ADVANCED PROBABILITY

I.F. BAILLEUL

Intentions

In a causal vision of the world, it is not clear what should be called a random Natural phenomenon. Within this framework, and at an intuitive level, probability theory quantifies our lack of knowledge on the causes of what we observe. In its relation with the empirical world, probability theory provides results of a subjective nature, and a change in our understanding of Nature may change this relation. Which mathematical (*i.e.* logical) model for random Natural phenomena should be adopted has been debated for long, and it was not before 1933 and Kolmogorov's work "*Foundations of the Theory of Probability*" that a model has been widely accepted. Building on the works of Lebesgue, Baire, Fréchet and others, Kolmogorov laid the foundations of probability theory on the ground of measure theory.

One can distinguish two levels in his theory: the random phenomenon itself is modelled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and the experimental observation process is modelled by a random variable or a family of random variables $(X_t)_{t\in T}$. The motion of a pollen grain in suspension at the surface of a glass of water will for instance be represented by the collection $(X_t)_{t\geqslant 0}$ of its positions as time goes.

The aim of this course is to introduce some of the most fundamental tools used in the study of random phenomena whose description involves infinitely many parameters.

Part I of the course tackles the problem of defining models of a given phenomenon for which experimental observations provide some constraints. The main question will thus be to define a "proper" probability on a measurable space $(\Omega, \mathcal{F})$ which assigns to some events, corresponding to the experimental events, a given probability. Two ways of constructing such probabilities will be explored: by using the general machinery of Caratheodory's extension theorem, or by constructing them as limits of other probabilities, defined in an elementary way. In this part, the (mathematical) observation process $(X_t)_{t\in T}$ will be considered globally, without paying attention to any notion of dynamics.

Part II of the course is devoted entirely to the dynamical description of a phenomenon; no attention will thus be paid to the probability space $(\Omega, \mathcal{F})$ itself. In most of the models we shall consider, $(X_t)_{t\in T}$ will be indexed by some sort of time; and time has an arrow. We shall explore in this second part what natural notions come out of this fact and some of their fundamental properties. Roughly speaking, as time passes, the observation process defines a dynamical system; like in deterministic dynamical systems, the knowledge of which quantities are preserved, increase or descrease, as time runs forward provides information on the dynamics. This role of "constant of motion" is played in the probabilistic setting by the notion of (sub/super-)martingale.

These notes are intended for use by students of the Mathematical Tripos at the University of Cambridge. Please send corrections to `i.bailleul@statslab.cam.ac.uk`.

The importance of Brownian motion in modern probability theory cannot be overstated. Not only is it the universal limit of many rescaled random walks (section 2.4), it is also the universal model for all continuous martingales, as will be seen in section 12.1. Chapter III opens with a section where we investigate the most fundamental properties of Brownian motion. To describe the most general martingales (in section 12.2) we shall introduce and study in section 11 the basic structure of Lévy processes.

Complements are added to each part, which present interesting facts related to each part; this is non-examinable material.

Contents

Part I. Static theory of stochastic processes 4
1. Construction of measures and random processes 5
1.1. Processes and sample space 5
1.2. Caratheodory's extension theorem 7
1.3. A convenient framework 9
1.4. Good modifications 14
2. Constructive approach in separable Banach spaces 16
2.1. Weak convergence on the set of probability measures on a metric space 16
2.2. Specific tools in finite dimension 17
2.3. Weak convergence in separable Banach spaces 20
2.4. Application: Universality of Brownian motion 24
3. Comments and exercises 26
3.1. References and comments 26
3.2. Exercises 28
4. Complements to part I 30
4.1. *Complement: Separable Banach spaces* 30
4.2. *Complement: Lebesgue measure on* $[0,1]$ 32
4.3. *Complement: Isomorphism of Borel probability spaces* 32
4.4. *Complement: Riesz representation theorem* 33
Part II. Dynamic theory of stochastic processes 35
5. Dynamics and filtrations 35
5.1. Conditional expectation 35
5.2. Filtrations 38
5.3. Martingales, supermartingales and submartingales 40
6. Discrete time martingale theory 41
6.1. Characterisation of supermartingales 41
6.2. Almost-sure and $\mathbb{L}^1$-convergence results 42
6.3. $\mathbb{L}^p$-convergence results 45
6.4. Applications 46
7. Continuous time martingale theory 51
8. Comments and exercises 53
8.1. Exercises 53
9. Complements to Part II 56
9.1. *Complement: Solving stochastic differential equations* 56
9.2. *Complement: Regular conditional probability* 58

Part III. Brownian motion, Lévy processes and martingales 60
10. Brownian motion 60
10.1. Different point of views on Brownian motion 60
10.2. Constructing martingales 63
10.3. Strong Markov property 65
10.4. Brownian motion and the Dirichlet problem 66
11. Lévy processes 68
11.1. Basics 68
11.2. Construction of Lévy processes 72
12. (...) and martingales 73
12.1. Representation of continuous martingales 73
12.2. Representation of general martingales 75
13. Comments and exercises 77
13.1. Exercises 77
14. Complement to Part III 79
14.1. *Complement: Infinite sums of infinitesimal independent random variables* 79
15. Solutions to the exercises 81
15.1. Exercises on part I 81
15.2. Exercises on part II 85
15.3. Exercises on part III 90
References 94
Index 95

Part I. Static theory of stochastic processes

Modern probability theory starts with the formalism of an experiment through the concept of ***abstract algebra***. This is in a sense the collection $\mathcal{Q}$ of questions we can ask about an experiment we are interested in, and which might be repeated; they are of the form: "Do you observe (that)?", shortly written "Observe (that)?" below. This collection of questions is supposed to enjoy the following *logical properties*.

- If questions "Observe (A)?" and "Observe (B)?" are in $\mathcal{Q}$ then the question "Observe (A and B)?" and "Observe (A or B)?" are meaningful and are in $\mathcal{Q}$. The following questions always have the same answers:
  - "Observe (A or (B and C))?" and "Observe ((A or B) and (A or C))?"
  - "Observe (A and (B or C))?" and "Observe ((A and B) or (A and C))?"
- $\mathcal{Q}$ contains a question "Observe ($\emptyset$)" whose answer is always "no" and a question "Observe (all)?" whose answer is always "yes". The following questions always have the same answers:
  - "Observe (A or $\emptyset$)?" and "Observe (A)?",
  - "Observe (A and $\emptyset$)?" and "Observe ($\emptyset$)?",
  - "Observe (A or all)?" and "Observe (all)?",
  - "Observe (A and all)?" and "Observe (A)?",etc.

Stone showed that any abstract logical structure as the above one can always be understood as a collection of questions of the form "Does this element of Ω belongs to A?", for some set Ω and A belonging to a collection $\mathcal{A}$ of parts of Ω stable by finite union, finite intersection, complementation, and containing the emptyset. The set $\mathcal{A}$ together with these operations is called a ***(concrete) algebra***[1]. This theorem gives a 'set representation' of the logical structure with which we comprehend Nature. As natural as it may appear, quantum mechanics has taught us that this representation has limits... and that Nature is subtler than that. Nonetheless, the benefits provided by such a view on Natural phenomena are tremendous and we shall adopt it without restriction.

We shall thus suppose given a set Ω, together with an algebra $\mathcal{A}$ of parts of Ω describing the elementary knowledge about some phenomenon we are questioning. Although no human being will ever be able to ask more than a finite number of questions during his life, it is a useful abstraction to think that since this number may be really large, we are actually able to ask countably many questions. This directly leads to the definition of a σ-algebra $\mathcal{F}$ of parts of Ω, which is the good setting in which defining a probability.

This formalism on which probability theory rests is due to Kolmogorov in his 1933 book "Foundations of the Theory of Probability". Although the advantages provided by this framework are numerous, you should keep in mind the following quotation from Kolmogorov's book on the interpretation of probability theory.

"*Even if the sets (events) of $\mathcal{A}$ can be interpreted as actual and (perhaps only approximately) observable events, it does not, of course, follow from this that the sets of the extended field $\mathcal{F}$ reasonnably admit an interpretation.*

Thus there is the possibility that while a field of probability $(\mathcal{A}, \mathbb{P})$ may be regarded as the image (idealized, however) of actual random events, the extended field of probability $(\mathcal{F}, \mathbb{P})$ will still remian merely a mathematical structure.

[1] Stone's theorem is well presented in T. Tao's post: http://terrytao.wordpress.com/2009/01/12/245b-notes-1-the-stone-and-loomis-sikorski-representation-theorems-optional.

Thus sets of $\mathcal{F}$ are generaly merely ideal events to which nothing corresponds in the outside world. However, if reasonning which utilizes the probabilities of such ideal events leads us toa determination of the probability of an actual event of $\mathcal{A}$, then, from an empirical point of view also, this determination will automatically fail to be contradictory."

1. Construction of measures and random processes

1.1. **Processes and sample space.** Interesting random natural phenomena are often described in terms of events defined by means of an infinite number of coordinates, as is the case for random sequences or random functions. They can be represented by a collection $(X_t)_{t\in T}$ of random variables[2], defined on some (potentially different) measurable space(s), and indexed by some set; the integers for random sequences, and $[0,1]$, say, for a random function from $[0,1]$ to any (measurable) space. The trajectory of a Markov chain is an important example of a process indexed by the integers.

Definition **1.** *A collection $(X_t)_{t\in T}$ of random variables, defined on some (potentially different) measurable space(s) is called a* ***process****; T will be referred to as the set of* ***coordinates****, or* ***index set****.*

In Kolmogorov's theory, a process is the mathematical abstraction of the experimental observation process. How can we define a *random* process? In practice, we generally face two kinds of situations, depending on which object is given as part of the model.

(1) A probability measure space $(\Omega, \mathcal{F}, \mathbb{P})$ is given and we have to define a process X on it satisfying some probabilistic requirements. This is sometimes easy but, more often, it requires some work as you will see in the course on stochastic calculus: defining stochastic integrals, solving stochastic differential equations are tasks of that type. We shall not encounter such a situation here, except in section 1.3.

(2) In other cases, when the measurable space $(\Omega, \mathcal{F})$ we are working with is nice enough, the definition of the process X is immediate but not the definition of a probability $\mathbb{P}$ on $(\Omega, \mathcal{F})$ which would give X the probabilistic properties we want it to have; this will be the case when we take as Ω the space of outcomes of the phenomenon under study.

This dichotomy is analogous to the situation an experimentor can meet: Given an experimental context, construct some measurement devices which will enable him/her to measure some given quantities, or, given some measurement devices, construct an experiment which will enable him/her to observe what he/she wants with his/her tools. We shall mainly explore situation **(2)** in the first part of this course, where we shall take as Ω the sample space of the phenomenon under study. It will for example be the space $\mathbb{R}^{\mathbb{N}}$ for a Markov chain on $\mathbb{R}$, the space $\mathbb{R}^{[0,1]}$ for a random function from $[0,1]$ to $\mathbb{R}$, or $\{0,1\}^E$ for the configuration space of a spin system over a set E. We first describe the σ-algebra $\mathcal{F}$ of observable events of these uncountable product spaces.

b) Product σ-algebra, or "What can we measure?". Suppose we model the experimental observation of a natural phenomenon by a collection $(X_t)_{t\in T}$ of random variables and denote by S_t the set of possible outcomes of X_t. We model the set of possible outcomes[3] of the phenomenon as a product $\prod_{t\in T} S_t$. This product space will be our Ω, with a generic element $\omega = (\omega_t)_{t\in T}$. If each set S_t has a σ-algebra of observable events, the

[2] That is, *measurable* functions from $(\Omega, \mathcal{F})$ to some $(\Omega', \mathcal{F}')$.

[3] More precisely, the set of experimentally accessible possible outcomes.

σ-algebra of observable events in the product space is generated by[4] the elementary events $\big\{(\omega_t)_{t\in T}\,;\,\omega_{t_1}\in A_1,\dots,,\omega_{t_n}\in A_n\big\}$, with $n\geqslant 1$ finite and each $A_i\in\mathcal{S}_{t_i}$; it is called the ***product σ-algebra***[5]. It will be the collection of sets to which we shall be able to associate a probability. Let us first describe this σ-algebra in some more concrete way.

• **The measurable space $\big(\mathbb{R}^{\mathbb{N}},\mathcal{B}(\mathbb{R}^{\mathbb{N}})\big)$.** Let us consider as an example the space $\mathbb{R}^{\mathbb{N}}$ of real-valued sequences; $T=\mathbb{N}$ and $S_t=\mathbb{R}$ for all $t\in T$. Introduce the metric $\rho(x,y)=\frac{|x-y|}{1+|x-y|}$ on $\mathbb{R}$; the open sets for ρ are countable unions of open intervals, as for the usual metric. Define on $\mathbb{R}^{\mathbb{N}}$ the metric

$$d(\omega,\omega')=\sum_{n\geqslant 1}2^{-n}\rho(\omega_n,\omega'_n),$$

where $\omega=(\omega_n)_{n\geqslant 1}$ and $\omega'=(\omega'_n)_{n\geqslant 1}$; the ***Borel σ-algebra*** of $\mathbb{R}^{\mathbb{N}}$ is the smallest σ-algebra of $\mathbb{R}^{\mathbb{N}}$ containing the open balls of the metric d.

LEMMA **2.** *The product σ-algebra of $\mathbb{R}^{\mathbb{N}}$ and its Borel σ-algebra coincide.*

PROOF – As both σ-algebras are defined by a collection fo elementary sets it suffices to prove that any of these is an element of the other σ-algebra. To start with, let us consider an elementary product event $B=\big\{\omega=(\omega_n)_{n\geqslant 1}\in\mathbb{R}^{\mathbb{N}}\,;\,\omega_{n(1)}\in A_1,\dots,,\omega_{n(p)}\in A_p\big\}$, with p finite and each A_i a Borel set of $\mathbb{R}$. By a monotone class argument, it suffices to consider the case where the A_i's are open intervals $(a_i-\varepsilon_i,a_i+\varepsilon_i)$. Prove that B can be written as a countable union of open balls of $\big(\mathbb{R}^{\mathbb{N}},d\big)$, in that case. To prove that open balls can be written as a union of elementary B's, mimic the 2-dimensional case, filling a circle with a union of squares[6]. $\triangleright$

• **The product space $\mathbb{R}^T$.** The following theorem shows that the product σ-algebra of any product space $\mathbb{R}^T$ is not richer than the product σ-algebra of $\mathbb{R}^{\mathbb{N}}$.

LEMMA **3.** *Let T be an uncountable set. To any event A of the product σ-algebra of $\mathbb{R}^T$ there corresponds a countable set of indices $(t_n)_{n\geqslant 1}$ and a Borel set B in $\mathbb{R}^{\mathbb{N}}$ such that*

$$A=\big\{\omega=(\omega_t)_{t\in T}\in\mathbb{R}^T\,;\,\big(\omega_{t_n}\big)_{n\geqslant 0}\in B\big\}. \tag{1.1}$$

PROOF – Denote by $\mathcal{E}$ the collection of subsets of $\mathbb{R}^T$ of the form (1.1). Given a sequence $(A_n)_{n\geqslant 1}$ of elements of $\mathcal{E}$ with corresponding indices $T^{(n)}$, set $T^{(\infty)}=\bigcup_{n\geqslant 1}T^{(n)}$; every A_n can be written

$$A_n=\big\{\omega\in\mathbb{R}^T\,;\,(\omega_{\tau_1},\omega_{\tau_2},\dots)\in B_n\big\},$$

where $\tau_i\in T^{(\infty)}$ and B_n is a Borel event of $\mathbb{R}^{\mathbb{N}}$. It follows that the collection $\mathcal{E}$ is a σ-algebra; as it contains the elementary product events, it contains the product σ-algebra. Conversely, given an event of the form (1.1), lemma 2 proves that it belongs to the product σ-algebra of $\mathbb{R}^T$; this establishes the conclusion of the theorem. $\triangleright$

[4]The σ-algebra generated by some family of parts of a set is the smallest σ-algebra containing the given family. It always exists as the family of all parts is a σ-algebra and the intersection of any collection of σ-algebras is a σ-algebra.

[5]You can think of the elementary events as rectangles in $\mathbb{R}^n$; this collection of sets is sufficient to describe all open sets, altough not all open sets are rectangles.

[6]Note that as ρ is bounded above by 1, an elementary event $\big\{(\omega_n)_{n\geqslant 1}\in\mathbb{R}^{\mathbb{N}}\,;\,\omega_{n(1)}\in A_1,\dots,\omega_{n(p)}\in A_p\big\}$ and the (infinite dimensional) cube $\big\{(\omega_n)_{n\geqslant 1}\in\mathbb{R}^{\mathbb{N}}\,;\,\omega_{n(1)}\in A_1,\dots,\omega_{n(p)}\in A_p,\omega_{n(p+1)}\in A_{p+1},\dots\big\}$ are within distance ε of one another provided p is big enough. So, fill the ball by cubes, and approximate cubes by elementary events.

• **The product space $\prod_{t\in T} S_t$.** In the general case where the sample space of the observed phenomenon is the product $\prod_{t\in T} S_t$ of possibly uncountably many measurable spaces $(S_t, \mathcal{S}_t)$, the description of its product σ-algebra $\bigotimes_{t\in T} \mathcal{S}_t$ is similar to the case of $\mathbb{R}^T$. Given a *countable* subset S of T, denote by $\mathcal{B}_S$ the σ-algebra on $\prod_{s\in S} S_s$ generated by its elementary product events. The proof of the following fact is identical to the proof of lemma 3.

THEOREM 4 (***Product σ-algebra***). *To any event A of the product σ-algebra of $\prod_{t\in T} S_t$ there corresponds a* countable *set S of indices and a measurable set B in $\prod_{s\in S} S_s$ such that*

$$A = \Big\{\omega = (\omega_t)_{t\in T} \in \prod_{t\in T} S_t \,;\, (\omega_s)_{s\in S} \in B\Big\}. \tag{1.2}$$

As the maps

$$X_t : \omega = (\omega_s)_{s\in T} \in \prod_{t\in T} S_t \mapsto \omega_t, \quad t \in T,$$

are measurable, by definition of the product σ-algebra, we define a process on the measurable space $\big(\prod_{t\in T} S_t, \bigotimes_{t\in T} \mathcal{S}_t\big)$ setting $X = (X_t)_{t\in T}$. It is called the ***coordinate process***.

DEFINITION **5.** *Our empirical knowledge of the investigated phenomenon provides us with an a priori set of values for the probability of the elementary events: $\mathbb{P}\big(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n\big)$. These quantities are called the **finite-dimensional laws** (or **distributions**) of the process.*

Under proper conditions, Caratheodory's theorem below gives us a mean to define the probability of any event of the product σ-algebra in an unambiguous way out of these quantities only.

1.2. **Caratheodory's extension theorem.** The main tool to construct abstractly probability measures is Caratheodory's extension theorem, of which we give a proof following J.L. Doob's exposition, in his book [Doo94]. Starting with the a priori datum of the "probability" of elementary events[7], it gives a sufficient condition under which this set function can be extended to a bigger set of events. Recall that an additive set function μ on an algebra is a real-valued function such that $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever A and B are disjoint elements of the algebra.

THEOREM **6** (Caratheodory's extension theorem). *Let $(\Omega, \mathcal{F})$ be a measurable space, $\mathcal{A} \subset \mathcal{F}$ be an algebra and $\mu : \mathcal{A} \to [0,1]$ be an* additive *function such that*

i) $\mu(\emptyset) = 0, \mu(\Omega) = 1$,

ii) (countable additivity on $\mathcal{A}$) If $A_1, A_2, \dots$ are disjoint sets of $\mathcal{A}$ with union in $\mathcal{A}$ then $\mu\big(\bigcup_{n\geqslant 1} A_n\big) = \sum_{n\geqslant 1} \mu(A_n)$.

Then μ has a unique extension into a probability measure on $\sigma(\mathcal{A})$.

Note that condition *ii)* is equivalent to condition

ii)' For any sequence $(A_n)_{n\geqslant 0}$ of sets of $\mathcal{A}$ decreasing to $\emptyset$ we have $\mu(A_n) \to 0$.

[7]Given by repeated measurements in a fixed experimental context.

Proof – *Uniqueness.* The collection of elements of $\sigma(\mathcal{A})$ on which two possible extensions coincide being a σ-algebra the two measures are equal on $\sigma(\mathcal{A})$ if they coincide on $\mathcal{A}$ by the monotone class theorem[8].

Existence. Denote by $\mathfrak{P}(\Omega)$ the family of subsets of Ω. The outer measure $\overline{\mu}$ associated with μ is a set function defined on $\mathfrak{P}(\Omega)$ by the formula

$$\overline{\mu}(B) = \inf\Big\{\sum_{n\geqslant 0}\mu(A_n)\,;\, B\subset\bigcup_{n\geqslant 0}A_n,\ \ A_n\in\mathcal{A}\Big\}.$$

$\overline{\mu}$ is easily seen to be increasing and countably sub-additive: $\overline{\mu}\big(\bigcup_{n\geqslant 0}B_n\big)\leqslant\sum_{n\geqslant 0}\overline{\mu}(B_n)$, for any sequence $(B_n)_{n\geqslant 0}$ of sets of Ω. Also, *as μ is countably additive on $\mathcal{A}$* we see[9] that $\mu(A)\leqslant\overline{\mu}(A)$ for $A\in\mathcal{A}$; as the converse inequality trivially holds, $\overline{\mu}$ and μ coincide on $\mathcal{A}$. Check that we define a pseudo-metric[10] on $\mathfrak{P}(\Omega)$ setting[11]

$$d(B,C) = \overline{\mu}(B\Delta C);$$

since $\overline{\mu}$ is sub-additive and $B\subset\big(B\Delta C\big)\cup C$ for any subsets B,C of Ω, we have

$$|\overline{\mu}(B)-\overline{\mu}(C)|\leqslant\overline{\mu}(B\Delta C)=d(B,C),$$

so $\overline{\mu}(B)=\overline{\mu}(C)$ if $d(B,C)=0$. Define $\mathcal{A}^{\overline{\mu}}$ as the collection of subsets B of Ω which can be approximated to any accuracy by elements of $\mathcal{A}$, using d-(pseudo)-distance.

Lemma **7.** *$\mathcal{A}^{\overline{\mu}}$ is a σ-algebra on which $\overline{\mu}$ is additive.*

Proof – • We start by proving the (finite) additivity of $\overline{\mu}$ on $\mathcal{A}^{\overline{\mu}}$ as we are going to use that fact in the proof that $\mathcal{A}^{\overline{\mu}}$ is a σ-algebra. Take two disjoint sets B and C in $\mathcal{A}^{\overline{\mu}}$, an $\epsilon>0$, and let A_B and A_C be elements of $\mathcal{A}$ such that $d\big(B,A_B\big),\ d\big(C,A_C\big)\ \leqslant\epsilon$. As they satisfy the inequality $\overline{\mu}(A_B\cap A_C)\leqslant 2\epsilon$, we have by sub-additivity of $\overline{\mu}$

$$\max\Big(d\big(B,A_B\backslash(A_B\cap A_C)\big),\ d\big(C,A_C\backslash(A_B\cap A_C)\big)\Big)\leqslant 3\epsilon.$$

It follows that

$$\begin{aligned}\overline{\mu}(B)+\overline{\mu}(C)\geqslant\overline{\mu}(B\cup C)&\geqslant\overline{\mu}\big(A\backslash(A_B\cap A_C)\,\cup\,A\backslash(A_B\cap A_C)\big)-3\epsilon\\&\geqslant\overline{\mu}(B)+\overline{\mu}(C)-5\epsilon,\end{aligned}$$

from which we get the conclusion as $\epsilon>0$ is arbitrary.

• $\mathcal{A}^{\overline{\mu}}$ is clearly stable by complementation; we check that $\mathcal{A}^{\overline{\mu}}$ is stable by countable disjoint union, this implies that $\mathcal{A}^{\overline{\mu}}$ is stable by countable union or intersection, and so is a σ-algebra. Given $\epsilon>0$ and a sequence $(B_n)_{n\geqslant 0}$ of disjoint elements of $\mathcal{A}^{\overline{\mu}}$; associate to each B_n an $A_n\in\mathcal{A}$ such that $d(A_n,B_n)\leqslant 2^{-n-1}\epsilon$. As $\overline{\mu}$ is finitely additive on $\mathcal{A}^{\overline{\mu}}$ we have $\sum_{n=0}^N\overline{\mu}(B_n)=\overline{\mu}\Big(\bigcup_{n=0}^N B_n\Big)\leqslant 1$, for all $N\geqslant 0$, so the sum $\sum_{n\geqslant N+1}\overline{\mu}(B_n)$ is less than ϵ for N large enough. For such a choice of N

$$\begin{aligned}d\Big(\bigcup_{n\geqslant 0}B_n,\bigcup_{n=0..N}A_n\Big)&\leqslant d\Big(\bigcup_{n\geqslant 0}B_n,\bigcup_{n=0..N}B_n\Big)+d\Big(\bigcup_{n=0..N}B_n,\bigcup_{n=0..N}A_n\Big)\\&\leqslant\overline{\mu}\Big(\bigcup_{n\geqslant N+1}B_n\Big)+\overline{\mu}\Big(\Big\{\bigcup_{n=0..N}B_n\Big\}\Delta\Big\{\bigcup_{n=0..N}A_n\Big\}\Big)\\&\leqslant\epsilon+\overline{\mu}\Big(\bigcup_{n=0..N}(B_n\Delta A_n)\Big)\leqslant\epsilon+\sum_{n=0}^N 2^{-n-1}\epsilon\leqslant 2\epsilon.\end{aligned}$$

[8] Algebras are π-systems.

[9] Replacing A_n by $A_n\backslash\bigcup_{k=0}^n A_k$ if necessary, we can suppose that the A_n's are disjoint.

[10] A metric for which two elements at null distance are not necessarily equal.

[11] $B\Delta C:=(B\cup C)\backslash(B\cap C)$.

As $\epsilon > 0$ can be chosen arbitrarily small this proves that $\bigcup_{n\geqslant 0} B_n$ is an element of $\mathcal{A}^{\overline{\mu}}$. $\odot$

$\overline{\mu}$ being increasing, additive and sub-countably-additive on $\mathcal{A}^{\overline{\mu}}\big(\supset \sigma(\mathcal{A})\big)$, it is countably-additive on $\mathcal{A}^{\overline{\mu}}$ (can you see why?); its restriction to $\sigma(\mathcal{A})$ provides the desired extension of μ. $\triangleright$

Despite its elegance, Caratheodory's theorem does not rule out all the difficulties as their remains to check conditions *i)* and *ii)* (or *ii)'*) if one wants to use it. The introduction of the following framework will help greatly in that task; it also provides a framework in which the use of Caratheodory's theorem is not necessary in some concrete situations.

1.3. **A convenient framework.** Although Caratheodory's extension theorem is a fantastic tool to construct probability measures as models of random phenomena, most of the time, it is not necessary to resort to the full strength of this abstract machinery as additional features can help us in our construction task. Indeed, problems can often be set in a topological framework where Ω is a topological space and $\mathcal{F}$ the σ-algebra generated by its open sets.

Definition **8.** *We say that two **measurable spaces** are **isomorphic** if$_{def}$ there exists a measurable bijection from one to the other with a measurable inverse*[12].

The interval $[0,1]$ will be equiped with its Borel σ-agebra $\mathcal{B}\big([0,1]\big)$, generated by the open sets.

Definition **9.** *A measurable space $(\Omega,\mathcal{F})$ is said to be a **Borel space** if$_{def}$ it is isomorphic to a measurable subset of* $[0,1]$.

Construction problems in Borel spaces are nothing more than constructions problems in the innocent framework $\big([0,1],\mathcal{B}\big([0,1]\big)\big)$. But powerful tools are available on the space $[0,1]$ which are not available in an abstract measurable space (basically compactness!, *i.e.* existence of limits of subsequences). We shall illustrate this fact in theorems 18 below where it is used together with Caratheodory's machinery to prove a general existence result. It will also be the framework of the approximation theory developped in section 2.

Theorem 10 below should convince you that the class of Borel spaces should be sufficient for your needs before long. It is proved in the Complement *Separable Banach spaces.*

Theorem **10.** *Any measurable subset of a separable Banach space is a Borel space.*

a) A first application: existence of sequences of independent random variables, construction of Markov chains. As a first example of how this property of a space can be used, let us see how one can construct on $[0,1]$, with Lebesgue measure Leb, a sequence of independently distributed random variables with values in some Borel spaces. As a first step let us construct a real-valued random variable with any given distribution. Given a probability measure μ in $\mathbb{R}$ denote by $F:\mathbb{R}\to[0,1]$ its distribution function $F(t)=\mu\big((-\infty,t]\big)$ and by $G:[0,1]\to\mathbb{R}$ its right inverse

$$G(u)=\inf\{t\in\mathbb{R}\,;\,F(t)\geqslant u\}$$

[12] In the same way as the inverse of a one-to-one continuous function may be non-continuous (can you find a counter-example?), the inverse of a one-to-one measurable function may be non-measurable.

with the convention that $\inf\emptyset = +\infty$; this is a càdlàg function characterized by the property[13]

$$u \leqslant F(t) \text{ iff } G(u) \leqslant t.$$

So if U is a uniform random variable in $[0,1]$

$$\mathbb{P}\big(G(U) \leqslant t\big) = \mathbb{P}\big(F(t) \geqslant U\big) = F(t).$$

THEOREM **11** (Existence of independent sequences)**.** *Given probability measures μ_n on some Borel spaces $(S_n, \mathcal{S}_n), n \geqslant 1$, we can construct on $\big([0,1], \mathcal{B}([0,1]), \mathrm{LEB}\big)$ a sequence $(X_n)_{n\geqslant 0}$ of independent random variables with respective distributions μ_n.*

PROOF – It is given under the form of an exercise.

