector. Show that the law of a Gaussian process is uniquely characterized by the numbers $E[X_t], t \geq 0$ and $\mathrm{Cov}\,(X_s, X_t)$ for $s, t \geq 0$.

Exercise 9.3.2
Let T be an exponential random variable with parameter $\lambda > 0$. Define

$$Z_t = \begin{cases} 0 & \text{if } t < T \\ 1 & \text{if } t \geq T \end{cases} , \quad \mathcal{F}_t = \sigma\{Z_s, 0 \leq s \leq t\}, \quad M_t = \begin{cases} 1 - e^{\lambda t} & \text{if } t < T \\ 1 & \text{if } t \geq T \end{cases} .$$

Show that $E[|M_t|] < \infty$ for every $t \geq 0$, and that $E[M_t \mathbb{1}_{\{T>r\}}] = E[M_s \mathbb{1}_{\{T>r\}}]$ for every $r \leq s \leq t$. Deduce that $(M_t, t \geq 0)$ is a càdlàg $(\mathcal{F}_t)$-martingale.

Is M bounded in L^1? Is it uniformly integrable? Is M_{T-} in L^1?

Exercise 9.3.3 Hazard function
Let T be a random variable in $(0, \infty)$ that admits a strictly positive continuous density f on $(0, \infty)$. Let $F(t) = P(T \leq t)$. Let

$$A_t = \int_0^t \frac{f(s)\mathrm{d}s}{1 - F(s)}, \qquad t \geq 0$$

to be the *hazard function of* T. Show that A_T has the law of an exponential random variable with parameter 1. As a hint, consider the distribution function $P(A_T \leq t), t \geq 0$ and write it in terms of the inverse function A^{-1}.

By letting $Z_t = \mathbb{1}_{t \geq T}, t \geq 0$ and $\mathcal{F}_t = \sigma\{Z_s, 0 \leq s \leq t\}$, prove that $(Z_t - A_{T \wedge t}, t \geq 0)$ is a càdlàg martingale with respect to $(\mathcal{F}_t, t \geq 0)$.

The next exercises are designed to (hopefully) help those of you who want to have a better insight on the nature of filtrations and events related to continuous-time processes.

Exercise 9.3.4
Let $\mathcal{C}_1$ be the product σ-algebra on $\Omega = C([0,1], \mathbb{R})$, i.e. the smallest σ-algebra that makes the applications $X_t : \omega \mapsto \omega(t)$ for $t \geq 0$ measurable.

Let $\mathcal{C}_2$ be the (more natural?) Borel σ-algebra on $C([0,1], \mathbb{R})$, when endowed with the uniform norm and the associated topology.

Show that $\mathcal{C}_1 = \mathcal{C}_2$.

Exercise 9.3.5
Let I be a nonempty real interval. Let $\Omega = \mathbb{R}^I$ be the set of all functions defined on I, which is endowed with the product σ-algebra $\mathcal{F}$, i.e. the smallest σ-algebra with respect to which $X_t : \omega \mapsto \omega(t)$ is measurable for every t. Show that

$$\mathcal{G} = \bigcup_{J \prec I} \sigma(X_s, s \in J)$$

is a σ-algebra, where $J \prec I$ stands for $J \subset I$ and J is countable. Deduce that $\mathcal{G} = \mathcal{F}$. Show that the set

$$\{\omega \in \Omega : s \mapsto X_s(\omega) \text{ is continuous}\}$$

is not measurable with respect to $\mathcal{F}$.

9.4 Weak convergence

Exercise 9.4.1
Let $(X_n, n \geq 1)$ be a sequence of independent random variables with uniform distribution on $[0,1]$. Let $M_n = \max(X_1, \ldots, X_n)$. Show that $n(1 - M_n)$ converges in distribution as $n \to \infty$, and determine the limit law.

Exercise 9.4.2
Let $(X_n, n \in \mathbb{N} \sqcup \{\infty\})$ be random variables defined on some probability space $(\Omega, \mathcal{F}, P)$, with values in a metric space (M, d).

1. Suppose that $X_n \to X_\infty$ a.s. as $n \to \infty$. Show that X_n converges to X_∞ in distribution.

2. Suppose that X_n converges in probability to X_∞. Show that X_n converges in distribution to X_∞.

Hint: use the fact that $(X_n, n \geq 0)$ converges in probability to X_∞ if and only if for every subsequence extracted from $(X_n, n \geq 0)$, there exists a further subsequence converging a.s. to X_∞.

3. If X_n converges in distribution to a *constant* $X_\infty = c$, then X_n converges in probability to c.

Exercise 9.4.3
Suppose given sequences $(X_n, n \geq 0), (Y_n, n \geq 0)$ of real-valued random variables, and two extra random variables X, Y, such that X_n, Y_n respectively converge in distribution to X, Y. Is it true that (X_n, Y_n) converges in distribution to (X, Y)? Show that this is true in the following cases

1. For every n, X_n and Y_n are independent, as well as X and Y.
2. Y is a.s. constant (*Hint*: use 3. in the previous exercise).