(1) Given a uniform random variable U on $[0,1]$ prove that the sequence of its binary expension is a Bernoulli sequence with parameter $\frac{1}{2}$.
(2) Deduce that there exists measurable functions $f_1, f_2, \dots$ from $[0,1]$ to itself such that the $f_n(U)$ are iid uniform on $[0,1]$.
(3) Let φ_i be an isomorphism between $(S_i, \mathcal{S}_i)$ and a Borel subset of $[0,1]$; define the probability ν_i on $[0,1]$ setting $\nu_i(A) = \mu_i\big(\varphi_i^{-1}(A)\big)$. Set, for $t \in [0,1]$

$$f_i(t) = \sup\big\{x \in [0,1]\,;\, \nu_i\big([0,x]\big) < t\big\}.$$

Why is this function measurable? Prove that if V is uniformly distributed in $[0,1]$ then $f_i(V)$ has law ν_i. Finish the proof.

$\triangleright$

As a by-product of the above result we are able to construct effectively any Markov chain in a proper way. Suppose we are given for each $x \in \mathbb{R}$ a probability measure $p(x,.)$ on $\mathbb{R}$.[14]

DEFINITION **12.** *A* ***discrete time Markov chain with transition kernel*** $\{p(x,.)\}_{x\in\mathbb{R}}$ ***and initial distribution*** ν *is a process $(X_n)_{n\geqslant 0}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that we have for any $n \geqslant 0$, and any (Borel) sets $A_0, \dots, A_n$ of $\mathbb{R}$*

$$\mathbb{P}\big(X_0 \in A_0, \dots, X_n \in A_n\big) = \int \cdots \int \nu(dx_0)\, \mathbf{1}_{A_1}(x_1) p(x_0, dx_1) \cdots \mathbf{1}_{A_n}(x_n) p(x_{n-1}, dx_n).$$

PROPOSITION **13.** *Given any transition kernel $\big\{p(x,.)\big\}_{x\in\mathbb{R}}$ and any initial distribution ν their exist a Markov chain with the corresponding characteristics.*

PROOF – Denote by g and f_x the right inverses of the distribution functions of ν and μ_x respectively,

$$g(u) = \inf\big\{z \in \mathbb{R}\,;\, \nu\big((-\infty, z]\big) \geqslant u\big\}, \quad f_x(u) = \inf\big\{z \in \mathbb{R}\,;\, p\big(x, (-\infty, z]\big) \geqslant u\big\}, \quad u \in [0,1],$$

and let $(U_n)_{n\geqslant 0}$ be a sequence of iid uniform random variables on $[0,1]$, whose existence is guaranteed by theorem 11. I leave you to check that the induction formula $X_0 = g(U_0)$ and

$$X_{n+1} = f_{X_n}(U_n)$$

defines a Markov chain with transition kernel $\big\{p(x,\cdot)\big\}_{x\in\mathbb{R}}$ and initial distribution ν.[15] $\triangleright$

[13] The left inverse of F is defined by the formula $H(u) = \sup\{t \in \mathbb{R}\,;\, F(t) < u\}$. Let D be the image by F of the countable collection of intervals where f is constant. The two inverses G and H coincide outside D.

[14] The construction below works equally well with any Borel space as a state space of the Markov chain.

[15] To be really clean we should make the hypothesis that $p(x,.)$ depends measurably on x, a detail which we shall leave aside.

Theorem 11 provides us with a reservoir of iid random variables; they can be used not only to construct discrete time random processes, as Markov chains, but also continuous time random processes.

b) A second application: Wiener measure and Brownian motion. The space $\mathcal{C}\big([0,1],\mathbb{R}\big)$ can be seen from two point of views, either as a subset of the product $\mathbb{R}^{[0,1]}$, or as a metric space $\big(\mathcal{C}\big([0,1],\mathbb{R}\big),\|.\|_\infty\big)$. Each picture has its own σ-algebra of observable events: the trace on $\mathcal{C}\big([0,1],\mathbb{R}\big)$ of the product σ-algebra, and the Borel σ-algebra of $\big(\mathcal{C}\big([0,1],\mathbb{R}\big)\|.\|_\infty\big)$, generated by the open balls. We shall prove later, in proposition 32, that the two σ-algebras coincide, making it the natural object to consider; denote it by $\mathcal{W}$ and write W for $\mathcal{C}\big([0,1],\mathbb{R}\big)$. Refering to §**a)** of the introduction, we construct in this paragraph a continuous time random process using the point of view **(2)**: the coordinate process $(X_t)_{t\in[0,1]}$ is naturally defined on $(W,\mathcal{W})$ setting $X_t : \omega \in W \mapsto \omega_t$ for each $t \in [0,1]$. So, turning X into a *random* process amounts to constructing a probability measure $(W,\mathcal{W})$. We construct here what it probably the most fundamental of all such measures: Wiener measure.

DEFINITION **14.** *A **Wiener measure** on $(W,\mathcal{W})$ is a probability measure $\mathbb{P}$ such that*

- $X_0 = 0$, *$\mathbb{P}$-almost-surely ,*
- *the process X has independent increments,*
- $X_t - X_s \sim \mathcal{N}(0, t-s)$ *for all $s < t$.*

THEOREM **15.** *There exists a unique Wiener measure on $(W,\mathcal{W})$.*

The uniqueness statement comes from the fact that the above three conditions define uniquely the probability of the elementary events $\big\{X_{t_1} \in A_1, \cdots, X_{t_n} \in A_n\big\}$, for (Borel) subsets A_i of $\mathbb{R}$ (can you see why?). As these events generate the product σ-algebra, which coincides with $\mathcal{W}$, the probability $\mathbb{P}$, if it exists, is uniquely determined by its values on these elementary events.

Denote by $\mathbb{D}$ the set of dyadic rationals in $[0,1]$ and write $\mathbb{D}_n$ for $\{k2^{-n}\,;\, k = 0..2^n\}$. The following existence proof of Wiener measure takes advantage of the following two facts.

- If one can construct on *some* probability space $(\Omega,\mathcal{F},\mathbb{Q})$ an almost-surely continuous process Y satisfying $\mathbb{Q}$-almost-surely some requirements then, denoting by $\mathbb{P}$ the image measure of $\mathbb{Q}$ by Y[16], the coordinate process X on $(W,\mathcal{W})$ satisfies $\mathbb{P}$-almost-surely the same requirements.
- It is easy to construct a "Wiener measure" on the space of functions from $\mathbb{D}_n$ to $\mathbb{R}$, for any $n \geqslant 1$.

PROOF – Use theorem 11 to construct on the probability space $\big([0,1],\mathcal{B}\big([0,1]\big),\textsc{Leb}\big)$ a countable collection $\big\{X_i^n\,;\, 1 \leqslant i \leqslant 2^{n-1}, n \geqslant 1\big\}$ of centered Gaussian random variables with variance 1. Define inductively a sequence $B_t^{(n)}$ of random continuous functions specifying their values on the points of $\mathbb{D}_n$ and interpolating linearly in between.

- $B^{(0)}(0) = 0$ and $B^{(0)}(1) = X^0$;
- supposing $B^{(n-1)}$ has been constructed and has independent $\mathbb{D}_{n-1}$-increments

$$\Big\{B^{(n-1)}_{(k+1)2^{-(n-1)}} - B^{(n-1)}_{k2^{-(n-1)}}\,;\, 0 \leqslant k \leqslant 2^{-(n-1)} - 1\Big\},$$

[16] $\mathbb{P}$ is a probability on $(W,\mathcal{W})$.

set $B_t^{(n)} = B_t^{(n-1)}$ for all $t \in \mathbb{D}_{n-1}$, and for $s = k2^{-(n-1)} + 2^{-n}$ set

$$B_s^{(n)} = \frac{1}{2}\Big(B^{(n-1)}_{k2^{-(n-1)}} + B^{(n-1)}_{(k+1)2^{-(n-1)}}\Big) + 2^{-\frac{n+1}{2}} X_k^n, \quad 1 \leqslant k \leqslant 2^n$$

The increments $B_s^{(n)} - B^{(n)}_{k2^{-(n-1)}}$ and $B^{(n)}_{(k+1)2^{-(n-1)}} - B_s^{(n)}$ being Gaussian, we check their independence showing they have null covariance; they have variance 2^{-n}. These two increments being constructed from $B^{(n-1)}_{k2^{-(n-1)}} - B^{(n-1)}_{(k+1)2^{-(n-1)}}$ and X_k^n they are independent of increments over intervals disjoint from $\big(k2^{-(n-1)}, (k+1)2^{-(n-1)}\big)$. An increment $B_t^{(n)} - B_s^{(n)}$ will thus have a centered Gaussian law with variance $t-s$, for $t, s \in \mathbb{D}_n$.
Now, by Borel-Cantelli's lemma, for any $c > \sqrt{2\log 2}$ there exists Leb-almost-surely an integer n_0 such that for all $n \geqslant n_0$ and all $0 \leqslant k \leqslant 2^n$ we have $|X_k^n| \leqslant c\sqrt{n}$. For such n's we thus have $\|B^{(n)} - B^{(n-1)}\|_\infty \leqslant c\sqrt{n}\,2^{-\frac{n}{2}}$, from which it follows that the sequence of continuous functions $\big(B^{(n)}\big)_{n\geqslant 0}$ converges almost-surely uniformly to some continuous (random) function $(B_t)_{t\in[0,1]}$. It is defined on the probability space $\big([0,1], \mathcal{B}\big([0,1]\big), \mathrm{Leb}\big)$. We check that the process B has independent Gaussian increments; this proves the existence of Wiener measure by the remarks preceding the beginning of the proof.
Given times $0 \leqslant t_0 < t_1 < \cdots < t_n$, approximate each t_i by a sequence t_i^k of dyadics. Write $\mathbb{E}_L$ for the expectation under Lebesgue measure. Use bounded convergence and the Leb-almost-sure continuity of B to write for any real-valued bounded continuous function f on $\mathbb{R}^n$

$$\begin{aligned}
\mathbb{E}_L\big[f(B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}})\big] &= \lim_{k+\infty} \mathbb{E}_L\big[f(B_{t_1^k} - B_{t_0^k}, \dots, B_{t_n^k} - B_{t_{n-1}^k})\big] \\
&= \lim_{k+\infty} \mathbb{E}_L\Big[f\Big(\sqrt{t_1^k - t_0^k}\,N_1, \dots, \sqrt{t_n^k - t_{n-1}^k}\,N_n\Big)\Big] \\
&= \mathbb{E}_L\big[f\big(\sqrt{t_1 - t_0}N_1, \dots, \sqrt{t_n - t_{n-1}}N_n\big)\big].
\end{aligned}$$

where $N_1, \dots, N_n$ are iid $\mathcal{N}(0,1)$ defined on $\big([0,1], \mathcal{B}\big([0,1]\big), \mathrm{Leb}\big)$. One reads on that formula that B has independent Gaussian increments. $\triangleright$

Definition **16.** *A $(W, \mathcal{W})$-valued random variable defined on some probability space is said to be a* ***Brownian motion*** *if its law is Wiener measure.*

c) Existence of random sequences and random processes. The preceding two paragraphs make it clear that it is not always necessary to resort to Caratheodory's extension theorem to define interesting random processes. Yet, it remains the best tool to deal with more general and abstract situations. As emphasized at the end of section 1.2, on eis left with checking the non-trivial condition *ii)* or *ii)'* of Caratheodory's theorem if one wants to apply it. Borel spaces provide a good framework in which proving *ii)'*, or rather its contraposition. This is typically done as follows.

Given a decreasing sequence $(A_n)_{n\geqslant 0}$ such that $\mu(A_n)$ is bounded below by some positive constant ε, approximate each A_n from inside by a compact K_n. A careful choice gives a decreasing sequence of "compact" sets whose measure is bounded below by $\frac{\varepsilon}{2}$. The intersection of finitely many of them having positive ("pre-")measure is thus non-empty so, by compactness, their intersection is non-empty; hence $\bigcap_{n\geqslant 0} A_n \supset \bigcap_{n\geqslant 0} K_n \neq \emptyset$, which proves *ii)'*. These "compact" sets are what the "Borel hypothesis" provides us with.

We are going to illustrate this approach in the following framework, which is well suited to deal with (Markov chains and more) general random sequences.

DEFINITION **17.** *Given measurable spaces $(S_i, \mathcal{S}_i)$, we say that a* ***sequence of probability measures*** *μ_n on $\prod_{i=0..n} S_i$ (^{17}) is* ***projective*** *if$_{def}$*

$$\mu_{n+1}(\cdot \times S_{n+1}) = \mu_n(\cdot), \quad n \in \mathbb{N}.$$

Projective families of probabilities are models of discrete time random processes with memory. If for instance all the S_i's are identical, equal to S, and $(\mu_n)_{n \geqslant 0}$ is determined by a family of transition kernels[18] $\big\{\mu_x(.)\,;\, x \in S\big\}$ via the formula

$$\mu_n(A_0 \times \cdots \times A_n) = \int_{A_0} \nu_0(dx_0) \int_{A_1} \mu_{x_0}(dx_1) \ldots \int_{A_n} \mu_{x_{n-1}}(dx_n),$$

then μ_n is the law of the first n positions of a Markov chain on $(S, \mathcal{S})$. In the above general model the law of the $(n+1)^{\text{th}}$-position of the process may depend not only on the n^{th} position of the process but also on all its history up to time n. Equip the infinite product $\prod_{i \geqslant 0} S_i$ with its product σ-algebra; denote by $\mathcal{S}_0 \otimes \cdots \otimes \mathcal{S}_n$ the product σ-algebra of $\prod_{i=0..n} S_i$.

THEOREM **18** (Existence of random sequences – Daniell)**.** *Let $\big((S_i, \mathcal{S}_i)\big)_{i \geqslant 0}$ be a sequence of Borel spaces. Given a projective sequence of probability measures μ_n on $\prod_{i=0..n} S_i$, there exists a probability $\mathbb{P}$ on the product σ-algebra of $\prod_{i \geqslant 0} S_i$ such that*

$$\mathbb{P}\Big(E \times \prod_{i \geqslant n+1} S_i\Big) = \mu_n(E)$$

for any $E \in \mathcal{S}_0 \otimes \cdots \otimes \mathcal{S}_n$, and $n \in \mathbb{N}$.

PROOF – We use Caratheodory's extension theorem; point *i)* is clear. The algebra $\mathcal{A} = \big\{E \times \prod_{i \geqslant n+1} S_i\,;\, E \in \mathcal{S}_0 \otimes \cdots \otimes \mathcal{S}_n\big\}$ generates the product σ-algebra of $\prod_{i \geqslant 0} S_i$. Let $(A_n)_{n \geqslant 0}$ be a decreasing sequence of elements of $\mathcal{A}$; we can suppose without loss of generality that $A_n = E_n \times \prod_{i \geqslant n+1} S_i$ and $E_n \in \bigotimes_{i=0..n} \mathcal{S}_i$. We prove the contraposition of condition *ii)'* of Caratheodory's theorem: if $\mathbb{P}(A_n) = \mu_n(A_n)$ is bounded below by some positive δ then $\bigcap_{n \geqslant 0} A_n$ cannot be empty. Let $\varepsilon > 0$ be given.

Denote by φ_n an isomorphism of $\prod_{i=0..n} S_i$ to a Borel subset of $[0,1]$ and denote by ν_n the image measure of μ_n by φ_n. As ν_n is inner regular (exercice 5), there exists a compact subset K_n of $\varphi_n(E_n)$ such that

$$\nu_n\big(\varphi(E_n) \backslash K_n\big) \leqslant 2^{-n}\epsilon,$$

i.e.

$$\mathbb{P}\Big(A_n \backslash \Big\{\varphi_n^{-1}(K_n) \times \prod_{i \geqslant n+1} S_i\Big\}\Big) \leqslant 2^{-n}\epsilon.$$

Writing V_n for $\varphi_n^{-1}(K_n) \times \prod_{i \geqslant n+1} S_i$ and setting $W_n = V_1 \cap \cdots \cap V_n$, it follows that

$$\mathbb{P}\big(A_n \backslash W_n\big) \leqslant \epsilon,$$

so W_n cannot be empty provided $\varepsilon < \delta$ (it has positive probability under that condition). Choose for each n a point m_n in W_n. As the sets W_n are decreasing, all the points $m_{n+1}, m_{n+2}, \ldots$ belong to W_n, and the φ_n-projection of their first n coordinates lie in the compact K_n, so have a converging sub-sequence. A diagonal extraction then provides a

[17] Equiped with its product σ-algebra $\mathcal{S}_0 \otimes \cdots \otimes \mathcal{S}_n$.

[18] $p(x, \cdot)$ is for every $x \in S$ a probability on $(S, \mathcal{S})$; the quantity $p(x, A)$ represents the probability starting from x to jump in A. For a discrete space S the matrix $\{p(x,y)\}_{x,y \in S}$ is the usual transition matrix of a Markov chain.

subsequence of m_n converging to a point m belonging to all the $W_n (\subset A_n)$, proving that $\bigcap_{n\geqslant 0} A_n$ is not empty. ▷

Kolmogorov's general existence theorem below gives a version of Daniell's theorem 18 which works on any product space $\prod_{t\in T} S_t$. The conceptual improvement is almost-null as a generic element of the product σ-algebra of $\prod_{t\in T} S_t$ is defined by requirements on *countably* many coordinates, as theorem 4 makes it clear. Kolmogorov's theorem is thus an almost-straightforward consequence of Daniell's theorem; details of its proof can be found in the proof of theorem 6.16, in Kallenberg's book [Kal02].

Given finite sets of indices $I \subset J$, denote by the same letter A an event of $\prod_{t\in I} S_t$, considered also as an event of $\prod_{t\in J} S_t$. Denote by $\mathbf{T}$ the set of finite subsets of T. A family of measures μ_I on $\prod_{t\in I} S_t$, $I \in \mathbf{T}$, is said to be *projective* if$_{\text{def}}$ $\mu_J(A) = \mu_I(A)$ for any event A as above, and any finite sets of indices $I \subset J$.

THEOREM **19** (Existence of processes – Kolmogorov)**.** *Let T be any index set and $(S_t)_{t\in T}$ be a family of Borel spaces. Given a projective family of probability measures μ_I on $\prod_{t\in I} S_t$ there exists a unique probability measure on $\prod_{t\in T} S_t$ with projection μ_I on each $\prod_{t\in I} S_t$, $I \in \mathbf{T}$.*

d) Limits of the abstract machinery. However powerful such general results may be, they remain unsufficient to provide models of real-valued *continuous* random paths. Try for instance to define such a process $X = (X_t)_{t\in[0,1]}$ as a random variable with values in $\mathbb{R}^{[0,1]}$ equiped with its product σ-algebra.

PROPOSITION **20.** *The subset $\mathcal{C}\big([0,1],\mathbb{R}\big)$ of $\mathbb{R}^{[0,1]}$ is not measurable*[19].

PROOF – The main reason for this is that the concept of continuity involves a continuum of conditions whereas any elements on the product σ-algebra contains only information on what happens at countably many times. Use theorem 4 to give a neat proof. ▷

To overcome this difficulty in defining random continuous functions as models of random Natural phenomena we shall rely on the idea that continuous functions are determined by their values on a countable set of times. Our construction of Wiener measure and Brownian motion relied on this idea. The next section gives a clear example of this philosophy; later, in section 6 on martingales, we shall construct continuous time martingales from their rational skeleton...

1.4. **Good modifications.**

DEFINITION **21.**
- *Two random processes $(X_t)_{t\in T}$ and $(\widetilde{X}_t)_{t\in T}$, indexed by the same set T of indices, are said to be a* ***modification*** *of one another if they have the same finite dimensional laws*[20]*: $\mathbb{P}(\widetilde{X}_t = X_t) = 1$ for any $t \in T$.*
- *$(X_t)_{t\in T}$ and $(\widetilde{X}_t)_{t\in T}$ are said to be* ***indinstiguishable*** *if $\mathbb{P}(\forall\, t \in T, \widetilde{X}_t = X_t) = 1$.*

The previous definition assumes that the event $\{\forall\, t \in T, \widetilde{X}_t = X_t\}$ is measurable, which does not hold for any index set or any pair of processes. This notion will only be used in a context where this problem does not happen. To be indistinguishable is a much stronger requirement than to be modifications of one another; however, these two notions coincide

[19]That is, it does not belong to the product σ-algebra.

[20]This condition is sufficient to have $\mathbb{P}(\forall i,\ \widetilde{X}_{t_i} = X_{t_i}) = 1$, for any finite collection of indices t_i.

if the index set is countable, or if the two processes are right-continuous with value in a Hausdorff topological space; prove that fact.

Caratheodory's theorem typically provides us with processes for which natural requirements, like continuity of the sample paths, have no meaning. Yet, if this process can be controlled in some way, it admits a modification with good sample paths properties. From an experimental point of view, working with a given process or a modification of it does not make any difference as the only quantities we can measure are the elementary probabilities $\mathbb{P}\big(X_{t_0} \in A_0, \dots, X_{t_n} \in A_n\big)$, whose values do not depend on which modification of X we are working with. As you will see in exercise ??, two processes which are modification of one another may have quite different pathwise properties; this leaves some freedom to choose the best version of a process for our needs.

This section provides the basic example of a modification procedure due to Kolmogorov. Recall we denote by $\mathbb{D}$ the set of dyadic rationals in $[0,1]$ and write $\mathbb{D}_n$ for $\{k2^{-n}\,;\,k = 0..2^n\}$.

Theorem **22** (Kolmogorov's criterion). *Let $p \geqslant 1$ and $\beta > 1/p$. Suppose $X = (X_t)_{t\in\mathbb{D}}$ is a real-valued process defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and such that*

$$\mathbb{E}\big[|X_s - X_t|^p\big] \leqslant C\,|s-t|^{p\,\beta}, \quad \textit{for all } s,t \in \mathbb{D}$$

for some finite constant C. Then, for all $\alpha \in \big[0, \beta - \frac{1}{p}\big)$, there exists a random variable *$C_\alpha \in \mathbb{L}^p$ such that one has almost surely*

$$|X_s - X_t| \leqslant C_\alpha |s-t|^\alpha, \quad \textit{for all } s,t \in \mathbb{D}.$$

As a consequence, and given any $\alpha \in \big[0, \beta - \frac{1}{p}\big)$, the process X has an α-Hölderian modification defined on $[0,1]$.

Proof – For $s,t \in \mathbb{D}$ with $s<t$, let $m \geqslant 0$ be the only integer such that $2^{-(m+1)} < t-s \leq 2^{-m}$. The interval $[s,t)$ contains at most one interval $\big[r_{m+1}, r_{m+1} + 2^{-(m+1)}\big)$ with $r_{m+1} \in \mathbb{D}_{m+1}$. If so, each of the intervals $[s, r_{m+1})$ and $\big[r_{m+1} + 2^{-(m+1)}, t\big)$ contains at most one interval $\big[r_{m+2}, r_{m+2} + 2^{-(m+2)}\big)$ with $r_{m+2} \in \mathbb{D}_{m+2}$. Repeating this remark up to exhaustion of the dyadic interval $[s,t)$ by such dyadic sub-intervals, we see that

$$|X_t - X_s| \leqslant 2 \sum_{n \geq m+1} S_n,$$

where $S_n = \sup_{t\in\mathbb{D}_n} |X_{t+2^{-n}} - X_t|$. So we have

$$\frac{|X_t - X_s|}{(t-s)^\alpha} \leqslant 2 \sum_{n \geqslant m+1} S_n\, 2^{(m+1)\alpha} \leqslant C_\alpha$$

where $C_\alpha = 2\sum_{n\geqslant 0} 2^{n\alpha} S_n$. But as

$$\mathbb{E}[S_n^p] \leqslant \mathbb{E}\Big[\sum_{t\in\mathbb{D}_n} |X_{t+2^{-n}} - X_t|^p\Big] \leqslant 2^n C (2^{-n})^{p\,\beta},$$

it follows that

$$\|C_\alpha\|_p \leqslant 2 \sum_{n\geqslant 0} 2^{n\alpha} \|S_n\|_p \leqslant 2C \sum_{n\geqslant 0} 2^{\left(\alpha - \beta + \frac{1}{p}\right)n} < \infty,$$

which proves that C_α is almost-surely finite. Use then the Hölder-continuity of X on $\mathbb{D}$ to extend it to $[0,1]$ in an unambiguous way, setting $Y_t = X_t$ for $t \in \mathbb{D}$, and

$$Y_t = \lim_{s\to t,\, s\in\mathbb{D}} X_s.$$

for $t \in [0,1]\backslash\mathbb{D}$. This defines a measurable function of ω (as a limit of measurable functions), so that $(Y_t)_{t\in[0,1]}$ defines a random process; it has by construction α-Hölder paths. ▷

2. Constructive approach in separable Banach spaces

We have seen in section 1 how one can construct in a more or less abstract way probability spaces and random processes. A different approach to the construction problem is taken in this section. Starting with probabilities on some space defined in an elementary way, we construct new probabilities as limits of such elementary probabilities; the above construction of Wiener measure as a limit of elementary probability measures corresponding to random piecewise linear continuous functions is an archetype of such a procedure. Our first task will be to explain what we mean by the limit of a sequence of probability measures. Before investigating further the general case we shall see in section 2.2 how this convergence notion works in $\mathbb{R}$. The general case is addressed in section 2.3. We shall see in section 2.3 **b)** how to characterize the compact sets of the set of probability measures. As is the case of the compact segment $[0,1]$, general compact sets have the property that any sequence of its points have a converging subsequence. This setting is thus ideal to construct some objects as limits of other objects[21]. With a view to constructing random continuous time functions, we shall see in 2.3 **c)** how the theory works in the space $\mathcal{C}\big([0,1],\mathbb{R}^n\big)$. We shall finally illustrate the whole section in 2.3 **d)** by proving Donsker's amazing invariance principle: *any nice random walk, properly rescaled, "is" a Brownian motion.*

Those of you who are not familiar with metric spaces can think about $\mathbb{R}^d$ throughout the whole section.

2.1. Weak convergence on the set of probability measures on a metric space.

Prior to the notion of limit is the notion of neigbourhood; it is the datum of the "neighbouring" relations amongst the elements of a given space. We define such a notion below on the space of probability measures of a metric space.[22]

Notations. • Given a measurable space $(A,\mathcal{A})$ denote by $\mathcal{P}(A)$ the set of probability measures on $\big(A,\mathcal{A}\big)$.

• Given a metric space (S,d), recall that the Borel σ-algebra of S is the σ-algebra generated by the open balls of S; denote it by $\mathcal{S}$.

• Write $\mathcal{C}_b(S)$ for the set of bounded real-valued continuous functions on S, and (f,μ) for $\int f(x)\,\mu(dx)$, if μ is any finite measure on $(S,\mathcal{S})$.

The following definition formalizes the fact that we want to declare two probability measures μ and ν on S close if the integrals of sufficiently many continuous bounded functions against μ and ν are close.

Definition **23.** *The $\mathcal{C}_b(S)^*$**-topology**[23] on $\mathcal{P}(S)$ is defined by the following basis of neighbourhoods of a point $\mu \in \mathcal{P}(S)$:*

$$\Big\{\big\{\nu \in \mathcal{P}(S)\,;\, \big|(f_i,\mu)-(f_i,\nu)\big| < a_i,\ 1 \leqslant i \leqslant n\big\}\,;\, n \geqslant 1, f_i \in \mathcal{C}_b(S), a_i > 0\Big\}.$$

[21]The existence of a limit object is usually a difficult question, except, precisely, when we are working in a compact set!

[22]Definitions 23 and 24 below apply on any topological space, not necessarily metric.

[23]Analysts call it the "weak* topology".

So[24] *a sequence* $(\mu_n)_{n\geqslant 0}$ *of probability measures* $\mathcal{C}_b(S)^*$*-converges to* μ *iff* $(f,\mu_n)\to(f,\mu)$ *for all* $f\in\mathcal{C}_b(S)$*. We shall adopt the notation* $\mu_n \overset{\mathcal{C}_b(S)^*}{\longrightarrow} \mu$*.*

Definition **24.** *An* S*-valued sequence of random variables* $(X_n)_{n\geqslant 0}$*, defined on some probability space* $(\Omega,\mathcal{F},\mathbb{P})$*, is said to* ***converge weakly*** *to* X *if*$_{def}$ $\mathbb{E}\big[f(X_n)\big]$ *converges to* $\mathbb{E}\big[f(X)\big]$ *as* n *goes to infinity, for any* $f\in\mathcal{C}_b(S)$*. We write* $X_n \overset{w}{\to} X$*.*

As is clear from the definition, X_n converges weakly to X iff its distribution $\mathcal{C}_b(S)^*$-converges to the distribution of X. For that reason the $\mathcal{C}_b(S)^*$-topology[25] is usually also called the ***weak topology***, and $\mathcal{C}_b(S)^*$-convergence called weak convergence.