Exercise 9.4.4
Let m be a probability measure on $\mathbb{R}$. Define, for every $n \geq 0$,

$$m_n(\mathrm{d}x) = \sum_{k \in \mathbb{Z}} m([k2^{-n}, (k+1)2^{-n}))\delta_{k2^{-n}}(\mathrm{d}x),$$

where $\delta_z(\mathrm{d}x)$ denotes the Dirac mass at z. Show that m_n converges weakly to m.

Exercise 9.4.5
1. Let $(X_n, n \geq 1)$ be independent exponential random variables with mean 1. Define

$S_n = X_1 + \ldots + X_n$, and determine without computation the limit of $P(S_n \le n)$ as $n \to \infty$ (*Hint*: which theorem could be useful here?).

2. Determine also without computation the limit of $\exp(-n)\sum_{k=0}^n (k!)^{-1} n^k$.

Hint: recall that the Poisson law with parameter $\lambda > 0$ is the probability distribution on $\mathbb{Z}_+$ that puts mass $e^{-\lambda}\lambda^n/n!$ on the integer n. Then if X, Y are two random variables which are independent and with respective laws that are Poisson with parameters λ, μ, then $X + Y$ has a Poisson law with parameter $\lambda + \mu$. Using this, make the formula look like question 1.

Exercise 9.4.6
Let $(Y_n, n \ge 0)$ be a sequence of random variables so that Y_n follows a Gaussian $\mathcal{N}(m_n, \sigma_n^2)$ law, and suppose that Y_n weakly converges to some Y as $n \to \infty$. Show that there exist $m \in \mathbb{R}$ and $\sigma^2 \ge 0$ so that $m_n \to m, \sigma_n^2 \to \sigma^2$, and that Y is Gaussian $\mathcal{N}(m, \sigma^2)$.

Hint: Use characteristic functions, and first show that the variance converges.

Exercise 9.4.7
Let $d \ge 1$.

1. Show that a finite family of probability measures on $\mathbb{R}^d$ is tight.

2. Assuming Prokhorov's theorem for probability measures on $\mathbb{R}^d$, show that if $(\mu_n, n \ge 0)$ is a sequence of non-negative measures on $\mathbb{R}^d$ which is tight (for every $\varepsilon > 0$ there is a compact $K \subset \mathbb{R}^d$ such that $\sup_{n\ge 0} \mu_n(\mathbb{R}^d \setminus K) < \varepsilon$) and such that

$$\sup_{n\ge 0} \mu_n(\mathbb{R}^d) < \infty,$$

then there exists a subsequence n_k along which μ_n weakly converges to a limit μ (i.e. $\mu_{n_k}(f)$ converges to $\mu(f)$ for every bounded continuous f).

9.5 Brownian motion

Exercise 9.5.1
Recall that a Gaussian process $(X_t, t \ge 0)$ in $\mathbb{R}^d$ is a process such that for every $t_1 < t_2 < \ldots < t_k \in \mathbb{R}_+$, the vector $(X_{t_1}, \ldots, X_{t_k})$ is a Gaussian random vector. Show that the (standard) Brownian motion in $\mathbb{R}^d$ is the unique Gaussian process $(B_t, t \ge 0)$ with $E[B_t] = 0$ for every $t \ge 0$ and $\mathrm{Cov}\,(B_s, B_t) = (s \wedge t)\, I_d$ for every $s, t \ge 0$.

Exercise 9.5.2
Let B be a standard real-valued Brownian motion.

1. Show that a.s.,

$$\limsup_{t\downarrow 0} \frac{B_t}{\sqrt{t}} = \infty\,, \qquad \liminf_{t\downarrow 0} \frac{B_t}{\sqrt{t}} = -\infty.$$

2. Show that $B_n/n \to 0$ a.s. as $n \to \infty$. Then show that a.s. for n large enough, $\sup_{t\in[n,n+1]} |B_t - B_n| \le \sqrt{n}$, and conclude that $B_t/t \to 0$ a.s. as $t \to \infty$.

3. Show that the process

$$B'_t = \begin{cases} tB_{1/t} & \text{if } t > 0 \\ 0 & \text{if } t = 0 \end{cases}, \qquad t \geq 0$$

is a standard Brownian motion (*Hint*: Use Exercise 9.5.1).

4. Use this to show that

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{t}} = \infty\,, \qquad \liminf_{t\to\infty} \frac{B_t}{\sqrt{t}} = -\infty.$$

Exercise 9.5.3 Around hitting times

Let $(B_t, t \geq 0)$ be a standard real-valued Brownian motion.

1. Let $T_x = \inf\{t \geq 0 : B_t = x\}$ for $x \in \mathbb{R}$. Prove that T_x has same distribution as $(x/B_1)^2$, and compute its probability distribution function.

2. For $x, y > 0$, show that

$$P(T_{-y} < T_x) = \frac{x}{x+y}\quad , \quad E[T_x \wedge T_{-y}] = xy.$$

3. Show that if $0 < x < y$, the random variable $T_y - T_x$ has same law as T_{y-x}, and is independent of $\mathcal{F}_{T_x}$ (where $(\mathcal{F}_t, t \geq 0)$ is the natural filtration of B).

Hint: the three questions are independent.

Exercise 9.5.4

Let $(B_t, t \geq 0)$ be a standard real-valued Brownian motion. Compute the joint distribution of $(B_t, \sup_{0\leq s\leq t} B_s)$ for $t \geq 0$.

Exercise 9.5.5

Let $(B_t, t \geq 0)$ be a standard Brownian motion, and let $0 \leq a < b$.

1. Compute the mean and variance of

$$X_n := \sum_{k=1}^{2^n} \left(B_{a+k(b-a)2^{-n}} - B_{a+(k-1)(b-a)2^{-n}}\right)^2.$$

2. Show that X_n converges a.s. and give its limit.

3. Deduce that a.s. there exists no interval $[a, b]$ with $a < b$ such that B is Hölder continuous with exponent $\alpha > 1/2$ on $[a, b]$, i.e. $\sup_{a\leq s,t\leq b}(|B_t - B_s|/|t-s|^\alpha) < \infty$.

Exercise 9.5.6

Let $(B_t, t \geq 0)$ be a standard Brownian motion. Define $G_1 = \sup\{t \leq 1 : B_t = 0\}$ and $D_1 = \inf\{t \geq 1 : B_t = 0\}$.

1. Are these random variables stopping times? Show that G_1 has same distribution as D_1^{-1}.

2. By applying the Markov property at time 1, compute the law of D_1. Deduce that of G_1 (it is called the arcsine law).

Exercise 9.5.7
Let $(B_t, t \geq 0)$ be a standard Brownian motion, and let $(\mathcal{F}_t, t \geq 0)$ be its natural filtration. Determine all the polynomials $f(t, x)$ of degree less than or equal to 3 in x such that $(f(t, B_t), t \geq 0)$ is a martingale.

Exercise 9.5.8
Let $(B_t, t \geq 0)$ be a standard Brownian motion in $\mathbb{R}^3$. We let $R_t = 1/|B_t|$.

1. Show that $(R_t, t \geq 1)$ is bounded in L^2.
2. Show that $E[R_t] \to 0$ as $t \to \infty$.
3. Show that $R_t, t \geq 1$ is a supermartingale. Deduce that $|B_t| \to \infty$ as $t \to \infty$, a.s.

Exercise 9.5.9 Zeros of Brownian motion
Let $(B_t, t \geq 0)$ be a standard real-valued Brownian motion. Let $\mathcal{Z} = \{t \geq 0 : B_t = 0\}$ be the set of zeros of B.

1. Show that it is closed, unbounded and has zero Lebesgue measure a.s.

2. By using the stopping times $D_q = \inf\{t \geq q : B_t = 0\}$ for $q \in \mathbb{Q}_+$, show that $\mathcal{Z}$ has no isolated point a.s.

Exercise 9.5.10
Let $W_0(\mathrm{dw})$ denote Wiener's measure on $\Omega_0 = \{\mathrm{w} \in \mathcal{C}([0,1]) : \mathrm{w}(0) = 0\}$, and define a new probability measure $W_0^{(a)}$ on Ω_0 by

$$\frac{\mathrm{d}W_0^{(a)}}{\mathrm{d}W_0}(\mathrm{w}) = \exp(a\mathrm{w}(1) - a^2/2).$$

1. Show that under $W_0^{(a)}$, the canonical process $X_t : \mathrm{w} \mapsto \mathrm{w}(t)$ remains Gaussian, and give its distribution.

2. Show that $W_0(\{f \in \Omega_0 : \|f\|_\infty < \varepsilon\}) > 0$ for every $\varepsilon > 0$, where $\|f\|_\infty = \sup_{0 \leq t \leq 1} |f(t)|$.

3. Show that for every non-empty open set $U \subset \Omega_0$, one has $W_0(A) > 0$. *Hint:* First note that any such U contains the ε-neighborhood of a function f which is piecewise linear, for some $\varepsilon > 0$.