Proposition **25** (Characterisation of $\mathcal{C}_b(S)^*$-convergence, Alexandrov)**.** *The following propositions are equivalent:*

(1) $\mu_n \overset{\mathcal{C}_b(S)^*}{\longrightarrow} \mu$,
(2) $(f,\mu_n)\to(f,\mu)$ *for every bounded uniformly continuous function* f*,*
(3) $\mu(O)\leqslant \underline{\lim}\,\mu_n(O)$ *for all open set* O *of* S*,*
(4) $\overline{\lim}\,\mu_n(F)\leqslant\mu(F)$ *for all closed set* F *of* S*,*
(5) $\mu_n(B)\to\mu(B)$ *for all Borel set* B *with* $\mu(\partial B)=0$*.*

Proof – We make a circular proof starting with the implication (1) ⇒ (2) ⇒ (3). The first implication is obvious. Given an open set O define the function $f_k(x)=1\wedge k\,d(x,O^c)$: it is bounded, k-Lipschitz (hence uniformly continuous), smaller than $\mathbf{1}_0$, and increases pointwise to $\mathbf{1}_O$; so we have $(f_k,\mu_n)\leqslant\mu_n(O)$. Letting n go to ∞ and then taking the limit $k\to\infty$ gives (3). Propositions (3) and (4) are clearly equivalent. Assume (4) and let B be any element of $\mathcal{S}$.

$$\mu\big(\overset{\circ}{B}\big)\leqslant\underline{\lim}\,\mu_n(B)\leqslant\overline{\lim}\,\mu_n(B)\leqslant\mu\big(\overline{B}\big)$$

As left and right members of theses inequalities coincide if $\mu(\partial B)=0$, proposition (5) follows. Last, supposing (5), notice that it is sufficient to prove

$$\overline{\lim}\,(f,\mu_n)\leqslant(f,\mu) \tag{2.1}$$

for any continuous function to get (1): apply it to f and $-f$ to get $(f,\mu_n)\to(f,\mu)$. As the set $E:=\big\{t\in\mathbb{R}\,;\,\mu\big(f^{-1}(\{t\})\big)\neq 0\big)\big\}$ is at most countable one can find a decreasing sequence $(f_\ell)_{\ell\geqslant 0}$ of simple functions[26] $f_\ell=\sum t_{i+1}\mathbf{1}_{f\in[t_i,t_{i+1})}$ converging μ-almost-surely to f and with no t_i in E. We then have for every ℓ the inequality $\overline{\lim_n}\,(f,\mu_n)\leqslant\lim_n(f_\ell,\mu_n)=(f_\ell,\mu)$, from which (2.1) follows by sending ℓ to infinity. ▷

2.2. **Specific tools in finite dimension.** Before investigating further the general case we investigate in this section how the above definition specializes on $\mathbb{R}$.

[24]Those of you who are not familiar with general topology can skip the preceding definition and only keep in mind the following property.
[25]That is the datum of all the above neighbourhoods.
[26]This is actually a finite sum.

a) Distribution functions. Recall the construction of a real-valued random variable with any fixed distribution described in section 1.2. Given a probability measure μ in $\mathbb{R}$ denote by $F : \mathbb{R} \to [0,1]$ its distribution function $F(t) = \mu\big((-\infty, t]\big)$ and by $G : [0,1] \to \mathbb{R}$ its right inverse

$$G(u) = \inf\{t \in \mathbb{R}\,;\, F(t) \geqslant u\}$$

with the convention that $\inf \emptyset = +\infty$; this is a càdlàg function characterized by the property

$$u \leqslant F(t) \text{ iff } G(u) \leqslant t.$$

So if U is a uniform random variable in $[0,1]$

$$\mathbb{P}\big(G(U) \leqslant t\big) = \mathbb{P}\big(F(t) \geqslant U\big) = F(t).$$

This canonical way of constructing a random variable with distribution μ leads to the following useful representation, or coupling, theorem. Given a sequence $(\mu_n)_{n \geqslant 0}$ of probability measures on $\mathbb{R}$, define the random variables $X_n = G_n(U)$, where G_n is the right inverse of the distribution function F_n of μ_n.

Theorem **26** (Representation theorem, Coupling)**.** *Suppose the sequence $(\mu_n)_{n\geqslant 0}$ weakly converges to μ, then we can construct on $[0,1]$ some random variables X_n with distribution μ_n and X with distribution μ, such that X_n converges almost-surely to X.*

Proof – Define the random variables U, $X_n = G_n(U)$ and $X = G(U)$ on the probability space $\big([0,1], \mathcal{B}([0,1]), \text{Leb}\big)$. Denote by C the countable (why?) union of intervals where F is constant and by D its image by F (made up by at most countably many points). We prove that for $u \in [0,1]\backslash D$ the sequence $G_n(u)$ converges to $G(u)$. Given such a u let us take a (small) positive ε such that $G(u) + \varepsilon$ is a continuity point of F([27]). As $F_n\big(G(u) + \varepsilon\big)$ converges to $F\big(G(u)+\varepsilon\big) > u$ (by point (4) in Alexandrov's characterization, with the Borel set $(-\infty, G(u) + \varepsilon]$) we have $F_n\big(G(u) + \varepsilon\big) \geqslant u$ for n large enough, *i.e.* $G_n(u) \leqslant G(u) + \varepsilon$. As ε can be taken arbitrarily small this shows that $\overline{\lim}\, G_n(u) \leqslant G(u)$. In the same way, choosing a small $\varepsilon > 0$ such that $G(u) - \varepsilon$ is a continuity point of F, on gets a sequence $F_n\big(G(u) - \varepsilon\big)$ converging to $F\big(G(u) - \varepsilon\big) < u$; so $F_n\big(G(u) - \varepsilon\big) \leqslant u$ for n large enough, *i.e.* $G(u) - \varepsilon \leqslant \underline{\lim}\, G_n(u)$. ▷

Note that we have only used one fact in the above proof: $F_n(x) \to F(x)$ for all $x \in \mathbb{R}$ at which F is continuous[28]. We see that under this sole hypothesis the conclusion of the theorem implies that[29] $(f, \mu_n) = \mathbb{E}_L\big[f(X_n)\big]$ converges to $\mathbb{E}_L\big[f(X)\big] = (f, \mu)$, for any bounded continuous function f.

Corollary **27.** *A sequence $(\mu_n)_{n\geqslant 0}$ of probability measures on $\mathbb{R}$ converges to μ iff $F_n(x)$ converges weakly to $F(x)$ for all $x \in \mathbb{R}$ at which F is continuous.*

Theorem **28** (Prohorov's compactness theorem in dimension 1)**.** *Let $(\mu_n)_{n\geqslant 0}$ be a sequence of probability measures such that*

$$\forall\, \varepsilon\, \exists\, M_\varepsilon > 0\, \forall\, n \geqslant 0, \quad F_n(-M_\varepsilon) \leqslant \varepsilon \text{ and } 1 - F_n(M_\varepsilon) \leqslant \varepsilon.$$

Then the sequence $(\mu_n)_{n\geqslant 0}$ has a weakly convergent subsequence.

[27] F have only countably many discontinuity points.
[28] That is $F(x^-) = F(x)$, since F is always continuous on the right – can you see why?
[29] Recall $\mathbb{E}_L$ denotes the expectation under Lebesgue measure on $[0,1]$.

The condition of the statement means that the probability measures μ_n have their mass essentially concentrated on the compact set $[-M, M]$, uniformly in n. Such a sequence is said to be *tight*.

PROOF – Use Cantor's diagonalisation procedure to extract a subsequence such that $F_{n_k}(t)$ converges for each rational t to some $F(t)$. This limit function $F : \mathbb{Q} \to [0, 1]$ being increasing has a unique extension to $\mathbb{R}$ which is continuous on the right, with left limits[30]. Check that the convergence $F_{n(k)}(s) \to F(s)$ holds if s is a continuity point of F. By the hypothesis we can associate to any $\varepsilon > 0$ a positive M_ε such that the inequalities

$$F_n(-M_\varepsilon) \leqslant \varepsilon \text{ and } 1 - F_n(M_\varepsilon) \leqslant \varepsilon$$

hold for all $n \geqslant 0$. It follows that

$$F(s) \underset{s\to-\infty}{\longrightarrow} 0 \text{ and } F(s) \underset{s\to+\infty}{\longrightarrow} 1,$$

so F is the distribution function[31] of a probability measure μ and, as a consequence of Alexandrov's characterization, $\big(\mu_{n(k)}\big)_{k\geqslant 0}$ converges weakly to μ. ▷

b) Weak convergence and characteristic functions. Corollary 28 can be used to prove the following useful result due to Paul Lévy. Recall the characteristic function of a measure μ on $\mathbb{R}$ is its Fourier transform:

$$\psi(t) = \int e^{itx}\mu(dx).$$

THEOREM **29.** *Let $(\mu_n)_{n\geqslant 0}$ be a sequence of probability measures on $\mathbb{R}$, with characteristic functions ϕ_n. If the ϕ_n converge pointwise to some function ϕ, continuous at 0, the sequence $(\mu_n)_{n\geqslant 0}$ converges weakly to some probability measure μ and ϕ is the characterisitic function of μ.*

The proof is based on the following simple estimate of the tail of a random variable Y in terms of its characteristic function ψ.

$$\mathbb{P}\Big(|Y| \geqslant \frac{1}{h}\Big) \leqslant \frac{C}{2h}\int_{-h}^{h}\big(1 - \psi(t)\big)dt \tag{2.2}$$

for some constant C and every positive h. Indeed, apply Fubini's theorem to see that[32]

$$\frac{1}{2h}\int_{-h}^{h}\big(1 - \psi(t)\big)dt = \mathbb{E}\Big[1 - \frac{\sin(hY)}{hY}\Big];$$

because $1 - \frac{\sin(hY)}{hY}$ is non-negative, and no less than $1 - \sin 1$ on the set $\big\{|Y| \geqslant \frac{1}{h}\big\}$, we have

$$\frac{1}{2h}\int_{-h}^{h}\big(1 - \psi(t)\big)dt \geqslant (1 - \sin 1)\mathbb{P}\Big(|Y| \geqslant \frac{1}{h}\Big).$$

PROOF – We prove that the sequence $(\mu_n)_{n\geqslant 0}$ is tight; then any weakly convergent subsequence (whose existence is guaranteed by corollary 28) will have ϕ as a characteristic function. This will show that $(\mu_n)_{n\geqslant 0}$ can have only one limit, so it converges.

[30] Set for $s \in \mathbb{R}\backslash\mathbb{Q}$, $F(s) = \inf\{F(t)\,;\, t \in \mathbb{Q}, t > s\}$. Like any other increasing $[0, 1]$-valued function, F has at most countably many discontinuities.

[31] Use Caratheodory's extension theorem.

[32] We use the convention $\frac{\sin 0}{0} = 1$.

But applying inequality (2.2) to X_n we obtain by dominated convergence

$$\overline{\lim}\,\mathbb{P}\Big(|X_n| \geqslant \frac{1}{h}\Big) \leqslant \frac{(1-\sin 1)^{-1}}{2h}\int_{-h}^{h}\big(1-\phi(t)\big)dt.$$

It now suffices to use the continuity of ϕ in 0 to see that the right hand side can be made arbitrarily small for h small enough. $\triangleright$

The same result holds for $\mathbb{R}^n$-valued random variables; the proof of this statement is a cosmetic change of the preceding one[33].

2.3. **Weak convergence in separable Banach spaces.** We now come back to the general case and study in more details the notion of weak convergence. This notion was introduced to provide a framework in which talking about limits of measures and constructing probability measures (and thus random processes) as limits of other measures (resp. processes). Statements about the existence of a limit are precious statements as existence statements are rarely easy to prove. There is yet one exception to this empirical rule: one can always decide whether or not a sequence of probability measures on a finite set converges or not (at least computers can do that for us!). As compact sets of a metric space are finite up to any arbitrarily small accuracy, they appear as a good framework in which tackling our convergence problem.

a) Compact sets of a metric space. Let (S, d) be a metric space. A subset K of S is said to be ***compact*** if$_{\text{def}}$ it is closed and for any $\varepsilon > 0$ it can be covered by finitely many balls of radius ε. Think of a closed interval of $\mathbb{R}$. This image might be misleading though, as although compact sets are closed and bounded, these two conditions alone are generally not sufficient to ensure compactness of a set. Indeed, any infinite dimensional normed vector space has a non-compact unit closed ball[34]. It can be proved that a metric space S is compact iff any sequence of points of S has a converging subsequence.

Suppose now that (S, d) is a compact metric space and let us look at the space $\mathcal{P}(S)$ of probability measures on $(S, \mathcal{S})$.[35] It is easily seen, using Stone-Weierstrass theorem, that $\mathcal{C}(S)\big(= \mathcal{C}_b(S) \text{ here}\big)$ has a dense sequence[36], say $(f_p)_{p\geqslant 0}$. Given any sequence $(\mu_n)_{n\geqslant 0}$ of probability measures on S, we can construct by a diagonal argument a subsequence such that each integral $(f_p, \mu_{n(k)})$ converges as k goes to infinity. This implies that all the integrals $(f, \mu_{n(k)})$, $f \in \mathcal{C}(S)$, converges as k goes to infinity (can you see why?). In other

[33]The $\mathbb{R}^n$-version of Prohorov's compactness theorem needed for the above proof to work is proved in a much more general framework below.

[34]Denote by B the unit ball and by $\overline{B}$ its closure. Suppose $\overline{B}$ is compact and cover it by finitely many balls of radius $\frac{1}{2}$. Denoting by F the finite-dimensional vector space spanned by their centers, we have $\overline{B} \subset F + \frac{B}{2}$, implying $\frac{\overline{B}}{2} \subset F + \frac{B}{4}$, and so $\overline{B} \subset F + \frac{B}{4}$. An induction bootstraps this inclusion into $\overline{B} \subset F + \frac{B}{2^n}$, for all $n \geqslant 1$, out of which it follows that $\overline{B} \subset F$, as F is closed; this proves that the ambiant vector space needs to be finite dimensional. This theorem is due to F. Riesz.

[35]Recall the notations introduced at the beginning of section 2.1.

[36]For the supremum norm. Consult theorem (81.3) of [RW00] for instance.

terms[37], any sequence of probability measures on a compact metric space has a weakly converging subsequence. Also, introducing the metric $\delta(\mu,\nu) = \sum_{p\geqslant 0}\big|(f_p,\mu)-(f_p,\nu)\big| \wedge 2^{-p}$ on $\mathcal{P}(S)$, one sees that its balls define the same notion of neighbourhoods as the weak topology; so the space $\mathcal{P}(S)$, with its weak topology, is a compact metric space. This conclusion holds in particular when we take for (S,d) the compact space $\big([0,1]^{\mathbb{N}},d\big)$, where $d(x,x') = \sum_{n\geqslant 0} 2^{-n}|x_n - x'_n|$.

b) What is special about separable Banach spaces? First of all, this is a general enough framework to encompass most of everyday' spaces we want to work with: the space of (real-valued) sequences, continuous and càdlàg paths are separable Banach spaces.

At the same time, a lot of things are known on separable Banach spaces! It is proved in the Comment section *Separable Banach spaces* that any separable Banach space (S,d) is homeomorphic to a measurable subset of the compact metric space $[0,1]^{\mathbb{N}}$. This implies[38] that the space $\big(\mathcal{P}(S),\mathcal{C}_b(S)^*\big)$ of probability measures on S is homeomorphic to a subset of the nice compact metrizable space $\Big(\mathcal{P}\big([0,1]^{\mathbb{N}}\big),\mathcal{C}_b^*\big([0,1]^{\mathbb{N}}\big)\Big)$. So, any sequence of probability measures on S, seen as probability measures on $[0,1]^{\mathbb{N}}$, has a converging subsequence in $\mathcal{P}\big([0,1]^{\mathbb{N}}\big)$, whose limit may give some positive mass to the set $[0,1]^{\mathbb{N}}\backslash S$,[39] giving rise to a limit measure in S of mass less than 1. One introduces the following notion to prevent this phenomenon and obtain limit probability measures supported on the original space.

DEFINITION **30.** *A family $\mathfrak{A}$ of measures on $(S,\mathcal{S})$ is said to be **tight** if one can associate to any $\varepsilon>0$ a compact set K_ε of S such that*

$$\forall\,\mu\in\mathfrak{A},\quad \mu(K_\varepsilon^c)\leqslant\varepsilon.$$

c) Compactness in $\big(\mathcal{P}(S),\mathcal{C}_b(S)^*\big)$. The following theorem due to Prohorov characterizes a large class of compact sets of $\big(\mathcal{P}(S),\mathcal{C}_b(S)^*\big)$ in terms of tightness. It is the general counterpart of theorem 28.

THEOREM **31** (Compactness. Prohorov)**.** *Let (S,d) be a separable metric space and $\mathfrak{A}\subset\mathcal{P}(S)$.*

- *If the family $\mathfrak{A}$ is tight then it is relatively compact in $\big(\mathcal{P}(S),\mathcal{C}_b(S)^*\big)$.*
- *Suppose in addition that (S,d) is complete. Then the two properties are equivalent.*

PROOF – • Suppose the family $\mathfrak{A}$ is tight and let $\big(K_{\frac{1}{p}}\big)_{p\geqslant 1}$ be an increasing sequence of compact subsets of S for which $\mu\big(K^c_{\frac{1}{p}}\big)\leqslant\frac{1}{p}$, for all $\mu\in\mathfrak{A}$. Denote by φ the homeomorphism between (S,d) and a subset[40] of $[0,1]^{\mathbb{N}}$ constructed in theorem 39 of the Comments section. As each *compact* set $\varphi\big(K_{\frac{1}{p}}\big)$ is measurable, $\varphi\big(\bigcup_{p\geqslant 1}K_{\frac{1}{p}}\big)$ is also measurable, as a union of measurable

[37]We are skipping here a little argument. The subsequence can be chosen such that the integrals of any linear combination of the f_p's converge. This implies that the map $f_p\to\lim_k\big(f_p,\mu_{n(k)}\big)$ is a positive linear map, with unit norm. It can be extended to $\mathcal{C}(S)$ by a straightforward approximation argument, so the map $L: f\to\lim_k\big(f,\mu_{n(k)}\big)$ is a positive linear form on $\mathcal{C}(S)$ with unit norm. Riesz representation theorem ensures us that there exists a probability measure μ on S such that $L(f)=(f,\mu)$ for all $f\in\mathcal{C}(S)$. Riesz representation theorem is proved in Complement 4.
[38]See theorem (83.7) in [RW00].
[39]I am writing S here for its homeomorphic image in $[0,1]^{\mathbb{N}}$.
[40]This subset of $[0,1]^{\mathbb{N}}$ has no a priori reason to be measurable.

sets. Now, since all the measures $\mu \in \mathfrak{A}$ have support in $\bigcup_{p\geqslant 1} K_{\frac{1}{p}}$ it is harmless to replace S by $\bigcup_{p\geqslant 1} K_{\frac{1}{p}}$; we still denote it by S. The map φ is then a homeomorphism between (S,d) and a *measurable* subset of $[0,1]^{\mathbb{N}}$; we use this function to transfer any statement about (S,d) to a statement about a subset of $[0,1]^{\mathbb{N}}$.

We shall associate to any sequence $(\mu_n)_{n\geqslant 0}$ of $\mathcal{P}(S)$ the sequence $(\nu_n)_{n\geqslant 0}$ of its images by φ in $\mathcal{P}\big([0,1]^{\mathbb{N}}\big)$. Given $\varepsilon > 0$, each $\varphi(K_\varepsilon)$ is a compact subset of $[0,1]^{\mathbb{N}}$ with ν_n-measure no less than $1-\varepsilon$ for any n. But as $\big(\mathcal{P}\big([0,1]^{\mathbb{N}}\big), \mathcal{C}_b\big([0,1]^{\mathbb{N}}\big)^*\big)$ is compact there is a sub-sequence $\big\{\nu_{n(k)}\big\}_{k\geqslant 0}$ that $\mathcal{C}_b\big([0,1]^{\mathbb{N}}\big)^*$-converges to some Borel probability measure ν on $[0,1]^{\mathbb{N}}$. From Alexandrov's characterization the limit probability ν satisfies $\nu\big(\varphi(K_\varepsilon)\big) \geqslant 1-\varepsilon$, for all $\varepsilon > 0$, hence ν is concentrated on $\varphi(S)$. Defining μ as the image measure of ν by φ^{-1}, the function $f\circ\varphi^{-1}$ is continuous and bounded for any $f \in \mathcal{C}_b(S)$, so we have

$$(f, \mu_{n(k)}) = \big(f\circ\varphi^{-1}, \nu_{n(k)}\big) \to \big(f\circ\varphi^{-1}, \nu\big) = (f, \mu),$$

that is, $\mu_{n(k)} \overset{\mathcal{C}_b(S)^*}{\longrightarrow} \mu$.

• Suppose now in addition that (S,d) is complete and let $\mathfrak{A} = \{\mu_\ell\,;\, \ell \in \Lambda\}$ be a compact subset of $\big(\mathcal{P}(S), \mathcal{C}_b(S)^*\big)$. Let $(x_n)_{n\geqslant 0}$ be a dense sequence of (S,d) and define $O_n(r) = \bigcup_{k=1..n} B(x_k, r)$. Let us first prove that

($\bigstar$) *for any $\varepsilon > 0, r > 0$ there exists an integer $N(\varepsilon, r)$ such that $\mu\big(O_{N(\varepsilon,r)}\big) \geqslant 1-\varepsilon$, for any $\mu \in \mathfrak{A}$.*

Would assertion ($\bigstar$) be wrong, there would exist ε_0, r_0 and for each n and index $\ell_n \in \Lambda$ such that $\mu_{\ell_n}(O_n(r_0)) \leqslant 1-\varepsilon_0$. Any limit μ of a converging subsequence $\big(\mu_{\ell_{n(k)}}\big)_{k\geqslant 0}$ (we are in a compact!) would then verify for any $p \geqslant 0$

$$\mu\big(O_{n(p)}(r_0)\big) \leqslant \underline{\lim}\, \mu_{n(k)}\big(O_{n(k)}(r_0)\big) \leqslant 1-\varepsilon_0$$

since $O_{n(p)} \subset O_{n(k)}$ for $k \geqslant p$, and by Alexandrov's proposition 25; this would forbid the convergence $\mu\big(O_{n(p)}(r_0)\big) \underset{p+\infty}{\to} 1$, a contradiction.

Fix now $\eta > 0$ and set

$$K := \bigcap_{p\geqslant 1} \left(\bigcup_{k=1}^{N\left(2^{-p}\eta, \frac{1}{p}\right)} \overline{B}\Big(x_k, \frac{1}{p}\Big) \right).$$

K is a compact set which satisfies for any $\mu \in \mathfrak{A}$ the inequality

$$\mu(K) \geqslant 1 - \sum_{p\geqslant 1} \mu\left(S \setminus \bigcup_{k=1}^{N\left(2^{-p}\eta, \frac{1}{p}\right)} \overline{B}\Big(x_k, \frac{1}{p}\Big) \right) \geqslant 1 - \sum_{p\geqslant 1} 2^{-p}\eta = 1-\eta.$$

This proves the tightness of the family $\mathfrak{A}$ of measures. ▷

d) Continuous random processes. We specialize in this paragraph the above general theory to the case of measures on the space of continuous function from some interval I of $\mathbb{R}_+$ to some $\mathbb{R}^d$. We shall thus be working here on the separable Banach space $(S,d) = \big(\mathcal{C}\big(I, \mathbb{R}^d\big), \|\cdot\|_\infty\big)$.

We have noticed in section 1.3 **b)** that the space $\mathcal{C}(I, \mathbb{R}^d)$ can be seen from two natural point of views: as a subset of the product $(\mathbb{R}^d)^I$ or as the metric space $\big(\mathcal{C}\big(I, \mathbb{R}^d\big), \|\cdot\|_\infty\big)$. Each picture has its own notion of σ-algebra. The following proposition states that the two σ-algebras coincide, so there is no problem on which point of view is adopted. Recall we denote by $X_t : \omega \in \mathcal{C}\big(I, \mathbb{R}^d\big) \mapsto \omega_t, t \in I$, the coordinate process.

PROPOSITION **32.** *The σ-algebra on $\mathcal{C}\big(I, \mathbb{R}^d\big)$ generated by the coordinate process coincides with the Borel σ-algebra of $\big(\mathcal{C}\big(I, \mathbb{R}^d\big), \|\cdot\|_\infty\big)$.*

PROOF – The trace on $\mathcal{C}(I, \mathbb{R}^d)$ of the product σ-algebra is generated by the collection $\mathcal{A}$ of the elementary events $\{X_{t_1} \in A_1, ..., X_{t_n} \in A_n\}$, where $n \geqslant 1$ and the A_i's are open balls of $\mathbb{R}^d$. The Borel σ-algebra of $\big(\mathcal{C}\big(I, \mathbb{R}^d\big), \|\cdot\|_\infty\big)$ is generated by the collection $\mathcal{B}$ of the open balls $\big\{\omega \in \mathcal{C}\big(I, \mathbb{R}^d\big)\,;\, \|\omega - \omega_0\|_\infty < \epsilon\big\}$, for $\epsilon > 0$ and $\omega_0 \in \mathcal{C}(I, \mathbb{R}^d)$. To prove that the two σ-algebras coincide it suffices to prove that any element of $\mathcal{A}$ is in $\sigma(\mathcal{B})$ and any element of $\mathcal{B}$ is in $\sigma(\mathcal{A})$.
Let $C := \big\{\omega \in \mathcal{C}\big(I, \mathbb{R}^d\big)\,;\, X_{t_1}(\omega) \in A_1, ..., X_{t_n}(\omega) \in A_n\big\}$ be an elementary event, and denote by $(\omega_p)_{p\geqslant 1}$ a dense sequence of $\big(\mathcal{C}\big(I, \mathbb{R}^d\big), \|\cdot\|_\infty\big)$; denote by $(\omega_{p(k)})_{k\geqslant 1}$ the subset of it made up of those ω_p's which belong to C. Then, for each $\omega \in C$, you can find some k_j for which $\|\omega - \omega_{p(k_j)}\|_\infty \leqslant \frac{1}{j}$; this proves the first point.
To prove the second point, denote by $(t_n)_{n\geqslant 1}$ a dense sequence of I, and notice that $\{\omega\,;\, \|\omega - \omega_0\|_\infty < \epsilon\} = \bigcap_{n\geqslant 1}\big\{\omega\,;\, \omega_{t_1} \in B(\omega_0(t_1), \epsilon), ..., \omega_{t_n} \in B(\omega_0(t_n), \epsilon)\big\}$. ▷

Considering $\mathcal{C}(I, \mathbb{R}^d)$ as a subset of $(\mathbb{R}^d)^I$ leads to the following notion of convergence.

DEFINITION **33** (Convergence of finite-dimensional distributions)**.** • *Let $\mu_n, n \geqslant 0$ and μ be Borel probability measures on $\mathcal{C}(I, \mathbb{R}^d)$. We say that the* ***finite-dimensional distributions of*** *μ_n* ***converge to those of*** *μ if$_{def}$ for every finite collection $\{t_1, \ldots, t_p\}$ of times, and any bounded continuous function $f : \big(\mathbb{R}^d\big)^p \to \mathbb{R}$, we have*

$$\int f\big(X_{t_1}(\omega), \ldots, X_{t_p}(\omega)\big)\mu_n(d\omega) \to \int f\big(X_{t_1}(\omega), \ldots, X_{t_p}(\omega)\big)\mu(d\omega).$$

We write $\mu_n \overset{fd}{\to} \mu$.

• *Let $(Y^{(n)})_{n\geqslant 0}$ and Y be $\mathcal{C}(I, \mathbb{R}^d)$-valued random variable defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We shall write $Y^{(n)} \overset{fd}{\to} Y$ if$_{\text{def}}$*

$$\mathbb{E}\big[f\big(Y^{(n)}_{t_1}, \ldots, Y^{(n)}_{t_p}\big)\big] \to \mathbb{E}\big[f(Y_{t_1}, \ldots, Y_{t_p})\big]$$

for any bounded continuous function $f : \big(\mathbb{R}^d\big)^p \to \mathbb{R}$, any $p \geqslant 1$ and any finite collection $\{t_1, \ldots, t_p\}$ of times.

PROPOSITION **34.** *Some probability measures μ_n on $\mathcal{C}(I, \mathbb{R}^d)$ converge weakly to some probability measure μ iff the following conditions hold:*

- *the finite dimensional distributions of μ_n converge to those of μ,*
- *the family $(\mu_n)_{n\geqslant 0}$ is tight.*

PROOF – $\Rightarrow$ Since the map $\omega \in \mathcal{C}(I, \mathbb{R}^d) \mapsto F(\omega_{t_1}, \ldots, \omega_{t_n})$ is continuous for any $n \geqslant 1$, $t_1, \ldots, t_n$ and continuous function F, the weak convergence of μ_n to some μ implies the finite dimensional convergence of μ_n to μ. Also, any convergent sequence is tight (prove it).
$\Leftarrow$ Suppose the sequence $(\mu_n)_{n\geqslant 1}$ is tight; by the first part of Prohorov's compactness theorem 31, it is relatively compact. Any limit ν of a converging subsequence having the same finite dimensional distribution as μ, we must have $\nu = \mu$. This shows that μ is the only cluster point of the sequence $(\mu_n)_{n\geqslant 1}$, so $(\mu_n)_{n\geqslant 0}$ converges weakly to μ. ▷

Do exercise 14 to see that one can have finite dimensional convergence without weak convergence. Given a compact interval $[a,b]$ of the real line, Ascoli-Arzela's compactness criterion gives a characterization of compact sets of $\big(\mathcal{C}\big([a,b],\mathbb{R}^d\big), \|\cdot\|_\infty\big)$ in terms of ***modulus of continuity***

$$M_\omega(h) = \sup\big\{|\omega_t - \omega_s|\,;\, t,s\in[a,b],\ |t-s|\leqslant h\big\},\quad h>0.$$

Theorem **35** (Ascoli-Arzela's theorem). *A subset A of $\mathcal{C}\big([a,b],\mathbb{R}^d\big)$ is relatively compact iff the following two conditions hold:*

$$\sup\big\{|\omega_0|\,;\,\omega\in A\big\}<\infty,$$
$$\lim_{h\searrow 0}\sup_{\omega\in A} M_\omega(h) = 0.$$

Together with Prohorov's theorem it provides an easy to use characterisation of compact subsets of the set of probability measures on $\mathcal{C}\big([a,b],\mathbb{R}^d\big)$.