Exercise 9.5.11 Brownian bridge
Let $(B_t, 0 \leq t \leq 1)$ be a standard Brownian motion. We let $(Z_t^y = yt + (B_t - tB_1), 0 \leq t \leq 1)$ for any $y \in \mathbb{R}$, and call it the *Brownian bridge* from 0 to y. Let W_0^y be the law of $(Z_t^y, 0 \leq t \leq 1)$ on $\mathcal{C}([0,1])$. Show that for any non-negative measurable function $F : \mathcal{C}([0,1]) \to \mathbb{R}_+$ for $f(y) = W_0^y(F)$, we have

$$E[F(B)|B_1] = f(B_1)\,, \quad a.s.$$

Hint: find a simple argument entailing that B_1 is independent of the process $(B_t - tB_1, 0 \leq t \leq 1)$.

Explain why we can interpret W_0^y as the law of a Brownian motion 'conditioned to hit y at time 1'.

Exercise 9.5.12
Show that the Dirichlet problem on $D = B(0,1) \setminus \{0\}$ in $\mathbb{R}^d$, with boundary conditions $g(x) = 0$ for $|x| = 1$ and $g(x) = 1$ for $x = 0$, has no solution for $d \geq 2$.

Exercise 9.5.13 Dirichlet problem in the upper-half plane
Let $\mathbb{H} = \{(x, y) \in \mathbb{R}^2 : y > 0\}$. Let $(B_t, t \geq 0)$ be a Brownian motion started from x under the probability measure P_x, and let $T = \inf\{t \geq 0 : B_t \notin \mathbb{H}\}$.

1. Determine the law of B_T under P_x whenever $x \in \mathbb{H}$.

2. Show that if u is a bounded continuous function on $\overline{\mathbb{H}}$ which is harmonic on $\mathbb{H}$, then

$$u(x, y) = \int_{\mathbb{R}} \mathrm{d}z\, u(z, 0)\, \frac{1}{\pi} \frac{y}{(x - z)^2 + y^2} \mathrm{d}z.$$

9.6 Poisson measures, ID laws and Lévy processes

Exercise 9.6.1
Prove that the Poisson law with parameter $\lambda > 0$ is the weak limit of the Binomial law with parameters $(n, \lambda/n)$ as $n \to \infty$.

A factory makes 500,000 light bulbs in a day. On an average, 4 of these are defectuous. Estimate the probability that on some given day, 2 of the produced light bulbs were defectuous.

Exercise 9.6.2 The bus paradox
Why do we always feel we are waiting a very long time before buses arrive? This exercise gives in indication of why... well, if buses arrive according to a Poisson process.

1. Suppose buses are circulating in a city day and night since ever, the counterpart being that drivers do not officiate with a timetable. Rather, the times of arrival of buses at a given bus-stop are the atoms of a Poisson measure on $\mathbb{R}$ with intensity $\theta \mathrm{d}t$, where $\mathrm{d}t$ is Lebesgue measure on $\mathbb{R}$. A customer arrives at a fixed time t at the bus-stop. Let S, T be the two consecutive atoms of the Poisson measure satisfying $S < t < T$. Show that the average time $E[T - S]$ that elapses between the arrivals of the last bus before time t and the first bus after time t is $2/\theta$. Explain why this is twice the average time between consecutive buses. Can you see *why* this is so?

2. Suppose that buses start circulating at time 0, so that arrivals of buses at the station are now the jump times of a Poisson process with intensity θ on $\mathbb{R}_+$. If the customer arrives at time t, show that the average elapsed time between the bus before (time S) and after his arrival (time T) is $\theta^{-1}(2 - e^{-\theta t})$ (with the convention $S = 0$ if no atom has fallen in $[0, t]$).

Exercise 9.6.3
Prove Proposition 7.3.2.

Exercise 9.6.4
Check the *marking property* of Poisson random measures: if $M(\mathrm{d}x) = \sum_{i \in I} \delta_{x_i}(\mathrm{d}x)$ is

a Poisson random measure on $E, \mathcal{E}$ with intensity μ, and if $y_i, i \in I$ are i.i.d. random variables with law ν on some measurable space $(F, \mathcal{F})$, and independent of M, then $\sum_{i\in I} \delta_{(x_i,y_i)}(\mathrm{d}x\,\mathrm{d}y)$ is a Poisson random measure on $E \times F$ with intensity $\mu \otimes \nu$.

Exercise 9.6.5 Brownian motion and the Cauchy process
Let $(B_t = (B_t^1, B_t^2), t \geq 0)$ be a standard Brownian motion in $\mathbb{R}^2$ (i.e. $B_0 = 0$). Recall that the Cauchy law with parameter $a > 0$ has probability distribution function $a/(\pi(a^2 + x^2)), x \in \mathbb{R}$. We let

$$C_a = \inf\{t \geq 0 : B_t^2 = -a\}\,, \qquad a \geq 0.$$

Prove that the process $(B^1_{C_a}, a \geq 0)$ is a Lévy process such that C_a is a Cauchy law with parameter a for every $a > 0$. Does this remind you of a previous exercise?

Index