Corollary **36** (Characterization of weak convergence). *Let $X, X_1, X_2,\dots$ be $\mathbb{R}^d$-valued continuous random processes. Then $X_n \overset{w}{\to} X$ iff $X_n \overset{fd}{\to} X$ and*

$$\lim_{h\searrow 0}\ \overline{\lim_{n\to+\infty}}\ \mathbb{E}\big[M_{X_n}(h)\wedge 1\big] = 0. \tag{2.3}$$

Proof – It suffices from corollary 34 to prove that condition (2.3) is equivalent to tightness. That the former implies the latter comes from Ascoli-Arzela's theorem and dominated convergence. Conversely, assume (2.3) and fix $h>0$. Since each X_n is continuous, $M_{X_n}(h)\to 0$ almost-surely as $h\searrow 0$ for each n; as a consequence of condition (2.3) it is thus possible to find a sequence $(h_k)_{k\geqslant 0}$ such that

$$\sup_n \mathbb{P}\big(M_{X_n}(h_k) > 2^{-k}\big) \leqslant 2^{-k-1}h$$

for all $k\geqslant 0$. Also, as $X_n \overset{fd}{\to} X$, there exists compact subsets $K_1, K_2,\dots$ of $\mathbb{R}^d$ such that

$$\sup_n \mathbb{P}\big(X_n(a)\notin K_k\big)\leqslant 2^{-k-1}h$$

for all $k\geqslant 0$. So the set

$$B := \big\{x\in\mathcal{C}\big([a,b],\mathbb{R}^d\big)\,;\, x(a)\in K_k,\ M_x(h_k)\leqslant 2^{-k},\ \text{for all } k\geqslant 0\big\}$$

satisfies $\sup_n \mathbb{P}(X_n\notin B)\leqslant 2h$ and has compact closure from Ascoli-Arzela's theorem; this proves the tightness of the laws of X_n. ▷

2.4. **Application: Universality of Brownian motion.** The central limit theorem gives a universal status to the Gaussian law among the class of (Borel) probability measures on the line, with finite first two moments. Brownian motion enjoys a similar universal property.

Theorem **37** (Donsker's invariance theorem). *Let $(X_n)_{n\geqslant 0}$ be a sequence of iid centered real-valued random variables with unit variance; set for $n\geqslant 1$ and $t\in[0,1]$*

$$B_t^{(n)} = \frac{1}{\sqrt{n}}\Big(\sum_{1\leqslant k\leqslant nt} X_k + \big(nt-[nt]\big)X_{[nt]+1}\Big).$$

This is a space and time rescaled version of a linearly interpolated random walk; note the scaling $n^{-1/2}$ in space and n in time. Denote by $\mathbb{P}^{(n)}$ the law of this continuous random path. Then the sequence $\big(\mathbb{P}^{(n)}\big)_{n\geqslant 0}$ converges weakly to Wiener measure.

PROOF – The strategy is simple and follows the pattern described in corollary 36: *i)* establish the convergence of finite dimensional distributions and *ii)* prove the tightness of the sequence $\big(\mathbb{P}^{(n)}\big)_{n\geqslant 0}$ using the equi-continuity criterion (2.3).

i) We need to prove that for any $p\geqslant 1$, any choice of times $t_i\in\mathbb{R}_+$ and constants a_i, the random variables $\sum_{i=1..p} a_i B^{(n)}_{t_i}$ converge in law to $\sum_{i=1..p} a_i B_{t_i}$, where B is a Brownian motion. Setting $\Delta B^{(n)}_j = B^{(n)}_{t_j} - B^{(n)}_{t_{j-1}}$, with $t_0=0$, write

$$\sum_{i=1..p} a_i B^{(n)}_{t_i} = \sum_{j=1..p}\Big(\sum_{i=1..p} a_i\Big)\Delta B^{(n)}_j;$$

as each term $\Delta B^{(n)}_j$ converges in law to $B_{t_j}-B_{t_{j-1}}$ by the central limit theorem, the result follows from the independence of the random variables $B^{(n)}_{t_j}-B^{(n)}_{t_{j-1}}$ and $B_{t_j}-B_{t_{j-1}}$.

ii) We shall use the following simple estimate to verify tightness.

LEMMA **38** (Ottaviani). *For $n\geqslant 1$, set $S_n = X_1+\cdots+X_n$ and $S^*_n = \max\big\{|S_k|\,;\,1\leqslant k\leqslant n\big\}$. Then for any $r>1$ and $n\geqslant 1$*

$$(1-r^{-2})\,\mathbb{P}\big(S^*_n\geqslant 2r\sqrt{n}\big)\leqslant \mathbb{P}\left(\frac{|S_n|}{\sqrt{n}}\geqslant r\right).$$

PROOF – Define the random time T as $\inf\big\{n\geqslant 1 \;\; \frac{|S_n|}{\sqrt{n}}\geqslant 2r\big\}$. We shall justify later that one can apply the strong Markov property[41] to the random walk $(S_n)_{n\geqslant 0}$ at time T; it is used in the third inequality below.

$$\begin{aligned}\mathbb{P}\big(|S_n|\geqslant r\sqrt{n}\big) &\geqslant \mathbb{P}\big(S^*_n\geqslant 2r\sqrt{n},\ |S_n|\geqslant r\sqrt{n}\big)\geqslant \mathbb{P}\big(T\leqslant n,\ |S_n-S_T|\leqslant r\sqrt{n}\big)\\ &\geqslant \mathbb{P}\big(T\leqslant n\big)\min_{1\leqslant k\leqslant n}\mathbb{P}\big(|S_k|\leqslant r\sqrt{n}\big)\end{aligned}$$

The inequality of the lemma follows from Chebychev's inequality

$$\min_{1\leqslant k\leqslant n}\mathbb{P}\big(|S_k|\geqslant r\sqrt{n}\big)\geqslant \min_{1\leqslant k\leqslant n}\Big(1-\frac{k}{nr^2}\Big)\geqslant 1-r^{-2}.$$

⊙

The following rough estimate comes out as a consequence of Ottaviani's lemma[42].

$$\lim_{r+\infty}\ \overline{\lim_{n+\infty}}\ r^2\,\mathbb{P}\Big(\frac{S^*_n}{\sqrt{n}}\geqslant 2r\Big)\leqslant \lim_{r+\infty} r^2\,\mathbb{P}\big(\mathcal{N}(0,1)\geqslant r\big)=0 \tag{2.4}$$

As we have for any $h>0, t\in[0,1-h]$, and $\ell>0$ ([43])

$$\begin{aligned}\mathbb{P}\Big(\sup_{0\leqslant r\leqslant h}\big|B^{(n)}_{t+r}-B^{(n)}_t\big|\geqslant \ell\Big) &= \mathbb{P}\left(\sup_{0\leqslant r\leqslant h}\frac{\Big|\sum_{k=[nt]+1}^{[n(t+r)]}X_k+\big(n(t+r)-[n(t+r)]\big)X_{[n(t+r)]}\Big|}{\sqrt{n}}\geqslant \ell\right)\\ &\leqslant \mathbb{P}\Big(\frac{S^*_{[nh]+1}}{\sqrt{nh}}\geqslant\frac{\ell}{\sqrt{h}}\Big),\end{aligned}$$

[41]Independence of what happen after and before time T, conditionnally on what happens at time T. The proof given in section 10.3 for Brownian motion works equally well for a random walk.
[42]$\mathcal{N}(0,1)$ stands here for a centered Gaussian random variable with unit variance.
[43]The S^* in the inequality below is associated with the $X_{[nh]+i}, i\geqslant 1$; it has the same law as the S^* associated with the $X_i,\ i\geqslant 1$.

identity (2.4) implies that $\mathbb{P}\Big(\sup_{0\leqslant r\leqslant h}\big|B^{(n)}_{t+r}-B^{(n)}_t\big|\geqslant \ell\Big)=o(h)$, for each $\ell>0$, uniformly in $t\in[0,1]$ and $n\geqslant 0$. Cutting the interval $[0,1]$ into sub-intervals $\big[kh,(k+1)h\big]$ and noting that $M_{B^{(n)}}(h)\leqslant 2\max_k\Big\{\sup_{0\leqslant r\leqslant h}\big|B^{(n)}_{kh+r}-B^{(n)}_{kh}\big|\Big\}$, it follows that we have uniformly in $n\geqslant 0$

$$\begin{aligned}\mathbb{E}\big[M_{B^{(n)}}(h)\wedge 1\big]&=\int_0^\infty \mathbb{P}\big(M_{B^{(n)}}(h)\wedge 1\geqslant \ell\big)\,d\ell\leqslant \int_0^1 \mathbb{P}\big(M_{B^{(n)}}(h)\geqslant \ell\big)\,d\ell\\&\leqslant \int_0^1 \mathbb{P}\Big(2\max_k\Big\{\sup_{0\leqslant r\leqslant h}\big|B^{(n)}_{kh+r}-B^{(n)}_{kh}\big|\Big\}\geqslant \ell\Big)\,d\ell\leqslant \int_0^1 h^{-1}\,o(h)\,d\ell=o_{h,0^+}(1);\end{aligned}$$

we have used dominated convergence in the last equality, where $h^{-1}\,o(h)$ is a function of ℓ which is $o(1)$ as h decreases to 0. The above inequality proves that the equi-continuity condition (2.3) holds. ▷

3. Comments and exercises

3.1. **References and comments.** *Introduction.* Don't hesitate to read Kolmogorov's (small) treatise *Mathematical foundations of probability theory*, as it is amazing of modernity and clarity. Chapter 2 of Shiryaev's book [Shi96] is nice reading, as well as the introduction chapter of Gikhman and Skorokhod's book [GS04].

Section 1. • Kallenberg's book [Kal02], (Chap. 2, 6) contains all the material exposed in this section, with much more details. Chapters 2 and 3 of Doob's book [Doo94] are well worth being read. Chapter 3 of Rogers and Williams' book [RW00] is also an excellent source.

• Read Chapter 1 of [RW00] for an exciting and fascinating description of Brownian motion.

Section 2.1 • I can't see any better reference than the first chapter of Ikeda & Watanabe's book [IW89]. Doob's book [Doo94], Chap. 8, is also a valuable source for this section (and all measure theory). Dudley's book [Dud02], chap. 11, is also quite nice.

Section 2.2 • You will find the classical proof of Donsker's theorem using Skorokhod embedding in Chapter 1, section 8, of [RW00].

The following comments on measure theory might help you understand some subtle and potentially unnoticed points[44].

1. σ-additivity of a probability is not obvious. Set $\Omega=\mathbb{Q}\cap[0,1]$ and define on Ω the algebra $\mathcal{A}$ as the collection of disjoint (traces on $\mathbb{Q}$ of) intervals with rational ends (open or not at both ends). I leave you to check that we define an additive set function setting $\mathbb{P}\big(\{a,b\}\big)=b-a$ and $\mathbb{P}\big(\bigcup_{i=1}^n\{a_i,b_i\}\big)=\sum_{i=1}^n(b_i-a_i)$, for $\bigcup_{i=1}^n\{a_i,b_i\}\in\mathcal{A}$. As any singleton $\{r\}\in\mathcal{A}$ has null $\mathbb{P}$-measure and Ω is countable $\mathbb{P}$ cannot be σ-additive.

2. The coincidence of two probability measures on a given class does not always imply their coincidence on the σ-algebra generated by this class. Let Ω be any set and $\mathcal{C}$ be a collection of subsets of Ω. It is well-known that if $\mathcal{C}$ is stable by intersection then any two probabilities defined on $\big(\Omega,\sigma(\mathcal{C})\big)$ coinciding on $\mathcal{C}$ are actually

[44] Most of these remarks are borrowed from [Sto87].

equal (on $\sigma(\mathcal{C})$!). This is no longer the case if $\mathcal{C}$ is non stable by intersection as the following counter-example shows.

On a four point set $\Omega = \{a, b, c, d\}$ define

$$\mathbb{P}(a) = \mathbb{P}(d) = \mathbb{Q}(b) = \mathbb{Q}(c) = \frac{1}{6},$$
$$\mathbb{P}(b) = \mathbb{P}(c) = \mathbb{Q}(a) = \mathbb{Q}(d) = \frac{1}{3}.$$

Set $\mathcal{C} = \big\{\{a,b\},\{c,d\},\{a,c\},\{b,d\}\big\}$ and check that $\sigma(\mathcal{C})$ is the σ-algebra of all parts of Ω. Clearly, $\mathbb{P}$ and $\mathbb{Q}$ coincide on $\mathcal{C}$, yet they do not take the same values on the singletons $\{a\},\{b\},\{c\},\{d\}$.

3. Is Daniell's theorem obvious? Let us restate it with a slightly different point of view and in a special case sufficient for our needs. Identify each $\mathbb{R}^n$ as a subset of $\mathbb{R}^{\mathbb{N}}$ sending $x \in \mathbb{R}^n$ to $(x, 0, \cdots) \in \mathbb{R}^{\mathbb{N}}$; this identifies the Borel σ-algebra of $\mathbb{R}^n$ to a σ-algebra $\mathcal{F}_n$ of $\mathbb{R}^{\mathbb{N}}$, increasing with n. Let us then consider a projective sequence $(\mu_n)_{n\geqslant 1}$ of probability measures on $\mathbb{R}^n$ as a set function $\mathbb{P}$ on $\bigcup_{n\geqslant 1}\mathcal{F}_n$ equal on each $\mathcal{F}_n$ to μ_n. Daniell's theorem states that $\mathbb{P}$ can be extended to $\sigma\big(\bigcup_{n\geqslant 1}\mathcal{F}_n\big)$.

Given a space Ω, an increasing sequence of σ-algebras $\mathcal{F}_n$ in Ω and a set function $\mathbb{P}$ on $\bigcup_{n\geqslant 1}\mathcal{F}_n$ such that $\mathbb{P}$ is a consistently defined probability measure on each $(\Omega, \mathcal{F}_n)$, the set function $\mathbb{P}$ need not extend to a probability on $\sigma\big(\bigcup_{n\geqslant 1}\mathcal{F}_n\big)$.

Consider the (non-complete) space $\Omega = (0,1]$ and set $h_n(\omega) = \mathbf{1}_{(0,\frac{1}{n})}(\omega)$ for each $n \geqslant 1$ and $\omega \in \Omega$. Write $\mathcal{C}_n = \Big\{\emptyset, (0,\frac{1}{n}), \big[\frac{1}{n}, 1\big], (0,1]\Big\}$ for the σ-algebra generated by h_n and define $\mathcal{F}_n = \sigma(h_1, \cdots, h_n) = \Big\{\emptyset; (0,\frac{1}{n}), \big[\frac{1}{k}, \frac{1}{k-1}\big], k = n..2, \text{ and their unions}; (0,1]\Big\}$. Set $\mathbb{P}\big((0,1]\big) = 1$, and for $A \in \mathcal{F}_n$, with $A \neq (0,1]$ and $\mathbf{1}_A = a_n \mathbf{1}_{(0,\frac{1}{n})} + \sum_{k=n}^{2} b_k \mathbf{1}_{[\frac{1}{k},\frac{1}{k-1}]}$, with $a_n, b_k \in \{0,1\}$, set

$$\mathbb{P}(A) = a_n;$$

this probability has support in $(0, \frac{1}{n})$. Check that the $(\mathbb{P}_n)_{n\geqslant 1}$ are a consistent family of probabilities: $\mathbb{P}_{n+1}(A) = \mathbb{P}_n(A)$ for $A \in \mathcal{F}_n$. Would there exists a probability on $\sigma\Big(\bigcup_{n\geqslant 1}\mathcal{F}_n\Big)$ with restriction $\mathbb{P}_n$ to each $\mathcal{F}_n$, it should give unit mass to any interval $(0, \frac{1}{n})$ and satisfy at the same time the continuity property[45] $\lim_n \mathbb{P}\big((0,\frac{1}{n})\big) = 0$, a contradiction.

4. Measurable events. Let $(\Omega, \mathcal{F})$ be a *measurable space* whose σ-algebra is generated by some algebra $\mathcal{A}$. The definition of $\mathcal{F}$ as the smallest σ-algebra containing $\mathcal{A}$ is non-constructive, and it is quite tempting to believe that one can construct any element of $\mathcal{F}$ by repeated finite and countable set-theoretic operations starting from $\mathcal{A}$. Precisely, set $\mathcal{A}_0 = \mathcal{A}$ and define inductively $\mathcal{A}_{n+1}$ as the class of sets of Ω that consists of the sets of $\mathcal{A}_n$, their complements, and the finite and countable union of those. Surprisingly, this procedure does not exhaust all the elements of $\mathcal{F}$, and $\bigcup_{n\geqslant 1}\mathcal{A}_n$ is generally strictly included in $\mathcal{F}$! Consult chapter 2 of Dudley's book [Dud02] for a proof in $[0,1]$. What is true, yet, is that if we are working in a *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ then any measurable set is equal to a set of $\bigcup_{n\geqslant 1}\mathcal{A}_n$ *up to a set of null $\mathbb{P}$-measure*; this is a consequence of Caratheodory's extension theorem.

[45] Which is equivalent to σ-additivity.

To be written: comments on weak convergence in spaces of càdlàg paths

3.2. **Exercises.** **1.** Give a formal construction of a process whose dynamics corresponds to the following heuristic description. This is a variant of the symmetric random walk on $\mathbb{Z}^3$ which can never come back to any position where it has already been. Except from that requirement, it chooses each time its future location uniformly amongst the set of available nearest neighbours. If it has visited all its neighbours at some point, it stops and stays forever where it is.

2. Give a formal construction of a process whose dynamics corresponds to the following heuristic description. This time we are looking at a variant of the simple random walk in $\mathbb{Z}^3$ where the sites already visited gain attractiveness. If the process is at time n in x, it chooses its next location amongst the nearest neighbours $\{x_i\}_{i=1..6}$ of x, it jumps on x_i at time $n+1$ with probability proportional to $N_n(i)+1$, where $N_n(i)$ is the number of times that the process has visited site x_i by time n.

3. Let $\lambda > 0$. Can you construct on some probability space a sequence $(X_i)_{i\geqslant 1}$ of $\mathbb{R}^d$-valued random variables such that, if one writes $N(A)$ for $\sharp\{i\,;\, X_i \in A\}$ for each measurable set A of $\mathbb{R}^d$, then

- each random variable $N(A)$ is a Poisson random variable with parameter λ,
- for any n-uple of distinct sets A_i the random variables $N(A_i)$ are independent?

4. ***Gaussian processes.*** Let T be any index set. A real-valued random process $(X_t)_{t\in T}$ is said to be Gaussian if$_{\text{def}}$ for any $n \geqslant 1, t_1, \dots, t_n \in T$, $c_1, \dots, c_n \in \mathbb{R}$, the random variables $c_1X_{t_1} + \cdots + c_nX_{t_n}$ are Gaussian. It is said to be centered if$_{\text{def}}$ any X_t has null mean.

a) Prove that, if it exists, the distribution of a Gaussian process $(X_t)_{t\in T}$ is determined by the mean and covariance functions.

b) Let $\big(H, (\cdot,\cdot)\big)$ be a Hilbert space. A centered Gaussian process $(X_h)_{h\in H}$ with covariance $\mathbb{E}[X_hX_{h'}] = (h, h')$ (for all $h, h' \in H$) is called an isonormal Gaussian process. Suppose H is separable, and let $(e_n)_{n\geqslant 0}$ be a basis of H. Let $(G_n)_{n\geqslant 0}$ be a sequence of iid $\mathcal{N}(0,1)$. Prove that we define an isonormal Gaussian process associating to any $h = \sum_{n\geqslant 0} h^n e_n \in H$ the random variable $X_h = \sum_{n\geqslant 0} h^n G_n$.

c) **(i)** Taking for Hilbert space the space $\mathbb{L}^2(\mathbb{R}_+)$ and constructing $(X_h)_{h\in\mathbb{L}^2}$ as above, prove that the process $B_t = X_{\mathbf{1}_{[0,t]}}, t \geqslant 0$, has independent stationary Gaussian increments.

(ii) Prove that B has a modification which is continuous; this modification is thus a Brownian motion.

(iii) As a consequence, characterize Brownian motion as the unique centered Gaussian process with covariance $\mathbb{E}[X_sX_t] = \min(s,t)$.

(iv) *Scaling.* Given a Brownian motion B, prove that the process $X_t = tB_{\frac{1}{t}}$, $X_0 = 0$, is also a Brownian motion.

5. Let $\mathbb{P}$ be a probability measure on $[0,1]$, equipped with its Borel σ-algebra Bor.

a) Use a monotone class argument to prove that the collection $\mathcal{C}$ of measurable subsets B such that

$$\mathbb{P}(B) = \inf\big\{\mathbb{P}(O)\,;\, O \text{ open set containing } B\big\} = \sup\big\{\mathbb{P}(C)\,;\, C \text{ closed subset of } B\big\}$$

is a σ-algebra.

b) Deduce that *for any $\varepsilon > 0$ and any measurable set $A \in$*Bor *there exists a* compact *subset K of A such that $\mathbb{P}(A\backslash K) \leqslant \varepsilon$.* ($\mathbb{P}$ is said to be inner regular.)

6. Let (S, d) be a metric space. An S-valued sequence $(X_n)_{n\geqslant 0}$ of random variables converges in probability to X if$_{\text{def}}$ $\mathbb{P}\big(d(X_n, X) > \varepsilon\big) \underset{n+\infty}{\longrightarrow} 0$ for any $\varepsilon > 0$, or, equivalently (why?), if $\mathbb{E}[d(X_n, X) \wedge 1] \underset{n+\infty}{\longrightarrow} 0$.

a) Prove that if $(X_n)_{n\geqslant 0}$ converges almost-surely or in probability to X then it converges weakly to X.

b) Find a weakly converging sequence which does not converge in probability.

7. Denote by $\mathcal{B}_b(\mathbb{R})$ the set of real-valued bounded measurable functions on $\mathbb{R}$ and define the $\mathcal{B}_b(\mathbb{R})^*$-topology as in definition 16, with $\mathcal{B}_b(\mathbb{R})$ in place of $\mathcal{C}_b(\mathbb{R})$. What difference is there between the notions of $\mathcal{C}_b(\mathbb{R})^*$ and $\mathcal{B}_b(\mathbb{R})^*$ convergence?

8. Let $(\mu_n)_{n\geqslant 0}$ be a sequence of probability measures on $\mathbb{R}$. Prove that it converges weakly to some probability μ iff $(f, \mu_n) \to (f, \mu)$ for any continuous function *with compact support*.

9. Suppose $\mu_n \xrightarrow{d} \mu$. Prove that the characteristic function of μ_n converges uniformly on bounded sets of $\mathbb{R}$ to the characteristic function of μ.

10. *Equicontinuity and tightness.* Let $(\mu_n)_{n\geqslant 0}$ be a sequence of probability measures on $\mathbb{R}$ and $\{\phi_n\}_{n\geqslant 0}$ be the sequence of their characteristic functions. Prove that the sequence $(\mu_n)_{n\geqslant 0}$ is tight iff the family $\{\phi_n\}_{n\geqslant 0}$ is equicontinuous at 0.

11. *Glivenki-Cantelli lemma.* Use the representation $X_n = G_n(U)$ of a random variable given in §2.3 to prove the following statement, due to Glivenko and Cantelli. Given a sequence $(X_k)_{k\geqslant 0}$ of iid random variables with distribution F, denote by $\widehat{F}_n$ the empirical distribution of the n-uple $(X_1, \dots, X_n)$:

$$\widehat{F}_n(t) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \leqslant t}.$$

Prove that

$$\sup_{t\in\mathbb{R}} \big|\widehat{F}_n(t) - F(t)\big| \to 0$$

as n goes to ∞.

12. Use the almost-surely representation of weakly converging random variables (theorem 26) to answer part or all of the following questions.

a) Find a sequence $(X_n)_{n\geqslant 0}$ of real-valued random variables converging weakly but not in probability. Prove yet that if the weak limit is a *constant random variable* then the convergence holds in probability.

b) Use this result to prove the following fact, due to Slutski (and useful in statistics). Suppose $(X_n)_{n\geqslant 0}$ has values in an interval I and that there exists some constant m such that $\sqrt{n}(X_n - m)$ converges in law to a centered Gaussian random variable with variance σ^2. Let f be a differentiable function defined on I. Prove that $\sqrt{n}(f(X_n) - f(m))$ converges in law to a centered Gaussian arv with variance $\sigma^2\big(f'(m)\big)^2$.

13. Find a modification X of the constant process $Y \equiv 0$ which is not indistinguishable of Y.

14. The purpose of this exercice is to give an example in which we have convergence of finite-dimensional distributions without convergence in law.

a) Let $(S, d), (S', d')$ be metric spaces and $f : S \to S'$ be a continuous map. Let $(\mu_n)_{n\geqslant 0}$ be a weakly convergent sequence of probability measures on $(S, \mathcal{S})$, with limit μ. Prove that the image measure of μ_n by f converge weakly to the image measure of μ by f.

b) Set $f(t) = 1 - |t|$ for $|t| \leqslant 1$ and 0 elsewhere. Let U be a random variable carried by some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and uniformly distributed on $\left[\frac{1}{3}, \frac{2}{3}\right]$. For $\omega \in \Omega$ and $t \in [0,1]$, define for $n \in \mathbb{N}$,

$$X_n(t,\omega) := f\big(3^n(t - U_\omega)\big), \quad \text{and} \quad X(t,\omega) := 0.$$

Make a picture of $X_n(\cdot,\omega)$ for a fixed ω. Consider X_n and X as $\mathcal{C}\big([0,1],\mathbb{R}\big)$-valued random variables. Prove that X_n does not converge in law to X despite the almost-surely convergence $X_n(t) \to X(t)$ for every t. What is missing?

15. *Brownian motion conditionned to be equal to* 0 *at time* 1. Let $\mathbb{P}$ be Wiener measure on $\mathcal{C}\big([0,1]\big)$ and X the canonical coordinate process (a Brownian motion under $\mathbb{P}$). Given $\varepsilon > 0$, define the law $\mathbb{P}_\varepsilon$ of X conditionned to have value in $[0,\varepsilon]$ at its final time: $\mathbb{P}_\varepsilon(A) = \mathbb{P}\big(A \,|\, X_1 \in [0,\varepsilon]\big)$, for any Borel set A of $\mathcal{C}\big([0,1]\big)$. Define also $X_t^0 = X_t - tX_1$, for any $t \in [0,1]$, and denote by $\mathbb{P}_0$ the distribution of X^0. The aim of this problem is to prove that $\mathbb{P}_\varepsilon$ converges in distribution to $\mathbb{P}_0$. In this sense, X^0 represents a Brownian motion conditionned to have value 0 at time 1; it is called a ***Brownian bridge***. Recall why it is sufficient to prove that

$$\overline{\lim_{\varepsilon \searrow 0}}\, \mathbb{P}_\varepsilon(F) \leqslant P_0(F), \tag{3.1}$$

for any closed set F of $\mathcal{C}\big([0,1]\big)$.

a) Given any times $t_i \in [0,1]$ and real (measurable) sets $B, (B_i)_{i=1..n}$, $n \geqslant 1$, prove that we have

$$\mathbb{P}\big(X_{t_1}^0 \in B_1, \ldots, X_{t_n}^0 \in B_n, X_1 \in B\big) = \mathbb{P}\big(X_{t_1}^0 \in B_1, \ldots, X_{t_n}^0 \in B_n\big)\, \mathbb{P}(X_1 \in B).$$

Why does this imply that $\mathbb{P}\big(X^0 \in A \,|\, 0 \leqslant X_1 \leqslant \varepsilon\big) = \mathbb{P}\big(X^0 \in A\big)$, for any Borel set A of $\mathcal{C}\big([0,1]\big)$?

b) Show how to get (3.1) from that point.

4. Complements to part I

4.1. ***Complement: Separable Banach spaces.*** Recall that a Banach space is a complete metric space. The space $[0,1]^{\mathbb{N}}$, equipped with the distance $d(x,x') = \sum_{n \geqslant 0} 2^{-n}|x_n - x'_n|$ is for example a separable Banach space. Its universal role is emphasized by the following theorem[46].

Theorem **39.** *Any separable Banach space is* homeomorphic *to a measurable subset of* $[0,1]^{\mathbb{N}}$.

Proof – Given a separable metric space (E,d), denote by $(z_p)_{p \geqslant 0}$ a dense sequence of points of E and define for each $p \geqslant 0$

$$f_p(x) = \frac{d(x,z_p)}{1 + d(x,z_p)}, \quad x \in E;$$

this is a continuous (and hence measurable) $[0,1]$-valued function on E. Therefore, the formula

$$f(x) = \big(f_p(x)\big)_{p \geqslant 0}$$

defines a continuous injective function from E into $[0,1]^{\mathbb{N}}$ (check it). Supposing that $f(x_n)$ converges to $f(x)$, we must have $d(x_n,z_p) \to d(x,z_p)$ for each $p \geqslant 0$, from which we easily deduce that x_n converges to x. This proves that f^{-1} is continuous on $f(E)$, that is, f is a homeomorphism from E to $f(E)$.

[46]The next two theorems and their proofs are essentially taken from Appendix 1 from Dynkin and Yushkevich's book [DY79].

Suppose in addition that the space is complete, so that it is a separable Banach space. To see that $f(E)$ is a measurable subset of $[0,1]^{\mathbb{N}}$, recall that we have seen in the proof of Prohorov's theorem that E can be written as an increasing union of compact sets[47] K_n. As each $f(K_n)$ is a compact set of $[0,1]^{\mathbb{N}}$, by continuity, it is measurable. This shows that $f(E) = \bigcup_{n\geqslant 1} f(K_n)$ is measurable[48]. ▷

Theorem **40.** *The space* $[0,1]^{\mathbb{N}}$*, equipped with its Borel* σ*-algebra, is* isomorphic *to a measurable subset of* $[0,1]$*. As a consequence, any measurable subset of a separable Banach space*[49] *is a Borel space.*

Proof – Equip $\{0,1\}^{\mathbb{N}}$ with its product σ-algebra. It is easily seen that if φ is an isomorphism from a measurable space X into Y then the formula

$$(x_n)_{n\geqslant 0} \mapsto \big(\varphi(x_n)\big)_{n\geqslant 0}$$

defines an isomorphism from $X^{\mathbb{N}}$ into $Y^{\mathbb{N}}$. Theorem 40 will thus be established if we can

a) construct an isomorphism φ from $[0,1]$ into $\{0,1\}^{\mathbb{N}}$,

b) prove that the spaces $\big\{\{0,1\}^{\mathbb{N}}\big\}^{\mathbb{N}}$ and $\{0,1\}^{\mathbb{N}}$ are isomorphic,

c) prove that the space $\{0,1\}^{\mathbb{N}}$ is isomorphic to a measurable subset of $[0,1]$.

a) Denote by $\mathcal{D}$ the countable subset of $\{0,1\}^{\mathbb{N}}$ made up of sequences with only finitely many zeros. The fomula $\psi : \epsilon \mapsto \sum_{n\geqslant 1} \epsilon_n 2^{-n}$ defines an injective measurable map from $\{0,1\}^{\mathbb{N}}\backslash\mathcal{D}$ onto $[0,1]$. To show that its inverse map $\varphi : [0,1] \mapsto \{0,1\}^{\mathbb{N}}\backslash\mathcal{D}$ is also measurable it suffices to show that the preimages $\varphi^{-1}(\Gamma_k) = \psi(\Gamma_k)$ of $\Gamma_k = \{\epsilon \in \{0,1\}^{\mathbb{N}}\backslash\mathcal{D}\,;\, \epsilon_k = 0\}$ are measurable; this is clearly the case as $\psi(\Gamma_k) = \bigcup_{p=0..2^{k-1}-1} \big[\frac{2p}{2^k}, \frac{2p+1}{2^k}\big]$.

b) Given a sequence $\big(\epsilon^{(p)}\big)_{p\geqslant 0}$ of elements of $\{0,1\}^{\mathbb{N}}$, write $\epsilon^{(p)} = \big(\epsilon^{(p)}_n\big)_{n\geqslant 0}$ and set

$$\epsilon = \epsilon^{(0)}_0 \epsilon^{(0)}_1 \epsilon^{(1)}_0 \epsilon^{(0)}_2 \epsilon^{(1)}_1 \epsilon^{(2)}_0 \dots,$$

identifying $\mathbb{N}^2$ to $\mathbb{N}$. This defines a bijective map F from $\big\{\{0,1\}^{\mathbb{N}}\big\}^{\mathbb{N}}$ onto $\{0,1\}^{\mathbb{N}}$. Denote by $B^{(n)}_k$ the subsets of $\big\{\{0,1\}^{\mathbb{N}}\big\}^{\mathbb{N}}$ defined by the condition $\epsilon^{(n)}_k = 0$; these sets generate the product σ-algebra of $\big\{\{0,1\}^{\mathbb{N}}\big\}^{\mathbb{N}}$ and the sets $F\big(B^{(n)}_k\big)$ the product σ-algebra of $\{0,1\}^{\mathbb{N}}$. This proves that the maps F and F^{-1} are measurable, so F is an isomorphism.

c) We show that $\{0,1\}^{\mathbb{N}}$ can be mapped continuously and injectively into a measurable subset of $[0,1]$. To see that this map G is an isomorphism from $\{0,1\}^{\mathbb{N}}$ onto its image[50] it suffices to see that the elementary product events $\big\{\epsilon \in \{0,1\}^{\mathbb{N}}\,;\, \epsilon_n = 0\big\}$ are mapped onto measurable sets; this is the case as these events being compact sets, their image by the continuous map G are compact, hence measurable, subsets of $[0,1]$.

The map G is simply defined by the formula

$$G(\epsilon) = \sum_{n\geqslant 0} 2\epsilon_n 3^{-n-1};$$

its continuity and injective character are easily checked. ▷

[47]The compact set K constructed in the proof of theorem 31 is a typical element of this union, obtained by letting η decrease to 0. The completeness hypothesis on the space is needed to prove that the set K constructed in that proof is compact.

[48]With a little bit of extra work, it also shows that $f(E)$ is a countable intersection of open sets of $[0,1]^{\mathbb{N}}$.

[49]Equipped with the trace σ-algebra of the ambient space.

[50]To see in particular that $G\big(\{0,1\}^{\mathbb{N}}\big)$ is a measurable subset of $[0,1]$.

4.2. ***Complement: Lebesgue measure on*** $[0,1]$**.** Let $(S,\mathcal{S})$ be a Borel space. We have seen in the proof of theorem 10 that any $(S,\mathcal{S})$ is isomorphic (first to a measurable subset of $[0,1]$, by definition, and then) to a measurable subset of $\{0,1\}^{\mathbb{N}}$; so, constructing a probability measure on $(S,\mathcal{S})$ amounts to construct a (Borel) probability measure on $\{0,1\}^{\mathbb{N}}$. The enormous advantage of this space is that is has an extremelly simple generating algebra: the *countable* collection $\mathcal{A}$ of cylindrical sets[51]. As these sets are at the same time open and closed, and so compact, a finitely additive set function on $\mathcal{A}$ will automatically satisfy condition *ii)'* of Caratheodory's extension theorem.

THEOREM **41.** *Borel probability measures on* $\{0,1\}^{\mathbb{N}}$ *correspond bijectively to additive set functions on* $\mathcal{A}$*, equal to* 0 *on* $\emptyset$ *and* 1 *on* Ω*.*

Setting $\mu\big(\{0\}\big)=\frac{1}{2}$ and $\mu\big(\{1\}\big)=\frac{1}{2}$, it follows that the product probability measure $\mu^{\otimes\mathbb{N}}$ is well defined on the product σ-algebra of $\{0,1\}^{\mathbb{N}}$. The image measure of $\mu^{\otimes\mathbb{N}}$ by the map $(\varepsilon_n)_{n\geqslant 0}\to\sum\varepsilon_n 2^{-n-1}\in[0,1]$ is Lebsegue measure.

4.3. ***Complement: Isomorphism of Borel probability spaces.*** A Borel space $(S,\mathcal{S})$ is by definition isomorphic to a measurable subset of $[0,1]$. Theorem 43 below essentially states that any probability measure on $(S,\mathcal{S})$ can be constructed as the image measure of Lebesgue measure on $[0,1]$ by some "isomorphism". This means that *all the theory developed in this course has actually a unique framework:* $[0,1]$ *with Lebesgue measure*; in particular no abstract measure theory is needed. The statement of theorem 43 requires the following definition.

DEFINITION **42.** *Two probability sapce* $(\Omega,\mathcal{F},\mathbb{P})$ *and* $(\Omega',\mathcal{F}',\mathbb{P}')$ *are said to be* ***isomorphic modulo zero*** *if*$_{def}$ *there exists* $\Omega_0\in\mathcal{F}$*,* $\Omega'_0\in\mathcal{F}'$ *with* $\mathbb{P}(\Omega_0)=\mathbb{P}'(\Omega'_0)=1$ *and an isomorphism* f *between* Ω_0 *and* Ω'_0 *such that* $\mathbb{P}'$ *is the image measure of* $\mathbb{P}$ *by* f *([52]).*

We shall write λ for Lebesgue measure on $[0,1]$ and $\mathcal{D}$ for the λ-completion of its Borel σ-algebra.

THEOREM **43.** *Any Borel probability space* $(S,\mathcal{S},\mathbb{P})$*, without atoms, is isomorphic modulo zero to* $\big([0,1),\mathcal{D},\lambda\big)$*.*

PROOF – The proof is simple and starts by identifying $(S,\mathcal{S})$ to a measurable subset of $[0,1]$ and then to a measurable subset of $\{0,1\}^{\mathbb{N}}$ (as in the proof of theorem 40). We shall now consider $\mathbb{P}$ as a probability on the product σ-algebra $\mathcal{F}$ of $\{0,1\}^{\mathbb{N}}$. Adopt the notations $\mathcal{C}_p$ for $\{0,1\}^{[\![0,p]\!]}$ and $X_p:\{0,1\}^{\mathbb{N}}\to\{0,1\}$ for the p^{th} projection, $p\geqslant 0$. We are going to construct by induction for any $z=(z_0,\dots,z_p)\in\mathcal{C}_p$ an interval $I(z)=\big[\alpha(z),\beta(z)\big)$ of Lebesgue measure $\beta(z)-\alpha(z)=\mathbb{P}(X_0=z_0,\dots,X_p=z_p)$.

Set $I(0)=\big[0,\mathbb{P}(X_0=0)\big)$ and $I(1)=\big[\mathbb{P}(X_0=0),1\big)$. Suppose $I(z)$ was constructed for any $z\in\mathcal{C}_k, k\leqslant p$ and let $z=(z_0,\dots,z_p,z_{p+1})\in\mathcal{C}_{p+1}$; set $\widetilde{z}=(z_0,\dots,z_p)$.

- If $z_{p+1}=0$, set $\alpha(z)=\alpha(\widetilde{z})$ and $\beta(z)=\alpha(\widetilde{z})+\mathbb{P}(X_0=z_0,\dots,X_p=z_p,X_{p+1}=z_{p+1})$.
- If $z_{p+1}=1$, set $\alpha(z)=\alpha(\widetilde{z})+\mathbb{P}\big(X_0=z_0,\dots,X_p=z_p,X_{p+1}=z_{p+1}\big)$ and $\beta(z)=\beta(\widetilde{z})$.

Set then for any $n\geqslant 1$

$$B_n=\bigcup_{\widetilde{z}\in\mathcal{C}_{n-1}} I\big((\widetilde{z},1)\big);$$

[51]Where only a finite number of coordinates are specified.

[52]$\mathbb{P}'$ and $\mathbb{P}$ have to be understood as defined on the σ-algebras $\{\Omega'_0\cap A'\,;\,A'\in\mathcal{F}'\}$ and $\{\Omega_0\cap A\,;\,A\in\mathcal{F}\}$ respectively.

it is easily checked that $I(z) = I\big((z_0,\dots,z_n)\big) = B_0^{z_0} \cap \cdots \cap B_n^{z_n}$, where we write B^0 for $[0,1)\backslash B$ and B^1 for B. One has $\sup_{z\in\mathcal{C}_n} \lambda\big(I(z)\big) \underset{n+\infty}{\longrightarrow} 0$. Indeed, would the converse happen we could construct by induction an element $z \in \{0,1\}^{\mathbb{N}}$ such that $\lambda\big(I(z_0,\dots,z_n)\big) \geqslant \varepsilon$ for all $n \geqslant 0$ and a positive constant ε. We would then have on the one hand $\mathbb{P}\big(\{z\}\big) = 0$, since $\mathbb{P}$ has no atoms, and on the other hand

$$\mathbb{P}\big(\{z\}\big) = \lim_{n+\infty} \searrow \mathbb{P}(X_0 = z_0,\dots,X_n = z_n) = \lim_{n+\infty} \searrow \lambda\big(I(z_0,\dots z_n)\big) \geqslant \varepsilon,$$

leading to a contradiction. It follows that the family $\mathcal{B} = (B_n)_{n\geqslant 0}$ is a basis of the topology of $[0,1)$. Define

$$\phi_{\mathcal{B}} : x \in [0,1) \mapsto \big(\mathbf{1}_{B_n}(x)\big)_{n\geqslant 0 \in \{0,1\}^{\mathbb{N}}}$$

and check that $\mathbb{P}$ is the image measure of λ by $\phi_{\mathcal{B}}$: this map is an isomorphism modulo zero between $\big([0,1),\mathcal{D},\lambda\big)$ and $\big(\{0,1\}^{\mathbb{N}},\mathcal{F},\mathbb{P}\big)$. ▷

You will find in appendix 1 of Dynkin and Yushkevich's book [DY79], or chapter 13 of Dudley's book [Dud02], a clear and definitive account on Borel spaces. Up to isomorphism (and not only isomorphism modulo 0) there exists only three types of Borel spaces: the finite spaces, $\mathbb{N}$ and the interval $[0,1]$.

4.4. ***Complement: Riesz representation theorem.*** We show in this complement how the proof of Caratheodory's extension theorem given in section 1.2 quickly leads to F. Riesz representation theorem. Given a topological space $(X,\mathfrak{X})$, denote by $\mathcal{C}_c(X)$ the set of continuous real-valued functions on X with compact support, equipped with the supremum norm.

Theorem **44.** *Let $(X,\mathfrak{X})$ be a locally compact topological space and $E : \mathcal{C}_c(X) \to \mathbb{R}$ be a positive linear form of norm 1. Then there exists a probability measure P on the Borel σ-algebra of X such that $E(f) = \int f(x)\,P(dx)$, for all $f \in \mathcal{C}_c(X)$.*

Proof – We first check that the suitable analogues of the conditions of Caratheodory's theorem hold here. Condition *i)* states that $\mu(\emptyset) = 0$ and $\mu(\Omega) = 1$. Its analogue here, $E(0) = 0$ and $E(\mathbf{1}) = 1$, is guaranteed by the linearity and the positivity and unit norm of the operator E. Countable additivity of E on $\mathcal{C}_c(X)$ is automatic! Indeed, any decreasing sequence of elements of $\mathcal{C}_c(X)$ converging to 0 pointwise actually converges uniformly to 0. As E has unit norm, it follows that $E(f_n)$ decreases to 0 if $f_n \in \mathcal{C}_c(X)$ decreases pointwise to $0 \in \mathcal{C}_c(X)$.

We can now copy word by word our proof of Caratheodory's extension theorem, but with $\mathcal{C}_c^+(X)$ in the role of the algebra $\mathcal{A}$, the set of non-negative real-valued functions on X in the role of $\mathfrak{P}(\Omega)$, and the operation $f\Delta g := f + g - 2f \wedge g$ in the role of $A\Delta B$. The σ-algebra generated by a family $\mathfrak{B}_0$ of functions is the smallest class of functions $\mathfrak{B}$ containing $\mathfrak{B}_0$ and closed by pointwise passage to the limit.

As a result, E has a unique extension into a linear functional of norm 1 on the set of bounded[53] functions belonging to the σ-algebra generated by $\mathcal{C}_c(X)$. This σ-algebra is also generated by the indicators of sets of the form $f^{-1}\big((a,b)\big)$, $a < b$ reals. As the space is locally compact, it coincides with the σ-algebra of Borel-measurable bounded functions.

[53]Use the monotone class theorem for functions for the uniqueness part of the statement; it deals with bounded functions only.

Denote by P the restriction of E to the indicators of Borel sets. It remains to prove that $E(f) = \int f(x)\, P(dx)$. As f is a uniform limit of elementary functions $\sum a_i \mathbf{1}_{A_i}$, with $a_i \in \mathbb{R}$ and Borel sets A_i, the results follows from the definition of the integral with respect to P and the fact that the extension of E has unit norm. $\triangleright$

Part II. Dynamic theory of stochastic processes

Recall that Kolmogorov's view on Natural random phenomena is a two levels theory: the random phenomenon itself is modelled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the set of possible outcomes and $\mathcal{F}$ is the set of observable events, while our experimental observations are modelled by a process $(X_t)_{t\in T}$, whose index set corresponds to the different types of measures of the phenomenon one can make. A random surface will for instance be described by a random process with index set a subset of $\mathbb{R}^2$ (or $\mathbb{S}^2$ as we live on Earth).

The first part of the course was devoted to constructing model probability spaces and processes. This task being done, we are now going to study random processes for themselves without paying attention to the background $(\Omega, \mathcal{F})$ anymore. More specifically, we are going to study random processes indexed by some sort of time: $\{1, \ldots, n\}, \mathbb{N}$ or an interval of $[0, +\infty]$. In this framework, it is natural to enrich our description of Nature by adding to $(\Omega, \mathcal{F}, \mathbb{P})$ the information on everything which has happened up to time t. This information is encoded under the form of an increasing family $(\mathcal{F}_t)_{t\in T}$ of σ-algebras. Being non-ubiquitous, our knowlegge of the history of the world up to time t is only partial (we cannot observe everything, but information also needs some time to travel, may be damaged during that travel, we may only be able to understand part of it...), so will be represented by an *increasing* family of σ-algebras $\mathcal{G}_t \subset \mathcal{F}_t$. (Note the optimistic character of this model: we do not forget our past.)

How can we then understand some events on which we have only a partial information? The introduction of the concept of conditional expectation will provide a mathematical answer; it will also provide a conceptual framework in which talking about "constants of motion", increasing/decreasing predictions, under the form of martingales, sub/supermartingales. Our main task in this part will be to understand the asymptotic behaviour of these "constants of motion", first in a discrete time setting, and then in a continuous time setting.

No topological hypotheses are made on the measurable spaces $(\Omega, \mathcal{F})$ used in this part. Given a probability $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{F})$ and a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$, we write $\mathbb{L}^1(\mathcal{G})$ for the class of integrable functions which are $\mathcal{G}$-measurables. We simply write $\mathbb{L}^1$ for $\mathbb{L}^1(\mathcal{F})$.

5. Dynamics and filtrations

5.1. **Conditional expectation.** Let us come back a moment to the considerations of the introduction. We saw there that the mathematical abstraction of the logical process of experimental research is the concept of algebra. Imagine we study a phenomenon X, with associated algebra $\mathcal{A}$. We associate to the known information about X a sub-σ-algebra $\mathcal{B}$ of $\mathcal{A}$. It is everyday's task of scientists to ask what can be infered on the 'law' of the phenomenon from the knowledge of $\mathcal{B}$. What predictions can we make? Can we quantify their quality? etc. These questions are easier to handle mathematically in the idealised framework of measurable spaces, where algebras have been replaced by σ-algebras. Roughly speaking, we may ask: Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$, how well can we approximate an $\mathcal{F}$-measurable random variable by a $\mathcal{G}$-measurable random variable?

$\mathbb{L}^2$ spaces, with their Hilbert structure, provide a good framework in which talking about approximation; a discrete framework also provides a playground where intuition is easy to formalise; we shall start with it. The construction of conditional expectation given below will make it clear that both views coincide.

5.1.1. *Discrete case.* The discrete case consists of the datum of a countable partition of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into events A_n of non-null probability. Set $\mathcal{G} = \sigma(A_n\,;\, n \geqslant 0)$. Let $X \in \mathbb{L}^1(\mathcal{F})$. As $\mathcal{G}$-measurable functions are constant on each atom A_n of $\mathcal{G}$, any $\mathcal{G}$-measurable approximation of X is of the form $\sum \alpha_n \mathbf{1}_{A_n}$. It is natural[54] to choose α_n as the mean of X on A_n: the ***conditional expectation of*** X ***given*** $\mathcal{G}$ is the random variable

$$Y := \sum_{n \geqslant 0} \frac{\mathbb{E}[X \mathbf{1}_{A_n}]}{\mathbb{P}(A_n)} \mathbf{1}_{A_n}.$$

It is characterized by the properties

- Y is $\mathcal{G}$-measurable,
- Y is integrable and $\mathbb{E}[X 1_A] = \mathbb{E}[Y 1_A]$ for all $A \in \mathcal{G}$.

For $\mathcal{G}$ generated by an increasing sequence of discrete σ-algebras $\mathcal{G}_n$ we could try to define $\mathbb{E}[X|\mathcal{G}]$ as the limit of the $\mathbb{E}[X|\mathcal{G}_n]$ if it exists. Although this constructive approach works (see theorem 65 below), the above characterization suggests a simpler and more general definition/construction procedure in accordance with the $\mathbb{L}^2$ idea of best approximation as a projection.

5.1.2. *General case: Existence and uniqueness.*

Definition/Proposition **45.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Let $X \in \mathbb{L}^1(\mathcal{F})$. Then there exists a random variable Y such that*

(1) *Y is $\mathcal{G}$-measurable,*
(2) *Y is integrable and $\mathbb{E}[X 1_A] = \mathbb{E}[Y 1_A]$ for all $A \in \mathcal{G}$;*

two such random variables are equal $\mathbb{P}$-almost-surely.

Proof – *Uniqueness.* If Y' also satisfies conditions (1) and (2) the event $A = \{Y > Y'\}$ belongs to $\mathcal{G}$, so

$$\mathbb{E}[(Y - Y')\mathbf{1}_A] = \mathbb{E}[Y \mathbf{1}_A] - \mathbb{E}[Y' \mathbf{1}_A] = \mathbb{E}[X \mathbf{1}_A] - \mathbb{E}[X \mathbf{1}_A] = 0$$

by property (2). This equality prevents A from having a positive probability. We prove in the same way the $\mathbb{P}(Y < Y') = 0$.

Existence. Assume to begin that $X \in \mathbb{L}^2(\mathcal{F})$. Since $V := L^2(\mathcal{G})$ is a closed subspace of $\mathbb{L}^2(\mathcal{F})$, we have $X = Y + W$ for some $Y \in V$ and $W \in V^\perp$. Then, for any $A \in \mathcal{G}$, we have $\mathbf{1}_A \in V$, so

$$\mathbb{E}[X \mathbf{1}_A] - \mathbb{E}[Y \mathbf{1}_A] = \mathbb{E}[W \mathbf{1}_A] = 0.$$

Hence Y satisfies (1) and (2).

Assume now that X is any non-negative random variable. Then $X_n = X \wedge n \in \mathbb{L}^2(\mathcal{F})$ and $0 \leqslant X_n \uparrow X$ as $n \to \infty$. We have shown, for each n, that there exists $Y_n \in \mathbb{L}^2(\mathcal{G})$ such that, for all $A \in \mathcal{G}$,

$$\mathbb{E}[X_n \mathbf{1}_A] = \mathbb{E}[Y_n \mathbf{1}_A]$$

[54] We may be guided in this choice by the fact that for a random variable U with a second moment, $\mathbb{E}[U]$ is the constant which *minimizes* the quantity $\mathbb{E}\big[|U - c|^2\big]$, seen as a function of c.

and moreover that $0 \leqslant Y_n \leqslant Y_{n+1}$ a.s.. Set $Y = \lim_{n\to\infty} Y_n$, then Y is $\mathcal{G}$-measurable and, by monotone convergence, for all $A \in \mathcal{G}$,

$$\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Y\mathbf{1}_A].$$

In particular, if $\mathbb{E}[X]$ is finite then so is $\mathbb{E}[Y]$. Finally, for a general integrable random variable X, we can apply the preceding construction to X^- and X^+ to obtain Y^- and Y^+. Then $Y = Y^+ - Y^-$ satisfies (1) and (2). ▷

5.1.3. *Properties of conditional expectation.*

PROPOSITION **46** (Simple properties). 1. *Let π be a π-system generating $\mathcal{G}$. If $Y \in \mathbb{L}^1(\mathcal{G})$ satisfies $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$ for any $A \in \pi$, then $Y = \mathbb{E}[X|\mathcal{G}]$.*

2.
   - *For any $Z \in \mathbb{L}^\infty(\mathcal{G})$ we have $\mathbb{E}\big[Z\,\mathbb{E}[X|\mathcal{G}]\big] = \mathbb{E}[ZX]$.*
   - *$\mathbb{E}[X|\mathcal{G}] \geqslant 0$ if $X \geqslant 0$.*
   - *If $\mathcal{G}$ and $\sigma(X)$ are independent then Y is constant equal to $\mathbb{E}X$.*
   - *We have $\mathbb{E}[\alpha X + \beta X'|\mathcal{G}] = \alpha\,\mathbb{E}[X|\mathcal{G}] + \beta\,\mathbb{E}[X'|\mathcal{G}]$, for any $\alpha, \beta \in \mathbb{R}, X' \in \mathbb{L}^1$,*
3. *(Conditional Jensen's inequality) For any convex function f such that $f(X) \in \mathbb{L}^1(\mathcal{F})$ we have*
$$f\big(\mathbb{E}[X|\mathcal{G}]\big) \leqslant \mathbb{E}[f(X)|\mathcal{G}].$$
In particular, if $X \in \mathbb{L}^p$ for some $p \in [1, +\infty)$ then $\big\|\mathbb{E}[X|\mathcal{G}]\big\|_p \leqslant \|X\|_p$.
4. *(We can take out what is known) If Z is bounded and $\mathcal{G}$-measurable, then $\mathbb{E}[ZX|\mathcal{G}] = Z\,\mathbb{E}[X|\mathcal{G}]$ almost-surely.*

PROOF – 1. Use the monotone class theorem for functions.
2. Use the monotone convergence theorem to prove the first statement. For the second note that we can have $\mathbb{E}\big[\mathbf{1}_{\mathbb{E}[X|\mathcal{G}]<0}\mathbb{E}[X|\mathcal{G}]\big] = \mathbb{E}[\mathbf{1}_{\mathbb{E}[X|\mathcal{G}]<0}X] \geqslant 0$ only if $\mathbb{P}\big(\mathbb{E}[X|\mathcal{G}] < 0\big) = 0$. The two other properties are checked verifying that the asserted quantities satisfy the characterization of conditional expectation.
3. As f is convex it is the supremum of a countable family of affine functions:

$$f(x) = \sup_i\,(a_i x + b_i),\ x \in \mathbb{R}.$$

Hence, almost-surely , for all i,

$$a_i\,\mathbb{E}[X|\mathcal{G}] + b_i \leqslant \mathbb{E}[f(X)|\mathcal{G}],$$

that is $f\big(\mathbb{E}[X|\mathcal{G}]\big) \leqslant \mathbb{E}[f(X)|\mathcal{G}]$.
4. Check that $Z\,\mathbb{E}[X|\mathcal{G}]$ satisfies the properties (1) and (2). ▷

PROPOSITION **47** (Conditional versions of convergence theorems). 5. (Monotone convergence) *If one has almost-surely $0 \leq X_n \leqslant X$ then one has almost-surely $\mathbb{E}[X_n|\mathcal{G}] \leqslant \mathbb{E}[X|\mathcal{G}]$.*

6. (Fatou lemma) *If $X_n \geqslant 0$ for all n, then one has almost-surely* $\mathbb{E}[\liminf X_n|\mathcal{G}] \leqslant \underline{\lim}\,\mathbb{E}[X_n|\mathcal{G}]$.
7. (Dominated convergence) *If X_n converges almost-surely to X and $|X_n|$ is dominated by an integrable random variable for all n, then $\mathbb{E}[X_n|\mathcal{G}]$ converges alomst-surely to $\mathbb{E}[X|\mathcal{G}]$.*

PROOF – 5. If $0 \leqslant X_n$ increases almost-surely to some random variable X, then $\mathbb{E}[X_n|\mathcal{G}]$ increases almost-surely to some $\mathcal{G}$-measurable random variable U; so, by monotone convergence, for all $A \in \mathcal{G}$,

$$\mathbb{E}[X\mathbf{1}_A] = \lim \mathbb{E}[X_n\mathbf{1}_A] = \lim \mathbb{E}\big[\mathbb{E}[X_n|\mathcal{G}]\mathbf{1}_A\big] = \mathbb{E}[U\mathbf{1}_A]$$

Fatou lemma (6) and dominated convergence (7) follow by essentially the same arguments as in the original results. ▷

To state the fundamental property 10, recall that a family $(X_t)_{t\in T}$ of real-valued random variables is ***uniformly integrable*** if$_{\text{def}}$

$$\sup_{t\in T} \mathbb{E}\big[|X_t|\mathbf{1}_{|X_t|>m}\big] \to 0 \qquad \text{as } m \to +\infty.$$

Proposition **48** ($\mathbb{E}(X|\mathcal{G})$ as a function of $\mathcal{G}$)**.** 8. (Tower property) *If $\mathcal{H} \subset \mathcal{G}$, then*

$$\mathbb{E}\big[\mathbb{E}[X|\mathcal{G}]\big|\mathcal{H}\big] = \mathbb{E}[X|\mathcal{H}].$$

9. *If $\sigma(X, \mathcal{G})$ is independent of $\mathcal{H}$, then $\mathbb{E}\big[X\big|\sigma(\mathcal{G},\mathcal{H})\big] = \mathbb{E}[X|\mathcal{G}]$.*
10. (Uniform integrability) *Let $X \in \mathbb{L}^1$. Then the set of random variables Y of the form $Y = \mathbb{E}[X|\mathcal{G}]$, where $\mathcal{G} \subset \mathcal{F}$ is a σ-algebra, is uniformly integrable.*

Proof – 8. Just check conditions (1) and (2).

9. Using property 1 it is sufficient to check that we have

$$\mathbb{E}\big[\mathbb{E}[X|\sigma(\mathcal{G},\mathcal{H})]\mathbf{1}_{A\cap B}\big] = \mathbb{E}\big[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_{A\cap B}\big]$$

for any $A \in \mathcal{G}$ and $B \in \mathcal{H}$, as the set of such $A \cap B$ is a π-system generating $\sigma(\mathcal{G},\mathcal{H})$. But the left hand side equals

$$\mathbb{E}[X\mathbf{1}_{A\cap B}] \overset{\text{hyp.}}{=} \mathbb{E}\big[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_A\big]\mathbb{P}(B) = \mathbb{E}\big[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_{A\cap B}\big].$$

10. We can find $\delta > 0$ so that $\mathbb{E}\big[|X|\mathbf{1}_A\big] \leqslant \varepsilon$ whenever $\mathbb{P}(A) \leqslant \delta$. Then choose $\lambda < \infty$ so that $\mathbb{E}[|X|] \leqslant \lambda\delta$. Suppose $Y = \mathbb{E}[X|\mathcal{G}]$, then $|Y| \leqslant \mathbb{E}\big[|X|\big|\mathcal{G}\big]$. In particular, $\mathbb{E}[|Y|] \leqslant \mathbb{E}[|X|]$ so

$$\mathbb{P}(|Y| \geqslant \lambda) \leqslant \lambda^{-1}\mathbb{E}[|Y|] \leqslant \delta.$$

Then

$$\mathbb{E}\big[|Y|\mathbf{1}_{|Y|\geqslant\lambda}\big] \leqslant \mathbb{E}\big[|X|\mathbf{1}_{|Y|\geqslant\lambda}\big] \leqslant \varepsilon.$$

Since λ was chosen independently of $\mathcal{G}$, we are done. ▷

5.2. **Filtrations.**

Dynamics becomes real through the accumulation of knowledge as time passes[55]; filtrations are the probabilistic counterpart of this accumulation process.

5.2.1. *Generalities.*

Let I be a time index, it may be finite $\{1, \cdots, n\}$, countable $\mathbb{N}$, or an interval of $\mathbb{R}_+ \cup \{\infty\}$.

Definition **49.** *Let $(\Omega, \mathcal{F})$ be a measurable space. A **filtration** on $(\Omega,\mathcal{F})$ is a monotonic family $(\mathcal{F}_t)_{t\in I}$ of sub-σ-algebras of $\mathcal{F}$. We shall talk of the **filtered space** $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}\big)$.*

Filtrations are the mathematical counterpart of the accumulation/loss of knowledge about a phenomenon as time passes; we shall give in theorem 65 and 67 a quantitative version of this fact. Note that we do not require $\mathcal{F}_0$ to be trivial or $\mathcal{F}_\infty$ to be equal to $\mathcal{F}$ (if $\infty \in I$).

Definition **50.** *Let $X = (X_t)_{t\in I}$ be a random process defined on a measurable space $(\Omega,\mathcal{F})$. The **filtration generated by** X is defined by the formula*

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

[55] There is no dynamics without memory, which enables one to compare what happens at different times.

Given $t \in I$ we denote by $\bigvee_{s<t} \mathcal{F}_s$ the σ-algebra generated by[56] $\bigcup_{s<t} \mathcal{F}_s$. Be careful, filtrations have no reason to be a priori continuous on the left: we may have $\bigvee_{s<t} \mathcal{F}_s \subsetneq \mathcal{F}_t$. Think of a process which is constant on $[0,t)$ and has a random (non-null) jump at time t. We may as well have $\bigcap_{s>t} \mathcal{F}_s \neq \mathcal{F}_t$. These fact motivate the following definition. Given a filtration $\left(\mathcal{F}_t\right)_{t\in I}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, set for any $t \geqslant 0$

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

This defines a new (and bigger) filtration where we allow ourselves to look slightly ahead in time; it is continuous on the right.

Definition **51.**
- *Let $\left(\Omega, \mathcal{F}, \left(\mathcal{F}_t\right)_{t\geqslant 0}\right)$ be a filtered space. A random process $X = (X_t)_{t\in I}$ on $(\Omega, \mathcal{F})$ is said to be* ***adapted*** *to $(\mathcal{F}_t)_{t\geqslant 0}$ if$_{def}$ $\mathcal{F}^X_t \subset \mathcal{F}_t$ for all $t \in I$.*
- *Let $\left(\Omega, \mathcal{F}, \left(\mathcal{F}_n\right)_{n\geqslant 0}\right)$ be a filtered space. A random process $X = (X_n)_{n\geqslant 0}$ on $(\Omega, \mathcal{F})$ is said to be $\left(\mathcal{F}_n\right)_{n\geqslant 0}$****-previsible*** *if$_{def}$ $\mathcal{F}^X_n \subset \mathcal{F}_{n-1}$ for all $n \geqslant 1$ and $\mathcal{F}^X_0 \subset \mathcal{F}_0$.*

We shall just say previsible when the context is clear.

5.2.2. *Stopping times.* Recall first that a random time is an I-valued random variable; it indicates the moment at which some event of interest happens; the σ-algebra $\mathcal{F}_t$ corresponds to our knowledge of the world at time t. Although an event may happen at time t we may not be aware of it immediately; events of which we have immediate knowledge are called stopping times.

Definition **52.** *A* ***stopping time*** *is a random time T such that $\{T \leqslant t\} \in \mathcal{F}_t$ for any $t \in I$. It is equivalent to say that the process $\left(\mathbf{1}_{T\leqslant t}\right)_{t\in I}$ is adapted.*

Fundamental example of previsible process. Given a filtered space $\left(\Omega, \mathcal{F}, \left(\mathcal{F}_n\right)_{n\geqslant 0}\right)$ and a stopping time T, the process $\left(\mathbf{1}_{n\leqslant T}\right)_{n\geqslant 0}$ is previsible.

As above, denote by (S, d) a metric space.

Definition **53** (First entrance and hitting times)**.** *Let $(X_t)_{t\geqslant 0}$ be an S-valued process and Γ be a Borel subset of S. The* ***first entrance of*** *X* ***in*** *Γ is the random time $D_\Gamma = \inf\{t \geqslant 0\,;\, X_t \in \Gamma\}$; the* ***hitting time of*** *Γ* ***by*** *X is the random time $H_\Gamma = \inf\{t > 0\,;\, X_t \in \Gamma\}$.*

These two classes of random times will be our main examples of stopping times.

Proposition **54.** *Suppose $(X_t)_{t\geqslant 0}$ is an S-valued continuous random process, and let F and O be some subsets of S, respectively closed and open. Then,*

- *D_F is an $(\mathcal{F}_t)_{t\geqslant 0}$-stopping time;*
- *D_O and H_F are $(\mathcal{F}_{t^+})_{t\geqslant 0}$-stopping times.*

Proof – D_F: Since the map $x \in S \mapsto d(x, F)$ is continuous (it is Lipschitz), the functions $\omega \to d(X_q(\omega), F)$ are measurable, for all $q \in \mathbb{Q}_+$. For $t \geqslant 0$, we have by continuity,

$$\{D_F \leqslant t\} = \left\{\inf\{d(X_q, F)\,;\, q \in \mathbb{Q} \cap [0,t)\} = 0\right\},$$

from which the $(\mathcal{F}_t)_{t\geqslant 0}$-stopping time property follows.

[56]Recall that the union of two σ-algebra may not be a σ-algebra; find a counter-example.

D_O: $\{D_O < t\} = \bigcup_{q\in\mathbb{Q}\cap[0,t]}\{X_q \in O\} \in \mathcal{F}_t$. Note that we may fail to have $\{D_O \leqslant t\} \in \mathcal{F}_t$: if for some ω, $X_t(\omega) \leqslant 1$, for $t \in [0,1)$, and $X_1(\omega) = 1$, we cannot tell wether $D_{(1,+\infty)} = 1$ or not without looking slightly ahead in time.

H_F: I leave you this case as an exercise. The fact that we just have an $(\mathcal{F}_{t^+})_{t\geqslant 0}$-stopping time comes from the fact that we cannot tell at time 0 if H_F is 0 or not... ▷

Given a stopping time T set

$$\mathcal{F}_T = \big\{A \in \mathcal{F}\,;\, A \cap \{T \leqslant t\} \in \mathcal{F}_t \text{ for all } t \in I\big\}.$$

Check that if T is constant equal to t then $\mathcal{F}_T = \mathcal{F}_t$. Given a process X we shall set $X_T(\omega) = X_{T(\omega)}(\omega)$ whenever $T(\omega) < \infty$. We also define the ***stopped process*** X^T by $X^T_t = X_{t\wedge T}$. The following facts are easily proved from the definitions.

PROPOSITION **55.** *Let T and T' be stopping times on a filtered space $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}\big)$, and let $X = (X_t)_{t\in I}$ be an adapted process. Then*

- *$T \wedge T'$ is a stopping time,*
- *if $T \leqslant T'$ then $\mathcal{F}_T \subset \mathcal{F}_{T'}$,*
- *X^T is adapted, when $I = \mathbb{N}$.*

It is also easy to prove that $X_T\mathbf{1}_{T<\infty}$ is $\mathcal{F}_T$-measurable if T is countable; this is no longer automatic if the time index is uncountable. We shall state and prove a sufficient (and useful) condition to have the conclusion later in proposition 83; this condition will essentially mean that the process is determined by its restriction to a countable set of indices.

5.3. **Martingales, supermartingales and submartingales.** Constants of motion play a dominant role in the theory of differential equations (and so in classical mechanics): knowledge of a constant of motion reduces the dimension of a problem, so if sufficiently many independent constants of motion are known then the system is integrable (at least theoretically). Liapounov functions[57] usually also provide precious informations on the dynamics; they are for instance used to prove the stability of hyperbolic zeros of vector fields under perturbation. The probabilistic counterpart of these notions are martingales and sub/supermartingales. In our framework the dynamics is not provided by the datum of a differential equation but by the datum of a filtration representing our evolving knowledge of a system as time passes.

DEFINITION **56.**

- *Let $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}\big)$ be a filtered probability space. A* ***martingale*** *is an adapted integrable process $(M_t)_{t\in I}$ such that $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$ for any $s \leqslant t$.*
- *A* ***submartingale*** *is an adapted process $(M_t)_{t\in I}$ such that $\mathbb{E}[M_t|\mathcal{F}_s] \geqslant M_s$ for any $s \leqslant t$.*
- *A* ***supermartingale*** *is an adapted process $(M_t)_{t\in I}$ such that $\mathbb{E}[M_t|\mathcal{F}_s] \leqslant M_s$ for any $s \leqslant t$.*

So, roughly speaking, submartingales play the role of increasing functions and supermartingales the role of decreasing functions. Do martingales play well their role of constant of motion? Yes: We shall see later for instance that if $\Omega = \mathcal{C}(\mathbb{R}_+, \mathbb{R})$, $\mathcal{F}$ is the Borel σ-algebra, $X_t(\omega) = \omega_t$ is the coordinate process and $\mathcal{F}_t = \mathcal{F}^X_t$ then a probability

[57]A Liapounov function for a differential equation $\dot{y}_t = f(y_t)$ is a function g such that $g(y_t)$ is monotone along any solution.

on $(\Omega, \mathcal{F})$ is the Wiener measure iff the complex-valued process $\left(e^{i\lambda X_t + \frac{\lambda^2 t}{2}}\right)_{t\geqslant 0}$ is a martingale. Similarly, a Markov chain is completely determined by the datum of a certain family of martingales. These facts should warn you of the importance of this concept. As a consequence of the conditional version of Jensen's inequality the convex image of a martingale (resp. concave) is a submartingale (resp. supermartingale).

6. Discrete time martingale theory

Let us fix a filtration $\left(\mathcal{F}_n\right)_{n\geqslant 0}$ on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Adaptedness, previsibility, martingales... are defined with respect to this set up in this section.

6.1. **Characterisation of supermartingales.** Martingales (resp. sub/super-martingales) are defined above by the "projection identities" at deterministic times. The possibility to use similar identities with *random times* is the very reason why this class of processes will happen to be so powerful.

Theorem **57** (Optional stopping theorem (1))**.** *An adapted process X is a supermartingale iff one of the following conditions hold.*

1. *For all* bounded *stopping time T and* any *stopping time S*
$$\mathbb{E}[X_T|\mathcal{F}_S] \leqslant X_{S\wedge T}.$$
2. *For all* bounded *stopping times S, T, with $S \leqslant T$*
$$\mathbb{E}[X_T] \leqslant \mathbb{E}[X_S].$$

Proof – We make a circular argument proving statement 1 first. Suppose the stopping times S and T bounded above by a constant n; we can write
$$X_T = X_{S\wedge T} + \sum_{k=0..n} (X_{k+1} - X_k)\mathbf{1}_{S\leqslant k<T}.$$
To prove 1 amounts to prove that we have for any $A \in \mathcal{F}_S$
$$\mathbb{E}[X_T \mathbf{1}_A] \leqslant \mathbb{E}[X_{S\wedge T}\mathbf{1}_A].$$
But as $A \cap \{S \leqslant k < T\} \in \mathcal{F}_k$ we have
$$\mathbb{E}\left[(X_{k+1} - X_k)\mathbf{1}_{S\leqslant k<T}\right] \leqslant 0;$$
the result follows. Statement 2 is a consequence of 1. Last, to prove that X is a supermartingale if it enjoys property 2, take integers $p < q$, an event $A \in \mathcal{F}_p$, and set $T = p\mathbf{1}_A + q\mathbf{1}_{A^c}$. This formula defines a stopping time bounded by q, so
$$\mathbb{E}[X_q] \leq \mathbb{E}[X_T],$$
i.e. $\mathbb{E}[X_q] \leqslant \mathbb{E}[X_p\mathbf{1}_A + X_q\mathbf{1}_{A^c}]$, or $\mathbb{E}[X_q\mathbf{1}_A] \leqslant \mathbb{E}[X_p\mathbf{1}_A]$. ▷

We have a similar statement for martingales, with equalities instead of inequalities.

Corollary **58.** *Given a martingale $(M_n)_{n\geqslant 0}$ and a stopping time T, the stopped process M^T is a martingale.*

6.2. **Almost-sure and $\mathbb{L}^1$-convergence results.**

6.2.1. *Non-negative martingales.* Given a non-negative martingale $(M_n)_{n\geqslant 0}$ and two positive real numbers $a<b$ define the stopping time $\sigma_1 = \inf\{p\geqslant 0\,;\, M_p\leqslant a\}$ and define inductively the stopping times[58]

$$\tau_k = \inf\{p\geqslant \sigma_k\,;\, M_p\geqslant b\},\quad \sigma_{k+1} = \inf\{p\geqslant \tau_k\,;\, M_p\leqslant a\}.$$

The number of upcrossings from a to b by the martingale $(M_n)_{n\geqslant 0}$ is equal to

$$U_{a,b} = \sup\{k\,;\,\tau_k<\infty\}\in\mathbb{N}\cup\{\infty\}.$$

PROPOSITION **59** (Dubins)**.** *We have $\mathbb{P}(U_{a,b}\geqslant k)\leqslant \left(\frac{a}{b}\right)^k$, for any $k\geqslant 0$. In particular $U_{a,b}$ is almost-surely finite.*

PROOF – As we have $\{\sigma_k<\infty\}\subset\{\tau_{k-1}<\infty\}$ it is sufficient to prove that $\mathbb{P}(\tau_k<\infty)\leqslant \frac{a}{b}\mathbb{P}(\sigma_k<\infty)$. We know from the optional stopping theorem 57 that we have for any $n\geqslant 0$

$$\mathbb{E}[M_{\tau_k\wedge n}] = \mathbb{E}[M_{\sigma_k\wedge n}],$$

i.e.

$$\mathbb{E}[M_{\tau_k}\mathbf{1}_{\tau_k\leqslant n}] + \mathbb{E}[M_n\mathbf{1}_{\tau_k>n}] = \mathbb{E}[M_{\sigma_k}\mathbf{1}_{\sigma_k\leqslant n}] + \mathbb{E}[M_n\mathbf{1}_{\sigma_k>n}].$$

Since the first term on the left is $\geqslant b\,\mathbb{P}(\tau_k\leqslant n)$ and the first on the right is $\leqslant a\,\mathbb{P}(\sigma_k\leqslant n)$, we have

$$b\,\mathbb{P}(\tau_k\leqslant n)\leqslant b\,\mathbb{P}(\tau_k\leqslant n) + \mathbb{E}\big[M_n\mathbf{1}_{\sigma_k\leqslant n<\tau_k}\big]\leqslant a\,\mathbb{P}(\sigma_k\leqslant n)$$

as M is non-negative; the inequality of the proposition follows sending n to $+\infty$. ▷

As a consequence of Dubins' result, almost-surely $U_{a,b}$ is finite for all rationals $a<b$.

COROLLARY **60** (Almost-sure convergence)**.** *A non-negative martingale converges almost-surely to an integrable random variable.*

PROOF – If not there would exists positive rational numbers $a<b$ such that $U_{a,b}=\infty$ on an event of positive probability, contradicting Dubins' proposition. Denote by M_∞ the almost-sure limit of M_n. We prove the integrability of M_∞ applying Fatou lemma in the equality $\mathbb{E}[M_n]=\mathbb{E}[M_0]$. ▷

Let S_n denote the simple symmetric random walk on $\mathbb{Z}$, stopped at the random time T when it hits -1. The process $M_n = S_n+1$ is a non-negative martingale which converges almost surely to 0; yet M_n does not converge to 0 in $\mathbb{L}^1$, as $\mathbb{E}[M_n]=1$.

DEFINITION **61.** *A* ***martingale*** $(M_n)_{n\geqslant 0}$ *is said to be* ***closed*** *if$_{def}$ it converges almost-surely to an integrable random variable M_∞ for which we can write the martingale identity $M_n=\mathbb{E}[M_\infty|\mathcal{F}_n]$, for all $n\geqslant 0$.*

6.2.2. *Almost-sure convergence of supermartingales.* The proof of Dubins' proposition 59 makes a crucial use of the non-negativeness of the martingale. The price to pay to get rid of this hypothesis is to impose to the martingale to be bounded in $\mathbb{L}^1$. This actually works for supermartingales as is implied by the following result due to Doob[59]. Let $(X_n)_{n\geqslant 0}$ be a supermartingale. Given two real numbers $a<b$, denote by, $U_{a,b}(n)$ the number of upcrossings of X from a to b made by time n; almost surely $U_{a,b}(n)$ increases to $U_{a,b}$ as $n\to+\infty$. The stopping times τ_k,σ_k are defined as above.

[58]Check that they are indeed stopping times. We adopt the convention $\inf\emptyset=+\infty$.
[59]A result prior to Dubins' result.

THEOREM **62** (Doob's upcrossing inequality). *For any supermartingale X we have for any $n \geqslant 1$*

$$\mathbb{E}\big[U_{a,b}(n)\big] \leqslant \frac{\mathbb{E}\big[(X_n - a)^-\big]}{b-a}.$$

PROOF – Given $n \geqslant 1$, set

$$S = \sum_{k \geqslant 1} \big(X_{\tau_k \wedge n} - X_{\sigma_k \wedge n}\big).$$

As τ_k and σ_k are no less than k, only the first $U_{a,b}(n) + 1 (\leqslant n+1)$ terms may be non-null; so $S = \sum_{k=1}^{n+1} \big(X_{\tau_k \wedge n} - X_{\sigma_k \wedge n}\big)$. Each of the first $U_{a,b}(n)$ terms are no smaller than $b-a$ as they correspond to upcrossings. The last (potentially non-null) term, $X_n - X_{\sigma_{U_{a,b}(n)+1} \wedge n}$ is greater than or equal to $-(X_n - a)^-$. So we have

$$\sum_{k \geqslant 1}^{n+1} \big(X_{\tau_k \wedge n} - X_{\sigma_k \wedge n}\big) \geqslant U_{a,b}(n)\,(b-a) - (X_n - a)^-. \tag{6.1}$$

But as X is a supermartingale and $\tau_k \wedge n, \sigma_k \wedge n$ bounded stopping times, we have $\mathbb{E}[X_{\tau_k \wedge n}] \leqslant \mathbb{E}[X_{\sigma_k \wedge n}]$, by the optional stopping theorem 57; the result follows. $\triangleright$

The same reasonning as in corollary 60 proves the following extension of corollary 60.

THEOREM **63** (Almost-sure convergence theorem for supermartingales). *A supermartingale bounded in $\mathbb{L}^1$ converges almost-surely to an integrable random variable.*

6.2.3. *Closed martingales.* **a) Main theorem.** To state the following necessary and sufficient condition of closedness of a martingale recall that a sequence $(X_n)_{n \geqslant 0}$ of integrable random variables converges in $\mathbb{L}^1$ to some $X (\in \mathbb{L}^1)$ iff it converges in probability to X and is uniformly integrable. This fact is a well known application of Egorov's theorem[60]. You are asked to prove that result in the example sheet.

THEOREM **64** ($\mathbb{L}^1$-convergence theorem for martingales). *A martingale is closed iff it is uniformly integrable.*

PROOF – $\Rightarrow$ has been proved in proposition 48. To establish the converse it suffices to note that a uniformly integrable martingale $(M_n)_{n \geqslant 0}$ is bounded in $\mathbb{L}^1$, so it converges almost-surely (hence in probability) to some $M_\infty \in \mathbb{L}^1$, by theorem 63. As a consequence of the above mentionned result it converges in $\mathbb{L}^1$ to M_∞. Passing to the limit in the martingale identity yields $M_n = \mathbb{E}[M_\infty | \mathcal{F}_n]$. $\triangleright$

As a direct application of this criterion we obtain Lévy's famous convergence theorem.

THEOREM **65** (Lévy's 'upward' theorem and $0-1$ law). *For every $X \in \mathbb{L}^1$ the martingale $\mathbb{E}[X|\mathcal{F}_n]$ converges almost-surely and in $\mathbb{L}^1$ to $\mathbb{E}[X|\mathcal{F}_\infty]$. In particular, $\mathbb{P}(A|\mathcal{F}_n)$ converges almost-surely to $\mathbf{1}_A$, for every $A \in \mathcal{F}_\infty$.*

b) Applications. We present here two important applications of the above two results.

• We first show that the optional stopping time theorem can be applied with *any* stopping time when working with a closed martingale.

COROLLARY **66** (Optional stopping theorem (2)). *For any uniformly integrable martingale M and* any *stopping times S, T, we have*

$$\mathbb{E}[M_T|\mathcal{F}_S] = M_{S \wedge T}.$$

[60] Egorov's theorem states that almost-sure limits are uniform outside sets of arbitrarily small measure.

PROOF – We have already proved the result when T is bounded (theorem 57). For an unbounded stopping time, approach it by $T \wedge n$ and use theorem 57 to write

$$\mathbb{E}[M_{T\wedge n}|\mathcal{F}_S] = M_{S\wedge T\wedge n}. \tag{6.2}$$

By Lévy's upward theorem 65, the right hand side $M_{S\wedge T\wedge n} = \mathbb{E}\big[M_\infty\big|\mathcal{F}_{S\wedge T\wedge n}\big]$ converges (almost-surely and) in $\mathbb{L}^1$ to $\mathbb{E}\big[M_\infty\big|\mathcal{F}_{S\wedge T}\big] = M_{S\wedge T}$. As we bhave $M_{T\wedge n} = \mathbb{E}[M_n|\mathcal{F}_T]$, by the optional stopping theorem proved so far, we deduce from the uniform integrability of the martingale (hence its $\mathbb{L}^1$ convergence) that[61] $M_T = \mathbb{E}[M_\infty|\mathcal{F}_T]$. Also, since

$$\begin{aligned}
\big\|\mathbb{E}[M_{T\wedge n} - M_T|\mathcal{F}_S]\big\|_1 &\leqslant \mathbb{E}\Big[\mathbb{E}\big[|M_{T\wedge n} - M_T|\big|\mathcal{F}_S\big]\Big] \\
&\leqslant \mathbb{E}\Big[\mathbb{E}\big[|M_{T\wedge n} - M_T|\big|\mathcal{F}_S\big]\Big] \leqslant \mathbb{E}\Big[\mathbb{E}\Big[\mathbb{E}\big[|M_n - M_\infty|\big|\mathcal{F}_T\big]\Big|\mathcal{F}_S\Big]\Big] \\
&\leqslant \mathbb{E}\big[|M_n - M_\infty|\big]
\end{aligned}$$

the result follows on passing to the limit in (6.2). ▷

• *Martingales with respect to decreasing filtrations (backward martingales).* Let $\cdots \subset \mathcal{F}_{n+1} \subset \mathcal{F}_n \cdots \subset \mathcal{F}_0 \subset \mathcal{F}$ be a decreasing filtration and set $\mathcal{F}_\infty = \bigcap_{n\geqslant 0}\mathcal{F}_n$. A ***backward martingale*** is a sequence $(M_n)_{n\geqslant 0}$ of $\mathbb{L}^1$-random variables such that

$$M_n \text{ is } \mathcal{F}_n\text{-measurable and } \mathbb{E}[M_{n-1}|\mathcal{F}_n] = M_n.$$

The great difference with (usual) martingales is that backward martingales satisfy the identity

$$M_n = \mathbb{E}[M_0|\mathcal{F}_n]$$

for every $n \geqslant 0$. The sequence $(M_n)_{n\geqslant 0}$ is thus uniformly integrable and the $\mathbb{L}^1$-convergence theorem implies the following result[62].

THEOREM **67** (Lévy's 'downward' theorem). *For all $X \in \mathbb{L}^1$ the backward martingale $\mathbb{E}[X\,|\,\mathcal{F}_n]$ converges almost-surely and in $\mathbb{L}^1$ to $\mathbb{E}[X\,|\,\mathcal{F}_\infty]$. In particular, if $(Y_n)_{n\geqslant 0}$ is a sequence of independent random variables and $\mathcal{F}_n = \sigma(Y_p\,;\,p\geqslant n)$, we have $\mathbb{E}[X\,|\,\mathcal{F}_n] \to \mathbb{E}[X]$ for any integrable random variable X.*

COROLLARY **68** (Strong law of large numbers). *Let $(X_n)_{n\geqslant 0}$ be a sequence of independent and identically distributed random variables in $\mathbb{L}^1$. Then $\frac{X_1+\cdots+X_n}{n}$ converges almost-surely and in $\mathbb{L}^1$ to the constant random variable $\mathbb{E}[X_1]$.*

PROOF – Set $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$ for $n \geqslant 1$; define also the decreasing filtration

$$\mathcal{F}_n = \sigma(S_p\,;\,p\geqslant n) = \sigma(S_n, X_p\,;\,p\geqslant n+1).$$

Since X_1 is independent of $\sigma(X_p\,;\,p\geqslant n+1)$, we have $\mathbb{E}[X_1|\mathcal{F}_n] = \mathbb{E}[X_1|S_n]$ for all n. Now, by symmetry, $\mathbb{E}[X_k|S_n] = \mathbb{E}[X_1|S_n]$ for all $1\leqslant k\leqslant n$, so we have almost-surely $\mathbb{E}[X_1|\mathcal{F}_n] = \frac{S_n}{n}$. The sequence $\big(\frac{S_n}{n}\big)_{n\geqslant 0}$ if thus a backward martingale, so it converges almost-surely and in $\mathbb{L}^1$ to $\mathbb{E}[X_1|\mathcal{F}_\infty]$. As this random variable is also, for each $k\geqslant 0$, the limit of $\frac{X_k+\cdots+X_n}{n}$, it is measurable with respect to $\bigcap_{k\geqslant 0}\sigma(X_p\,;\,p\geqslant k)$. Since this σ-algebra is trivial under $\mathbb{P}$, by Kolmogorov's $0-1$ law, $\mathbb{E}[X_1|\mathcal{F}_\infty]$ is constant, equal to $\mathbb{E}[X_1]$. ▷

[61]The stopped martingale $(M_{T\wedge n})_{n\geqslant 0}$ is uniformly integrable if M is uniformly integrable.

[62]We have not used the fact that the filtration considered in section 6.2 is increasing to prove the results of that section. They also hold for decreasing filtrations.

6.3. **$\mathbb{L}^p$-convergence results.** Given a process $(X_n)_{n\geqslant 0}$, set $X^* = \sup_{n\geqslant 0}|X_n|$. The key to the $\mathbb{L}^p$-convergence results is Doob's $\mathbb{L}^p$ inequality (6.4) below. It provides a control of the behaviour of the whole trajectory in terms of its behaviour at fixed times. Doob's upcrossing inequality plaid the same role above, in the almost-surely convergence results.

THEOREM **69** (Doob's maximal inequality)**.** *Let X be a martingale or a non-negative submartingale. Then, for all $\lambda \geqslant 0$,*

$$\lambda\mathbb{P}(X^* \geqslant \lambda) \leqslant \sup_{n\geqslant 0}\mathbb{E}\big[|X_n|\big].$$

PROOF – As X^* is the increasing limit of $X_\ell^* := \sup_{n\leqslant \ell}|X_n|$ it suffices to prove that the inequality $\lambda\mathbb{P}(X_\ell^* \geqslant \lambda) \leqslant \sup_{n\leqslant\ell}\mathbb{E}\big[|X_n|\big]$ holds for all $\ell \geqslant 1$. Also, as $|X|$ is a non-negative submartingale, it suffices to consider the case where X is a non-negative submartingale, for which we prove that $\lambda\mathbb{P}(X_\ell^* \geqslant \lambda) \leqslant \mathbb{E}[X_\ell]$.

Set $T = \inf\{n \geqslant 0 : X_n \geqslant \lambda\} \wedge \ell$. Then $T \leqslant \ell$ is a bounded stopping time so, by the optional stopping theorem,

$$\mathbb{E}[X_\ell] \geqslant \mathbb{E}[X_T] = \mathbb{E}\big[X_T\mathbf{1}_{X_\ell^*\geqslant\lambda}\big] + \mathbb{E}\big[X_T\mathbf{1}_{X_\ell^*<\lambda}\big] \geqslant \lambda\mathbb{P}(X_\ell^* \geqslant \lambda) + \mathbb{E}\big[X_\ell\mathbf{1}_{X_\ell^*<\lambda}\big].$$

As X is non-negative it follows that we have

$$\lambda\mathbb{P}(X_\ell^* \geqslant \lambda) \leqslant \mathbb{E}\big[X_\ell\mathbf{1}_{X_\ell^*\geqslant\lambda}\big] \leqslant \mathbb{E}[X_\ell]. \tag{6.3}$$

▷

THEOREM **70** (Doob's $\mathbb{L}^p$-inequality)**.** *Let X be a martingale or non-negative submartingale. Then, for all $p > 1$ and $q = p/(p-1)$,*

$$\|X^*\|_p \leq q\sup_{n\geqslant 0}\|X_n\|_p. \tag{6.4}$$

PROOF – As above it suffices to consider the case of a non-negative submartingale indexed by the finite set $\{1,\dots,\ell\}$. We adopt the same notations. Fix $C < \infty$. By Fubini's theorem, equation (6.3) and Hölder's inequality,

$$\begin{aligned}\mathbb{E}[(X_\ell^*\wedge C)^p] &= \mathbb{E}\int_0^C p\lambda^{p-1}\mathbf{1}_{X_\ell^*\geqslant\lambda}\,d\lambda = \int_0^C p\lambda^{p-1}\mathbb{P}(X_\ell^*\geqslant\lambda)\,d\lambda \\ &\leqslant \int_0^C p\lambda^{p-2}\mathbb{E}\big[X_\ell\mathbf{1}_{X_\ell^*\geqslant\lambda}\big]\,d\lambda = q\,\mathbb{E}\big[X_\ell(X_\ell^*\wedge C)^{p-1}\big] \leqslant q\,\|X_\ell\|_p\,\|X_\ell^*\wedge C\|_p^{p-1}.\end{aligned}$$

Hence $\|X_\ell^*\wedge C\|_p \leqslant q\,\|X_\ell\|_p$ and the result follows by monotone convergence on letting $\ell \to \infty$. ▷

THEOREM **71** ($\mathbb{L}^p$-martingale convergence theorem for $p > 1$)**.**
1. *Let M be a martingale bounded in $\mathbb{L}^p$. Then M_t converges almost-surely and in $\mathbb{L}^p$ to some random variable $M_\infty \in \mathbb{L}^p$. Moreover, $M_n = \mathbb{E}[M_\infty|\mathcal{F}_n]$ a.s. for all n.*
2. *Suppose $Y \in \mathbb{L}^p(\mathcal{F}_\infty)$ and set $M_n = \mathbb{E}[Y|\mathcal{F}_n]$. Then $M = (M_n)_{n\geqslant 0}$ is a martingale bounded in $\mathbb{L}^p$ which converges almost-surely and in $\mathbb{L}^p$ to Y.*

PROOF – (1) As an $\mathbb{L}^p$-bounded martingale is also bounded in $\mathbb{L}^1$ the martingale M_n converges almost-surely to some M_∞, by the almost-sure martingale convergence theorem 63. By Doob's $\mathbb{L}^p$-inequality,

$$\|M^*\|_p \leqslant q\sup_{n\geqslant 0}\|M_n\|_p < \infty.$$

Since $|M_n - M_\infty|^p \leqslant (2M^*)^p$ for all n, we can use dominated convergence to deduce that M_n converges to M_∞ in $\mathbb{L}^p$. It follows that $M_n = \mathbb{E}[M_\infty|\mathcal{F}_n]$ almost-surely.

(2) Suppose now that $Y \in \mathbb{L}^p(\mathcal{F}_\infty)$ and set $M_n = \mathbb{E}[Y|\mathcal{F}_n]$. Then M is a martingale by the tower property and

$$\|M_n\|_p = \big\|\mathbb{E}[Y|\mathcal{F}_n]\big\|_p \leqslant \|Y\|_p$$

for all n, so M is bounded in $\mathbb{L}^p$. Hence M_n converges almost-surely and in $\mathbb{L}^p$, with limit $M_\infty \in \mathbb{L}^p(\mathcal{F}_\infty)$, say, and we can show that $M_\infty = Y$ a.s., as in the proof of Lévy's upward theorem 65. ▷

It is worth noting that one does not need Doob's results to analyse $\mathbb{L}^2$-martingales and that basic tools are sufficient in that case. This fact entirely comes from the elementary identity obtained by conditioning on $\mathcal{F}_p$

$$\mathbb{E}\big[(M_q - M_p)^2\big] = \mathbb{E}\big[M_q^2\big] - \mathbb{E}\big[M_p^2\big], \quad p < q. \tag{6.5}$$

As a consequence we see that the sequence $\big(\mathbb{E}[M_n^2]\big)_{n\geqslant 0}$ increases with n.

THEOREM **72.** *Let $(M_n)_{n\geqslant 0}$ be an $\mathbb{L}^2$-martingale. The following propositions are equivalent.*

(1) *$(M_n)_{n\geqslant 0}$ is bounded in $\mathbb{L}^2$,*
(2) *$(M_n)_{n\geqslant 0}$ converges almost-surely and in $\mathbb{L}^2$ to some $M_\infty \in \mathbb{L}^2$,*
(3) *$M_n = \mathbb{E}[M_\infty \,|\, \mathcal{F}_n]$ for some $M_\infty \in \mathbb{L}^2$.*

PROOF – (2) $\Rightarrow$ (1) $\Rightarrow$ $\mathbb{L}^2$-convergence: The first implication is obvious. For the second one, note that if $\big(\mathbb{E}[M_n^2]\big)_{n\geqslant 0}$ is bounded, it converges as it is increasing; it follows from identity (6.5) that $(M_n)_{n\geqslant 0}$ is a Cauchy sequence in (the complete space) $\mathbb{L}^2$, so it converges.
$\mathbb{L}^2$-convergence $\Rightarrow$ (3): $\mathbb{L}^2$-convergence implies $\mathbb{L}^1$-convergence...
(3) $\Rightarrow$ (2): The almost-sure convergence was established above in corollary 60 or theorem 63, the $\mathbb{L}^2$-convergence is a basic result of Hilbert space theory. ▷

6.4. **Applications.**

6.4.1. *Martingale characterization of Markov chains.* Let $(S, \mathcal{S})$ be a Borel probability space (*i.e.* nothing worst than a measurable subset of $[0,1]$) and let $\{p(x,\cdot)\,;\, x \in S\}$ be a transition kernel in S: $p(x,\cdot)$ is a probability measure on $(S,\mathcal{S})$ for every $x \in S$, and for any $A \in \mathcal{S}$ the function $p(\cdot, A)$ is measurable. The quantity $p(x,A)$ represents the probability starting from x to jump into A. We have seen an explicit construction of Markov chains in proposition 13. Daniell's theorem 18 provides another construction: it constructs a probability measure $\mathbb{P}$ on $\big(S^{\mathbb{N}}, \mathcal{S}^{\otimes\mathbb{N}}\big)$ with the prescribed finite dimensional laws, under which the coordinate process is a Markov chain with the given transition kernel. This probability $\mathbb{P}$ is the distribution of the Markov chain; it can be characterized in terms of martingales. Denote by $(X_n)_{n\geqslant 0}$ the coordinate process on $S^{\mathbb{N}}$ and by $\big(\mathcal{F}_n\big)_{n\geqslant 0}$ the induced filtration.

PROPOSITION **73.** *$(X_n)_{n\geqslant 0}$ is a Markov chain with transition kernel $\{p(x,\cdot)\,;\, x \in S\}$ iff for all bounded measurable function $f : S \to \mathbb{R}$ the process*

$$M_n^f = f(X_n) - f(X_0) - \sum_{k=0}^{n-1} \int_S \big(f(y) - f(X_k)\big)p(X_k, dy)$$

is an $\big(\mathcal{F}_n\big)_{n\geqslant 0}$-martingale.

This statement should be understood in the light of the following heuristic: martingales are the "constants of motion" of the dynamics; the above collection of martingales is big enough to characterize completely the dynamics. This is in accordance with what happens in deterministic dynamical systems.

PROOF – $\Rightarrow$: Note that $\int_S \big(f(y) - f(X_k)\big)p(X_k, dy) = \mathbb{E}\big[f(X_{k+1}) - f(X_k)\big|\mathcal{F}_k\big]$ is the mean jump of f between times k and $k+1$, knowing $\mathcal{F}_k$. Simply write

$$\mathbb{E}\big[M^f_{n+1}\big|\mathcal{F}_n\big] = \mathbb{E}\big[f(X_{n+1})\big|\mathcal{F}_n\big] - f(X_0) - \sum_{k=0}^{n-1}\mathbb{E}\big[f(X_{k+1}) - f(X_k)\big|\mathcal{F}_k\big]$$

$$= \mathbb{E}\big[f(X_{n+1}) - f(X_n)\big|\mathcal{F}_n\big] + f(X_n) - f(X_0) - \sum_{k=0}^{n-1}\mathbb{E}\big[f(X_{k+1}) - f(X_k)\big|\mathcal{F}_k\big] = M^f_n.$$

$\Leftarrow$: We only need to check that we have for any $n \geqslant 0$ and any $A \in \mathcal{S}$

$$\mathbb{P}(X_{n+1} \in A|\mathcal{F}_n) = p(X_n, A).$$

This directly comes from the martingale property of M^f_n for the function $f = \mathbf{1}_A(\cdot)$. ▷

6.4.2. *Radon-Nikodym theorem.* Let $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ be two probability measures on a measurable space $(\Omega, \mathcal{F})$. Recall that $\widetilde{\mathbb{P}}$ is said to be absolutely continuous with respect to $\mathbb{P}$ if$_{\text{def}}$ $\mathbb{P}(A) = 0$ implies $\widetilde{\mathbb{P}}(A) = 0$. It is a well known fact that this condition is equivalent to the following: For any $\varepsilon > 0$ there exists $\eta > 0$ such that for all $A \in \mathcal{F}$, the condition $\mathbb{P}(A) \leqslant \eta$ implies $\widetilde{\mathbb{P}}(A) \leqslant \varepsilon$; prove it.

THEOREM **74** (Radon–Nikodym theorem). *Let $(\Omega, \mathcal{F})$ be a measurable space such that the σ-algebra $\mathcal{F}$ is generated by an increasing sequence $\big(\mathcal{F}_n\big)_{n\geqslant 0}$ of finite σ-algebras. Let $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ be two probability measures on $(\Omega, \mathcal{F})$. Then $\widetilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$ iff there exists a non-negative random variable X such that $\widetilde{\mathbb{P}}(A) = \mathbb{E}[X\mathbf{1}_A]$ for all $A \in \mathcal{F}$.*

The random variable X, which is unique $\widetilde{\mathbb{P}}$-a.s., is called (a version of) the ***Radon-Nikodym derivative*** of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$. We write $X = d\widetilde{\mathbb{P}}/d\mathbb{P}$. The theorem extends immediately to finite measures by scaling, then to σ-finite measures by breaking Ω into pieces where the measures are finite. The assumption that $\mathcal{F}$ is countably generated can also be removed but we do not give the details here.

Without loss of generality, we shall write $\mathcal{F}_n = \sigma(A^n_1, \dots, A^n_{p_n})$, for disjoint sets A^n_i of positive $\widetilde{\mathbb{P}}$-probability.

PROOF – Recall the discussion on the construction of conditional expectation in the discrete case. In the same spirit, define the non-negative random variable

$$M_n = \sum_{k=1}^{p_n}\frac{\widetilde{\mathbb{P}}(A^n_k)}{\mathbb{P}(A^n_k)}\mathbf{1}_{A^n_k};$$

it satisfies the identity $\widetilde{\mathbb{P}}(A) = \mathbb{E}[M_n\mathbf{1}_A]$ for all $A \in \mathcal{F}_n$. As $\mathcal{F}_n$ is increasing[63], it follows that the process $(M_n)_{n\geqslant 0}$ is an $\big((\mathcal{F}_n)_{n\geqslant 0}, \mathbb{P}\big)$-martingale. We are going to show that it is uniforlmy integrable with respect to $\mathbb{P}$. By the $\mathbb{L}^1$-martingale convergence theorem, there will exist a random variable $X \geqslant 0$ such that $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[M_n\mathbf{1}_A]$ for all $A \in \mathcal{F}_n$. Define

[63]So that each A^n_k can be written as a union of A^{n+1}_j.

$\mathbb{Q}(A) = \mathbb{E}[X\mathbf{1}_A]$ for $A \in \mathcal{F}$. Then $\mathbb{Q}$ is a probability measure and $\mathbb{Q} = \widetilde{\mathbb{P}}$ on $\bigcup_n \mathcal{F}_n$, which is a π-system generating $\mathcal{F}$. Hence $\mathbb{Q} = \widetilde{\mathbb{P}}$ on $\mathcal{F}$.

To prove the uniform integrability of $(M_n)_{n\geqslant 0}$ with respect to $\mathbb{P}$ we use the above characterization of absolute continuity: as $\mathbb{P}(M_n > m) \leqslant \frac{\mathbb{E}[M_n]}{m} = \frac{1}{m} \leqslant \eta$ for m large enough, we have for such m's $\widetilde{\mathbb{P}}(M_n > m) \leqslant \varepsilon$, independently of n. So $\mathbb{E}[M_n \mathbf{1}_{M_n > m}] = \widetilde{\mathbb{P}}(M_n > m) \leqslant \varepsilon$ for all n. $\triangleright$

6.4.3. *Cameron-Martin theorem.* You are asked to prove the following statement in exercise.

Proposition **75.** *Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geqslant 0}, \mathbb{P})$ be a filtered probability space on which a non-negative martingale $(M_n)_{n\geqslant 0}$ is defined. Suppose a probability $\widetilde{\mathbb{P}}$ is defined on $(\Omega, \mathcal{F})$ such that $\widetilde{\mathbb{P}}(A) = \mathbb{E}[M_n \mathbf{1}_A]$ for all $A \in \mathcal{F}_n$ and all $n \geqslant 0$. Then $\widetilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$ iff the martingale M is uniformly integrable.*

As an application we are going to prove a result due to Cameron and Martin whose importance for stochastic analysis is difficult to overstate. We shall denote by γ the Gaussian measure on $\mathbb{R}$ (with density $(2\pi)^{-\frac{1}{2}} \exp\big(-\frac{x^2}{2}\big)$ with respect to Lebesgue measure) and by $\gamma^{\otimes \mathbb{N}}$ the product measure on $\mathbb{R}^{\mathbb{N}}$. Given $h \in \mathbb{R}^{\mathbb{N}}$ denote by τ_h the translation on $\mathbb{R}^{\mathbb{N}}$: $(x_k)_{k\geqslant 0} \to (x_k + h_k)_{k\geqslant 0}$, and by $\tau_h^* \gamma^{\otimes \mathbb{N}}$ the image measure of $\gamma^{\otimes \mathbb{N}}$ by τ_h; it is another measure on $\mathbb{R}^{\mathbb{N}}$. Last recall that two ***measures*** $\mathbb{P}$ and $\mathbb{Q}$ on a measurable space $(\Omega, \mathcal{F})$ are said to be ***equivalent*** if$_{\text{def}}$ they are absolutely continuous with respect to each other.

Theorem **76** (Cameron-Martin)**.** *The measures $\gamma^{\otimes \mathbb{N}}$ and $\tau_h^* \gamma^{\otimes \mathbb{N}}$ are equivalent iff $h \in \ell^2(\mathbb{N}) : \sum_{n\geqslant 0} h_k^2 < \infty$.*

Denote by $X_n : (x_k)_{k\geqslant 0} \to x_n$ the n^{th} coordinate map and write $\mathcal{F}_n$ for $\sigma(X_p\,;\, p \leqslant n)$; denote by X the identity map from $\mathbb{R}^{\mathbb{N}}$ to itself.

We shall denote by $\mathbb{E}$ the expectation operator with respect to $\gamma^{\otimes \mathbb{N}}$ and by $\widetilde{\mathbb{E}}$ the expectation operator with respect to $\tau_h^* \gamma^{\otimes \mathbb{N}}$, meaning nothing else than $\widetilde{\mathbb{E}}\big[f(X)\big] = \mathbb{E}\big[f(X+h)\big]$.

The proof will rely on the elementary identity below. Set

$$M_n = \exp\Big(\sum_{k=0}^n h_k X_k - \frac{1}{2} \sum_{k=0}^n h_k^2\Big).$$

The random process $(M_n)_{n\geqslant 0}$ is a martingale as all the X_k are Gaussian independent random variables. Given a bounded function $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ depending only on the first n coordinates, an elementary change of variable leads to the equality

$$\begin{aligned}
\widetilde{\mathbb{E}}\big[f(X)\big] = \mathbb{E}\big[f(X+h)\big] &= (2\pi)^{-\frac{n}{2}} \int f(x_1 + h_1, \dots, x_n + h_n)\, e^{-\frac{x_1^2 + \cdots + x_n^2}{2}} dx_1 \dots dx_n \\
&= (2\pi)^{-\frac{n}{2}} \int f(y_1, \dots, y_n)\, e^{\sum_{k=0}^n h_k y_k - \frac{1}{2}\sum_{k=0..n} h_k^2}\, e^{-\frac{y_1^2 + \cdots + y_n^2}{2}}\, dy_1 \dots dy_n \\
&= \mathbb{E}\big[M_n f(X)\big]
\end{aligned}$$

Proof – By proposition 75, the probabilities $\gamma^{\otimes \mathbb{N}}$ and $\tau_h^* \gamma^{\otimes \mathbb{N}}$ are equivalent iff the martingale $(M_n)_{n\geqslant 0}$ is uniformly integrable.

$\Leftarrow$: Let $p>0$. Supposing $h\in\ell^2(\mathbb{N})$ and replacing h by ph, we see immediately that

$$\mathbb{E}\Big[\exp\Big(p\sum_{k=0}^n h_kX_k-\frac{p^2}{2}\sum_{k=0}^n h_k^2\Big)\Big]=1,$$

so $\mathbb{E}[M_n^p]\leqslant\exp\big(\frac{p^2-p}{2}\sum_{k=0}^n h_k^2\big)\leqslant\exp\big(\frac{p^2}{2}\|h\|_2^2\big)$ and the martingale is bounded in $\mathbb{L}^p$ for any $p>1$; the result then follow from the $\mathbb{L}^p$-convergence theorem 71.
$\Rightarrow$: Suppose $(M_n)_{n\geqslant 0}$ uniformly integrable, then it converges almost-surely and in $\mathbb{L}^1$ to some non-negative random variable $M_\infty\in\mathbb{L}^1$, with $\mathbb{E}[M_\infty]=1$. Would we have $\|h\|_2=\infty$, then we would have $\mathbb{E}[M_n^p]\leqslant\exp\big(\frac{p^2-p}{2}\sum_{k=0}^n h_k^2\big)\underset{n+\infty}{\to}0$, for any $p\in(0,1)$; Fatou's lemma would imply $\mathbb{E}[M_\infty^p]=0$, a contradiction. $\triangleright$

Girsanov's theorem which you will encounter in any stochastic calculus course is nothing else than a variation of this theorem, despite its elaborated appearance. In this continuous time setting, the counterpart of theorem 76 will read as follows. Denote by $\mathbb{P}$ Wiener measure on $\mathcal{C}\big([0,1]\big)$ and by $(X_t)_{t\in[0,1]}$ the coordinate process.

Theorem **77.** *Let $H:[0,1]\to\mathbb{R}$ be a continuous function. The law under $\mathbb{P}$ of the "drifted" process $(X_t+H_t)_{t\in[0,1]}$ is abolutely continuous with respect to $\mathbb{P}$ iff there exists a function $h\in\mathbb{L}^2\big([0,1]\big)$ such that $H_t=H_0+\int_0^t h_s\,ds$, for all $t\in[0,1]$.*

The computation of the Radon-Nikodym derivative of the law of the drifted process with respect to Wiener measure involves a stochastic integral and is analogous to $\exp\Big(\sum_{k=0}^\infty h_kX_k-\frac{1}{2}\sum_{k=0}^\infty h_k^2\Big)$.

6.4.4. *A glimpse at the concentration of measure phenomenon.* Concentration of measure is the following phenomenon. Given a (Borel) probability $\mathbb{P}$ on a metric space, any Lipschitz function[64] X is close to its mean m on a set of (surprisingly) big probability: $\mathbb{P}(|X-m|\geqslant r)\leqslant\exp(-cr)$ or $\exp(-cr^2)$ for some positive constant c. This kind of inequality have a wide range of applications ranging from combinatorics, statistical physics to functional analysis and probability in Banach spaces. Although big progresses have been made recently, numerous open questions remain in this extremely lively area of research[65]. A breadth of different views and tools can lead to concentration results; in this section we give an example of how martingales can sometimes lead to them.

Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $\{\emptyset,\Omega\}\subset\mathcal{F}_1\subset\cdots\subset\mathcal{F}_n=\mathcal{F}$ be a filtration. Let Y be any real-valued integrable random variable and set for every $i\in[\![1,n]\!]$

$$D_i=\mathbb{E}[Y|\mathcal{F}_i]-\mathbb{E}[Y|\mathcal{F}_{i-1}]. \tag{6.6}$$

Note that $Y-\mathbb{E}[Y]=\sum_{i=1}^n D_i$.

Theorem **78.** *Suppose there exists some constants c_i such that one has almost-surely $|D_i|\leqslant c_i$ for all $i=1..n$; set $C^2=\sum_{i=1..n}c_i^2$. Then for $r\geqslant 0$*

$$\mathbb{P}\big(\big|Y-\mathbb{E}[Y]\big|\geqslant r\big)\leqslant 2e^{-\frac{r^2}{2C^2}}.$$

[64]That is, any one dimensional picture of the metric space in which geometry is not too much expanded.
[65]The heart of the phenomenon itself remains somewhat unclear.

Proof – Let $r \geqslant 0$. As $-Y$ satisfies the same hypothesis as Y it suffices to prove that $\mathbb{P}\big(Y - \mathbb{E}[Y] \geqslant r\big) \leqslant e^{-\frac{r^2}{2C^2}}$. Use for that Chebychev's exponential inequality

$$\mathbb{P}\big(Y - \mathbb{E}[Y] \geqslant r\big) \leqslant e^{-\lambda r}\mathbb{E}\big[e^{\lambda\sum_{i=1}^n D_i}\big],$$

introducing a non-negative parameter λ which will be optimize at the end. If we can bound above each $\mathbb{E}\big[e^{\lambda D_i}\big|\mathcal{F}_{i-1}\big]$ by some constant, by a repeated use of the tower property we shall be able give an upper bound for $\mathbb{E}\big[e^{\lambda\sum_{i=1}^n D_i}\big] = \mathbb{E}\Big[e^{\lambda\sum_{i=1}^{n-1} D_i}\,\mathbb{E}\big[e^{\lambda D_n}\big|\mathcal{F}_{n-1}\big]\Big]$, etc. But noting that $\lambda D_i = \frac{1+D_i/c_i}{2}\lambda c_i + \frac{1-D_i/c_i}{2}(-\lambda c_i)$ and using the convexity of the exponential map, we get for any $\lambda \geqslant 0$

$$e^{\lambda D_i} \leqslant \frac{1+D_i/c_i}{2}e^{\lambda c_i} + \frac{1-D_i/c_i}{2}e^{-\lambda c_i},$$

and so, as $\mathbb{E}[D_i|\mathcal{F}_{i-1}] = 0$,

$$\mathbb{E}\big[e^{\lambda D_i}\big|\mathcal{F}_{i-1}\big] \leqslant \cosh(\lambda c_i) \leqslant e^{\frac{\lambda^2 c_i^2}{2}}.$$

This leads to the estimate $\mathbb{E}\big[e^{\lambda\sum_{i=1}^n D_i}\big] \leqslant e^{-\lambda r + \frac{\lambda^2}{2}C^2}$; it remains to optimize over $\lambda \geqslant 0$ to get the result. ▷

In the following corollary the discrete space $\{0,1\}^n$ is endowed with the ℓ^1 metric $d(x,y) = \sum_{i=1}^n |x_i - y_i| = \#\{i \in [\![1,n]\!]\,;\, x_i \neq y_i\}$ and the uniform probability $\mathbb{P}$.

Corollary **79.** *Let $Y : \{0,1\}^n \to \mathbb{R}$ be a contraction: $|Y(x) - Y(y)| \leqslant d(x,y)$ for all $x, y \in \{0,1\}^n$. Then for any $r \geqslant 0$*

$$\mathbb{P}(Y \geqslant \mathbb{E}[Y] + r) \leqslant 2e^{-\frac{2r^2}{n}}.$$

Proof – Using the coordinate maps $X_k : (x_i)_{i=1..n} \to x_k$ on $\{0,1\}^n$ we define the filtration $(\mathcal{F}_k)_{k=1..n}$ setting $\mathcal{F}_0 = \big\{\emptyset, \{0,1\}^n\big\}$ and $\mathcal{F}_k = \sigma(X_p\,;\, 1 \leqslant p \leqslant k)$ for $1 \leqslant k \leqslant n$. Set for convenience $Y_k = \mathbb{E}[Y|\mathcal{F}_k]$ and define the martingale difference $D_k = Y_k - Y_{k-1}$ as above. Let us estimate $D_1 = Y_1 - \mathbb{E}[Y]$. Observe that Y_1 takes only two values: the average of Y on the faces $\{\pi_1 = 0\}$ and $\{\pi_1 = 1\}$. These averages cannot differ too much as we go from a point of one face to a point of the other changing only one coordinate, so that Y cannot change by more than 1 since it is a contraction. So we see that $|D_1| \leqslant \frac{1}{2}$. The same argument holds for estimating the other D_k; apply theorem 78 to conclude. ▷

The preceding proof also justifies the following statement. Let $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)_{i=1..n}$ be probability spaces and define $(\Omega, \mathcal{F}, \mathbb{P})$ as the product probability space.

Corollary **80.** *Let Y be an integrable function on $(\Omega, \mathcal{F})$ such that there exist constants c_i with $|Y(x) - Y(y)| \leqslant c_i$ if x and y differ only by their i^{th} coordinate. Then for any $r \geqslant 0$*

$$\mathbb{P}\Big(\big|Y - \mathbb{E}[Y]\big| \geqslant r\Big) \leqslant 2\exp\Big(-\frac{r^2}{2\sum_{i=1}^n c_i^2}\Big).$$

7. Continuous time martingale theory

Although section 6 was written in the setting of a filtered probability space with a filtration $\left(\mathcal{F}_n\right)_{n\geqslant 0}$, all the definitions given above are meaningful for a filtration indexed by any other countable subset I of $\mathbb{R}_+$, with ∞ to be understood as $\sup I$. Write $\mathbb{Q}_+$ for $\mathbb{Q}\cap\mathbb{R}_+$. Our basic setting to construct continuous time martingales will be a probability filtered space $\left(\Omega,\mathcal{F},\left(\mathcal{F}_t\right)_{t\in\mathbb{Q}_+},\mathbb{P}\right)$ on which a martingale $(M_t)_{t\in\mathbb{Q}_+}$ is defined; this is the skeleton of the coming extension. It is a remarkable result due to Doob that nothing else than the martingale property is needed to extend $(M_t)_{t\in\mathbb{Q}_+}$ to $\mathbb{R}_+$. Demanding continuity for $(M_t)_{t\geqslant 0}$ will be too much, we shall get càdlàg paths; càdlàg = continue à droite, limite à gauche = continuous on the right with left limit. We shall suppose that $\mathcal{F}_0$ *contains all the* $\mathbb{P}$*-null sets.*

Theorem **81** (Regularization of martingales. Doob)**.** *Let* $\left(\Omega,\mathcal{F},\left(\mathcal{F}_t\right)_{t\in\mathbb{Q}_+},\mathbb{P}\right)$ *be a probability filtered space and let* $(M_t)_{t\in\mathbb{Q}_+}$ *be an* $\left(\mathcal{F}_t\right)_{t\in\mathbb{Q}_+}$*-martingale. For* $t\in\mathbb{R}_+$ *set* $\mathcal{F}_{t^+}:=\bigcap_{s>t,\,s\in\mathbb{Q}_+}\mathcal{F}_s$*. Then one can construct on* $(\Omega,\mathcal{F},\mathbb{P})$ *an* $(\mathcal{F}_{t^+})_{t\geqslant 0}$*-martingale* $(\widetilde{M}_t)_{t\geqslant 0}$ *with càdlàg paths such that one has* $\mathbb{P}$*-almost-surely for all* $t\in\mathbb{Q}_+$

$$\mathbb{E}[\widetilde{M}_t|\mathcal{F}_t]=M_t.$$

Proof – Given real numbers $a<b$ denote by $U_{a,b}\big([0,N]\big)$ the number of upcrossings of $(M_t)_{t\in\mathbb{Q}_+}$ from a to b in $[0,N]$, and set $M_N^*=\sup_{t\in\mathbb{Q}_+\cap[0,N]}|M_t|$. By Doob's upcrossing inequality the $U_{a,b}\big([0,N]\big)$ are almost-surely finite[66] for all rational $a<b$ and $N\geqslant 0$; also, all the M_N^* are finite[67], for all $N\geqslant 0$, by Doob's maximal inequality. Denote by Ω_0 this event of probability 1 where all these quantities are finite; the following limits exists on Ω_0

$$\begin{aligned} M_{t^+} &= \lim_{s\downarrow t,\,s\in\mathbb{Q}_+} M_s, \quad t\geqslant 0\\ M_{t^-} &= \lim_{s\uparrow t,\,s\in\mathbb{Q}_+} M_s, \quad t>0. \end{aligned}$$

Define, for $t\geqslant 0$,

$$\widetilde{M}_t=\begin{cases} M_{t^+}, & \text{on } \Omega_0,\\ 0, & \text{otherwise.}\end{cases}$$

Then $\widetilde{M}$ is càdlàg and $\left(\mathcal{F}_{t^+}\right)_{t\geqslant 0}$-adapted. To prove that it is an $\left(\mathcal{F}_{t^+}\right)_{t\geqslant 0}$-martingale, given $s<t$ choose rationals $s_n<t_n$ decreasing to s and t respectively. By the convergence theorem for backward martingales, M_{s_n} (resp. M_{t_n}) converges almost-surely and in $\mathbb{L}^1$ to M_{s^+} (resp. M_{t^+}), so we have for any event $A\in\mathcal{F}_{s^+}$

$$\mathbb{E}[M_{s^+}\mathbf{1}_A]=\lim\mathbb{E}[M_{s_n}\mathbf{1}_A]=\lim\mathbb{E}[M_{t_n}\mathbf{1}_A]=\mathbb{E}[M_{t^+}\mathbf{1}_A],$$

i.e. $\mathbb{E}[M_{t^+}|\mathcal{F}_{s^+}]=M_{s^+}$, or $\mathbb{E}[\widetilde{M}_t|\mathcal{F}_{s^+}]=\widetilde{M}_s$. The projection property $\mathbb{E}[\widetilde{M}_t|\mathcal{F}_t]=M_t$, for $t\in\mathbb{Q}_+$, is also a direct consequence of the convergence theorem for backward martingales. $\triangleright$

Definition **82.** *Filtrations which are continuous on the right ($\mathcal{F}_{t^+}=\mathcal{F}_t$) and for which $\mathcal{F}_0$ contains the $\mathbb{P}$-null sets are said to satisfy the* ***usual conditions****.*

---

[66]Work first with a finite index set D, for which Doob's upcrossing lemma says us that $(M_t)_{t\in D}$ will almost-surely have only a finite number of upcrossings between any two rational times; let then increase D to $\mathbb{Q}_+$: a countable intersection of events of probability 1 being of probability 1 the result follows.

[67]Play the same game here as above.

Doob's regularization theorem shows that we do not lose much in restricting our attention to càdlàg martingales and filtrations satisfying the usual conditions. A comment is needed here, however: *all martingales are not continuous on the right, and quite venerable filtrations do not satisfy the usual conditions(!).* Let for example Y be a binomial random variable with parameter $\frac{1}{2}$. Define

$$X_t = \begin{cases} \frac{1}{2} & \text{for } 0 \leqslant t \leqslant \frac{1}{2}, \\ Y & \text{for } \frac{1}{2} < t \leqslant 1. \end{cases}$$

Set also

$$\mathcal{F}_t = \begin{cases} \{\emptyset, \Omega\} & \text{for } 0 \leqslant t \leqslant \frac{1}{2}, \\ \sigma(Y) & \text{for } \frac{1}{2} < t \leqslant 1. \end{cases}$$

Then X is an $(\mathcal{F}_t)_{t\in[0,1]}$-martingale who is not continuous on the right at time $\frac{1}{2}$. The filtration $(\mathcal{F}_t)_{t\in[0,1]}$ also fails to be continuous on the right at time $\frac{1}{2}$.

Doob's regularization procedure transforms in a non-trivial way a process: working on the canonical space of continuous functions from $\mathbb{R}_+$ to $\mathbb{R}$, with the coordinate process $(X_t)_{t\geqslant 0}$ and its filtration $(\mathcal{F}_t)_{t\geqslant 0}$, under Wiener measure, the process $M_t = \mathbf{1}_{X_t=1}$ is an $(\mathcal{F}_t)_{t\geqslant 0}$-martingale which is almost-surely equal to 0 for each fixed $t \geqslant 0$ and does not converge to 0 as time goes to infinity. The càdlàg regularization procedure "smoothes" these irregularities and gives as a regularized process the constant 0.

The càdlàg property of regularized martingales implies that their pathwise properties are entirely determined by their $\mathbb{Q}_+$-skeleton. It follows that ***all theorems of section 6 (Doob's inequalities, convergence, optional stopping theorems...) hold for càdlàg martingales, for a filtration satisfying the usual conditions***. As an example we give the details of the proof of the optional stopping theorem.

THEOREM **83** (Optional stopping theorem). *Let us work on a filtered probability space with a filtration satisfying the usual conditions, and let M be a càdlàg adapted process. Then the following are equivalent:*

1. *M is a martingale,*
2. *for all bounded stopping times T and all stopping times S, M_T is integrable and*

$$\mathbb{E}[M_T|\mathcal{F}_S] = M_{S\wedge T}, \tag{7.1}$$

3. *for all stopping times T, the stopped process M^T is a martingale,*
4. *for all bounded stopping times T, the random variable M_T is integrable and*

$$\mathbb{E}[M_T] = \mathbb{E}[M_0].$$

Moreover, if M is uniformly integrable, then 2 and 4 hold for all *stopping times T.*

PROOF – Suppose M is a martingale. Let S and T be stopping times, with T bounded, $T \leqslant t$ say. For $n \geqslant 0$, set

$$S_n = 2^{-n}\lceil 2^n S\rceil, \quad T_n = 2^{-n}\lceil 2^n T\rceil.$$

The random times S_n and T_n are stopping times decreasing to S and T respectively. Since $(M_t)_{t\geqslant 0}$ is right continuous, M_{T_n} converges almost-surely to M_T. By the discrete-time optional stopping theorem, $M_{T_n} = \mathbb{E}[M_{t+1}|\mathcal{F}_{T_n}]$ so $(M_{T_n})_{n\geqslant 0}$ is uniformly integrable and so M_{T_n} converges to M_T in $\mathbb{L}^1$; in particular, M_T is integrable. Similarly, $M_{S_n\wedge T_n}$ converges almost-surely and in $\mathbb{L}^1$ to $M_{S\wedge T}$. So, by the discrete-time optional stopping theorem again, we have for any $A \in \mathcal{F}_S \subset \mathcal{F}_{S_n}$

$$\mathbb{E}[M_{T_n}\mathbf{1}_A] = \mathbb{E}[M_{S_n\wedge T_n}\mathbf{1}_A].$$

On letting $n \to \infty$, we deduce that identity (7.1) holds. For the rest of the proof we argue as in the discrete-time case. $\triangleright$

8. Comments and exercises

References. Williams' book [Wil91] is certainly a good reference for this section part of the course; so is Rogers and Williams' book [RW00]. The book [BMP02] on martingales and Markov chains is an excellent source of worked out examples, under the form of solved exercises; spending some time with it will undoubtedly bring you some acquaintance with the subject.

Filtrations indexed by $\mathbb{R}_+$ are subtle and sometime mysterious objects. The following comments are here to guide your first steps in this suject.

1. Filtrations generated by a process. Let X be a process defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{F}^X$ the filtration it generates. The following two exmples clarify the relationship between the regularity of X and the regularity of $\mathcal{F}^X$.

a) *A continuous process can generate a discontinuous filtration.* Suppose Ω has at least two points, so we can define on it a non-constant real-valued random variable ξ. Set $X_t(\omega) = t\xi(\omega)$. One easily checks that $\mathcal{F}_0^X$ is trivial while $\mathcal{F}_t^X = \sigma(\xi)$ is non-trivial for $t > 0$; so $\mathcal{F}^X$ is not right continuous despite the continuity of X.

b) *A discontinuous process can generate a continuous filtration.* Given any discontinuous function $h : \mathbb{R}_+ \to \mathbb{R}$ set $X_t(\omega) = h(t)$ for all $\omega \in \Omega$ and $t \geqslant 0$. The σ-algebra $\mathcal{F}_t^X$ is then trivial for all $t \geqslant 0$, so the filtration is continuous.

2. Usual assumptions.

8.1. **Exercises. 1. a)** Suppose (U, V) is an $\mathbb{R}^2$-valued random variable with a density function $f_{U,V}(u, v)$ with respect to Lebesgue measure on $\mathbb{R}^2$. Then (why?) U has a density function f_U with respect to Lebesgue measure on $\mathbb{R}$, given by

$$f_U(u) = \int_{\mathbb{R}} f_{U,V}(u, v)\, dv.$$

The conditional density function of V given U is defined by the formula

$$f_{V|U}(v|u) = \frac{f_{U,V}(u, v)}{f_U(u)}$$

where $0/0 = 0$, by convention. Given a bounded measurable function $h : \mathbb{R} \to \mathbb{R}$, set

$$g(u) = \int_{\mathbb{R}} h(v) f_{V|U}(v|u)\, dv.$$

Prove that $g(U) = \mathbb{E}\big[h(V)|\sigma(U)\big]$.

b) Let (U, V) be an $\mathbb{R}^2$-valued Gaussian random variable with null mean and covariance matrix Λ. Find $\mathbb{E}[V|\sigma(U)]$.

c) More generally, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and U, V be two integrable real-valued random variables defined on it. Prove that there exists a measurable function $g : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}\big[V \,|\, \sigma(U)\big] = g(U)$.

2. Let $(\Omega, \mathcal{F}, \big(\mathcal{F}_n\big)_{n \geqslant 0}, \mathbb{P})$ be a filtered probability space. Suppose T is a stopping time such that for some $N \in \mathbb{N}$ and some $\epsilon > 0$ we have almost-surely for every $n \geqslant 0$

$$\mathbb{P}(T \leqslant n + N | \mathcal{F}_n) \geqslant \epsilon.$$

Prove by induction that

$$\mathbb{P}(T > kN) \leqslant (1 - \epsilon)^k, \ k \in \mathbb{N},$$

and deduce that $\mathbb{E}[T]$ is finite.

3. a) Find an example of a measurable space $(\Omega, \mathcal{F})$, with two filtrations $(\mathcal{F}_t)_{t \geqslant 0}$ and $(\mathcal{G}_t)_{t \geqslant 0}$ on it, on which there exists a random time T which is an $(\mathcal{F}_t)_{t \geqslant 0}$-stopping time but not a $(\mathcal{G}_t)_{t \geqslant 0}$-stopping time. Note that we do not need a probability measure to talk about stopping times.

b) Let us work on the canonical space $\mathcal{C}\big([0,1], \mathbb{R}\big)$, with its coordinate process and the induced filtration $(\mathcal{F}_t)_{t \in [0,1]}$. Given any $a \in [0,1]$ and $\omega \in \mathcal{C}\big([0,1], \mathbb{R}\big)$, define the last zero of ω before time a as

$$\gamma_a(\omega) = \max\{s \in [0, a]\, ;\, \omega_s = 0\}.$$

Show that the random variable γ_a is $\mathcal{F}_a$-measurable, while the event $\{\gamma_a < t\}$ is not in $\mathcal{F}_t$ for any $t < a$.

4. Let $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geqslant 0}\big)$ be a filtered space, and T be a stopping time. **a)** Recall the definition of the σ-algebra $\mathcal{F}_T$: it consists of those events whose occurence or no-occurence can be decided from what we know up and including time T. It seems tempting then to re-define it as $\bigcap_{s \leqslant T} \mathcal{F}_s$. What goes wrong with that "definition"?

b) Let S be another stopping time. Prove that $\mathcal{F}_{S \wedge T} = \sigma(\mathcal{F}_S, \mathcal{F}_T)$.

5. Prove that a sequence of integrable random variables X_n converges in $\mathbb{L}^1$ to some $X \in \mathbb{L}^1$ iff X_n converges in probability to X and the family $(X_n)_{n \geqslant 0}$ is uniformly integrable.

6. Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on a measurable space $(\Omega, \mathcal{F})$. Prove that $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$ iff for any $\epsilon > 0$ there exists an $\eta > 0$ such that for all $A \in \mathcal{F}$, $\mathbb{Q}(A) \leqslant \eta$ implies $\mathbb{P}(A) \leqslant \epsilon$.

7.. Let $(\Omega, \mathcal{F}, \big(\mathcal{F}_n\big)_{n \geqslant 0}, \mathbb{P})$ be a filtered probability space on which a non-negative martingale $(M_n)_{n \geqslant 0}$ is defined. Suppose a probability $\mathbb{Q}$ is defined on $(\Omega, \mathcal{F})$ such that $\mathbb{Q}(A) = \mathbb{E}[M_n \mathbf{1}_A]$ for all $A \in \mathcal{F}_n$ and all $n \geqslant 0$. Prove that $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$ iff the martingale M is uniformly integrable.

8. Kolmogoro'v $0 - 1$ law. *The following result was used in the proof of the strong law of large numbers given in the notes and does not use martingale theory.* Let $(\mathcal{F}_n)_{n \geqslant 0}$ be the filtration generated by some process $(X_n)_{n \geqslant 0}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define the tail σ-algebra of the process as the sub-σ-algebra $\mathcal{T} = \bigcap_{n \geqslant 0} \sigma\big(X_k\, ;\, k \geqslant n\big)$ of $\mathcal{F}$ on Ω. Suppose all the X_n are independent under $\mathbb{P}$. Prove that any event of $\mathcal{T}$ has $\mathbb{P}$-probability 0 or 1. Does the result remains true if we do not suppose the X_n are independent?

9. *We give here the details of the proof of Lévy's upward convergence theorem, 65, seen as a corollary of the $\mathbb{L}^1$-convergence theorem, 64.*

a) Prove that any $\mathcal{F}_\infty$-measurable *bounded* random variable is an $\mathbb{L}^1$-limit of elements of $\mathbb{L}^1\big(\mathcal{F}_n\big)$.

b) Using then an approximation argument, prove that $\bigcup_{n \geqslant 0} \mathbb{L}^1\big(\mathcal{F}_n\big)$ is dense in $\mathbb{L}^1\big(\mathcal{F}_\infty\big)$.

c) Give a neat proof of Lévy's upward convergence theorem and 0-1 law.

d) Remark that this theorem provides a *constructive* approach of $\mathbb{E}[X|\mathcal{F}_\infty]$ when the filtration $\left(\mathcal{F}_n\right)_{n\geqslant 0}$ is made up of finite σ-algebras (why?). We say in this case that the *measurable space* $(\Omega, \mathcal{F}_\infty)$ is *separable*.

(i) Prove that a measurable space $(\Omega, \mathcal{F}_\infty)$ is separable iff $\mathbb{L}^1(\mathcal{F}_\infty)$ is separable (*i.e.* has a dense sequence).

(ii) Prove that Borel spaces are always separable.

10. Simple random walks. Let $(X_n)_{n\geqslant 0}$ be a sequence of iid random variables with law $p\delta_1 + q\delta_{-1}$, and $S_n = X_1 + \cdots + X_n$. Denote by $\mathcal{F}_n$ the filtration generated by $X_1, \dots, X_n$. All notions are relative to this filtration. Given $a < 0, b > 0$ and $x \in \mathbb{R}$, define the stopping times $T_{ab} = \inf\{n \geqslant 1\,;\, S_n = a \text{ or } b\}$ and $T_x = \inf\{n \geqslant 1\,;\, S_n = x\}$.

a) *Case $p > q$.* Prove first that the random times T_b and T_{ab} are almost-surely finite.

(i) Prove that the process $\left(\frac{q}{p}\right)^n$ is a martingale, and deduce the law of $S_{T_{ab}}$.

(ii) Using the martingale $S_n - n(p-q)$ (prove it), find $\mathbb{E}[T_{ab}], \mathbb{E}[T_b]$ and $\mathbb{E}[T_a]$.

b) *Case $p = q = \frac{1}{2}$: Symmetric random walk.* Prove first that T_{ab} is almost-surely finite, and using the martingale S_n find the law of $S_{T_{ab}}$. Using then the martingale $S_n^2 - n$ (prove it), find $\mathbb{E}[T_{ab}], \mathbb{E}[T_b]$ and $\mathbb{E}[T_a]$.

c) *Case $p \geqslant q$* For $\lambda \in \mathbb{R}$ set $\phi(\lambda) = pe^{\lambda} + qe^{-\lambda}$. Prove that $Y_n := e^{\lambda S_n}\phi(\lambda)^{-n}$ is a martingale. Deduce from that the generating function of T_b and find back $\mathbb{E}[T_b]$.

11. Branching processes. a) In this question we consider a Galton-Watson branching process in which the number of children of each individual is 0 or 2, equally probably. Denote by Z_n the size of the n^{th} generation, starting zith $Z_0 = 1$. Prove that $(Z_n)_{n\geqslant 0}$ is a martingale (with respect to its own filtration), and that it converges almost-surely to 0.

b) Consider now a general case in chiwhc the distribution of the number of children takes values in $\mathbb{N}$ and is integrable. Denote by μ its mean. We chall write $Z_{n+1} = X_1^{(n+1)} + \cdots + X_{Z_n}^{(n+1)}$, where the $X_i^{(n+1)}$ are iid, conditionally on Z_n.

(i) Prove that $M_n = \frac{Z_n}{\mu^n}$ is a martingale.

(ii) Prove that $(Z_n)_{n\geqslant 0}$ converges almost-surely to 0 if $\mu \leqslant 1$.

(iii) Suppose $\mu > 1$. Prove that $(M_n)_{n\geqslant 0}$ converges almost-surely to a finite limit M_∞. Setting $p = \mathbb{P}(M_\infty = 0)$, prove that p^{Z_n} converges almost-surely to $\mathbf{1}_{M_\infty = 0}$ and describe the behaviour of $(Z_n)_{n\geqslant 0}$ in terms of M_∞.

c) Suppose in this question that the offspring distribution is not only integrable but also has a finite variance σ^2. Prove that $(M_n)_{n\geqslant 0}$ is a martingale bounded in $\mathbb{L}^2$ and that it cannot converge to 0 almost-surely. What is the variance of M_∞?

12. Find an example of a martingale which converges almost-surely but not in $\mathbb{L}^1$.

13. Let M be a martingale bounded in $\mathbb{L}^1$ and T be a stopping time. Prove that M_T is in $\mathbb{L}^1$. Give an example where $\mathbb{E}[M_T] \neq \mathbb{E}[M_0]$.

14. Let $f : [0,1] \to \mathbb{R}$ be a Lipschitz function, and denote by f_n the simplest piecewise linear function agreeing with f on $\mathbb{D}_n = \left\{k2^{-n}\,;\, k = 0..2^n\right\}$. Set $M_n = f_n'$ outside $\mathbb{D}_n$. Introducing a proper filtered probability space, prove that M_n converges Lebesgue-almost-surely and in $\mathbb{L}^1(\textsc{Leb})$ to some bounded f_∞' which satisfies $f(t) = \int_0^t f_\infty'(s)\,ds$ for any $t \in [0,1]$.

15. Recall the construction of the isonormal Gaussian process X indexed by a separable Hilbert H given in exercise 4 in example sheet 1. Take $h \in H$. Prove that the series defining X_h converges almost-surely and in $\mathbb{L}^2(\mathbb{P})$.

16. Let a be a real constant. Let $\mathbb{P}$ denote Wiener measure on $\mathcal{C}\big([0,1], \mathbb{R}\big)$, X the coodinate process and $\mathbb{P}^1$ the law of the process $\{X_t + at\}_{t\in[0,1]}$. Prove that $\mathbb{P}^1$ is absolutely continuous with respect to $\mathbb{P}$ and find $\frac{d\mathbb{P}^1}{d\mathbb{P}}$.

17. Equip the symmetric group $\mathfrak{S}_n$ with the Hamming distance: $d(\sigma,\tau) = \#\{i \in [\![1,n]\!]\,;\,\sigma_i \neq \tau_i\}$ and the uniform probability. Prove that for any function $f : \mathfrak{S}_n \to \mathbb{R}$, and any $r \geqslant 0$, we have $\mathbb{P}\big(\big|f(\sigma) - \mathbb{E}[X]\big| > r\big) \leqslant 2e^{-\frac{r^2}{2n}}$.

18. Let $(\mathcal{F}_t)_{t\geqslant 0}$ be the filtration generated by the coordinate process on $\mathcal{C}\big([0,1],\mathbb{R}\big)$. Prove that this filtration is not continuous on the right.

19. a) Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, and let $\mathcal{G}_1,\mathcal{G}_2,\dots$ and $\mathcal{G}$ be some sub-σ-algebras independent under $\mathbb{P}$. Prove that any event of $\bigcap_{n\geqslant 1}\sigma(\mathcal{G}_n,\mathcal{G}_{n+1},\dots;\mathcal{G})$ coincides almost-surely with an event of $\mathcal{G}$.

b) Denote now by $\mathcal{F}_t$ the filtration generated by a Brownian motion defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$. As we have seen in the preceding exercise, the $\mathbb{P}$-null sets may cause unexpected and unpleasant things. Denote by $\mathcal{N}$ the σ-algebra of $\mathbb{P}$-null sets, and replace each $\mathcal{F}_t$ by $\mathcal{G}_t := \sigma(\mathcal{N},\mathcal{F}_t)$. Set as usual $\mathcal{G}_{t+} = \bigcap_{s>t}\mathcal{G}_s$, for all $t \geqslant 0$. Using ***a)*** and the independence of the increments of Brownian motion, prove that the σ-algebras $\mathcal{G}_t$ and $\mathcal{G}_{t+}$ coincide up to $\mathbb{P}$-null sets.

9. Complements to Part II

We show in the first complement how ideas from martingale theory can be used to *give some meaning and solve stochastic differential equations*, without using the machinery of stochastic integrals.

The second complement is dedicated to elucidate the question: *Is the conditional expectation operator an integral with respect to a random measure?*

9.1. ***Complement: Solving stochastic differential equations.***

Let $(B_t)_{0\leqslant t\leqslant 1}$ be a Brownian motion defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$. Given two functions $b,\sigma \neq 0$, and a starting point $x_0 \in \mathbb{R}$, define for every integer $n \geqslant 1$ a process $\big(Y^n(t)\big)_{0\leqslant t\leqslant 1}$ setting $Y^n(0) = x_0$, and for $\frac{k-1}{n} < t \leqslant \frac{k}{n}$, $k \in \{1,\dots,n\}$,

$$Y^n_t = Y^n_{\frac{k-1}{n}} + b\big(Y^n_{\frac{k-1}{n}}\big)\left(t - \frac{k-1}{n}\right) + \sigma\big(Y^n_{\frac{k-1}{n}}\big)\left(B_t - B_{\frac{k-1}{n}}\right). \tag{9.1}$$

When $\sigma = 0$ this dynamics is nothing else than the Euler approximation of the differential equation $\dot{x}_t = b(x_t)$. A well-known corollary of Ascoli-Arzela's compactness theorem states that the Euler approximations has a converging subsequence whose limit is *a* solution of the differential equation. The following theorem says the same in our stochastic context. Recall we denote by $(W,\mathcal{W})$ the space $\mathcal{C}\big([0,1],\mathbb{R}\big)$ with its Borel σ-algebra and write X for the coordinate process. We suppose σ non-identically null.

Theorem **84.** *Suppose the functions b and σ are bounded. Then the laws $\mathbb{P}^n$ of Y^n form a tight sequence of probability measures on $(W,\mathcal{W})$.*

This statement and Prohorov's compactness theorem ensure us that the sequence of $\mathbb{P}^n$'s has at least one weak limit $\mathbb{Q}$, say. It seems reasonnable to say that ***under*** $\mathbb{Q}$ ***the coordinate process on*** $(W,\mathcal{W})$ **solves** ***the stochastic differential equation***

$$dx_t = b(x_t)\,dt + \sigma(x_s)\,dB_t, \tag{9.2}$$

where dB_s is a Brownian increment over a time interval ds, with variance equal to ds.

Let $\mathbb{E}^n$ be the expectation operator associated with the probability $\mathbb{P}^n$. The following proposition is the heart of the proof of theorem 84.

Proposition **85.** *Suppose there exists a positive constant C such that we have*

$$\mathbb{E}^n\big[|X_t - X_s|^4\big] \leqslant C|t-s|^2 \tag{9.3}$$

for all s, t in $[0,1]$ and $n \geqslant 1$. Then the sequence $\big(\mathbb{P}^n\big)_{n\geqslant 1}$ is tight.

Proof – Kolmogorov's regularity criterion states that if $\mathbb{E}^n\big[|X_t - X_s|^4\big] \leqslant C|t-s|^2$ for some constant C and all $s,t \in [0,1]$ then there exists a random variable $C(\omega)$ in $\mathbb{L}^4\big(\mathbb{P}^n\big)$ such that we have $\mathbb{P}^n$-almost-surely $|X_t - X_s| \leqslant C(\omega)|t-s|^\alpha$, for all $s,t \in [0,1]$ and any $\alpha \in \big[0, \frac14\big)$. This implies in particular that the modulus of continuity $M_X(\delta)$ of X is $\mathbb{P}^n$-almost-surely bounded above by $C(\omega)\delta^\alpha$, so we have $\mathbb{E}^n\big[M_X(\delta)^4\big] \leqslant \mathbb{E}^n[C(\omega)]\,\delta^\alpha$ for all $n \geqslant 0$. As the proof of Kolmogorov's criterion provides an upper bound for $\mathbb{E}^n\Big[\,\big|C(\omega)\big|^4\Big]$ depending only on the constant C of (9.3) we actually have $\mathbb{E}^n\big[M_X(\delta)^4\big] \leqslant C'\delta^4$ for some constant C'. This inequality implies the equi-continuity condition (2.3) of corollary 36: $\lim\limits_{\delta\downarrow 0}\overline{\lim\limits_{n}}\,\mathbb{E}^n\big[M_X(\delta)\wedge\ 1\big] = 0$. As X_0 is $\mathbb{P}^n$-almost-surely equal to 0, it follows from the Ascoli-Arzela theorem that the probabilities $\mathbb{P}^n$ have support in a compact set of W. ▷

We are now going to see that condition (9.3) can be obtained as a simple application of martingale ideas.

Lemma **86.** *There exists a positive constant C' such that we have*

$$\mathbb{E}\big[|Y^n_t - Y^n_s|^4\big] \leqslant C'\,|t-s|^2 \tag{9.4}$$

for all $s,t \in [0,1]$ and all $n \geqslant 1$.

Proof – Denote by $(\mathcal{F}_t)_{t\in[0,1]}$ the filtration on $(\Omega, \mathcal{F})$ generated by B. The process M_t containing all the explicit Brownian terms in the definition of Y^n and defined for $t \in \big(\frac{k-1}{n}, \frac{k}{n}\big]$ by

$$M_t := \sum_{j=0}^{k-2} \sigma\Big(Y^n_{\frac{j}{n}}\Big)\big(B_{\frac{j+1}{n}} - B_{\frac{j}{n}}\big) + \sigma\Big(Y^n_{\frac{k-1}{n}}\Big)\big(B_t - B_{\frac{k-1}{n}}\big)$$

is an $(\mathcal{F}_t)_{t\in[0,1]}$-martingale. This is easily checked by induction. Also, conditioning successively on $\mathcal{F}_{\frac{k-1}{n}}, \mathcal{F}_{\frac{k-2}{n}}, \dots$ we see that

$$\mathbb{E}\big[|M_t|^2\big] \leqslant A^2 t,$$

where A denotes an upper bound for σ. Clearly, the same proof gives $\mathbb{E}\big[|M_t - M_s|^2\big] \leqslant A^2|t-s|$. It is not harder to prove that $\mathbb{E}\Big[\,\big|Y^n_t - Y^n_s\big|^4\,\Big] \leqslant 9A^4\,|t-s|^2$; it suffices to do it for $s=0$. Write σ_j for $\sigma\Big(Y^n_{\frac{j}{n}}\Big)$. Expanding the sum defining M_t and keeping only the terms with non-vanishing expectation we get

$$\begin{aligned}\mathbb{E}\big[\,|M_t|^4\,\big] = \mathbb{E}\left[\sum_{j=0}^{k-1}\sigma_j^4\big(B_{t\wedge\frac{j+1}{n}} - B_{t\wedge\frac{j}{n}}\big)^4\right] &+ 6\,\mathbb{E}\left[\sum_{0\leqslant j<\ell\leqslant k-1}\sigma_j^2\sigma_\ell^2\big(B_{t\wedge\frac{j+1}{n}} - B_{t\wedge\frac{j}{n}}\big)^2\big(B_{t\wedge\frac{\ell+1}{n}} - B_{t\wedge\frac{\ell}{n}}\big)^2\right] \\ &+ 12\,\mathbb{E}\left[\sum_{0\leqslant j<\ell<m\leqslant k-1}\sigma_j\sigma_\ell\sigma_m^2\big(B_{t\wedge\frac{j+1}{n}} - B_{t\wedge\frac{j}{n}}\big)\big(B_{t\wedge\frac{\ell+1}{n}} - B_{t\wedge\frac{\ell}{n}}\big)\big(B_{t\wedge\frac{m+1}{n}} - B_{t\wedge\frac{m}{n}}\big)^2\right]\end{aligned}$$

The first term is bounded above by $3A^4t^2$ and the sum of the two other terms is equal to

$$6\sum_{\ell=0}^{k-1}\mathbb{E}\Big[M^2_{\frac{\ell-1}{n}}\sigma_\ell^2\big(B_{t\wedge\frac{\ell+1}{n}} - B_{t\wedge\frac{\ell}{n}}\big)^2\Big].$$

By conditioning with respect to $\mathcal{F}_{\frac{\ell}{n}}$ in each expectation we get the upper bound

$$6A^2 \sum_{\ell=0}^{k-1} \Big(t \wedge \frac{\ell+1}{n} - t \wedge \frac{\ell}{n}\Big)\, \mathbb{E}\Big[\,\big|M_{\frac{\ell-1}{n}}\big|^2\Big] \leqslant 6A^2\, t\, \mathbb{E}\Big[\,\big|M_{\frac{\ell-1}{n}}\big|^2\Big] \leqslant 6A^4\, t^2.$$

Fix $s < t$ in $[0,1]$ and let k and k' be the integer parts of $n\,t$ and $n\,s$ respectively. Using the elementary inequality $(a+b)^4 \leqslant 8\,(a^4+b^4)$ and Jensen's inequality we get

$$\mathbb{E}\big[|Y^n_t - Y^n_s|^4\big] \leqslant 8\,\mathbb{E}\Big[\,\Big|\frac{1}{n}\sum_{j=k'}^{k-2} b\big(Y^n_{\frac{j-1}{n}}\big) + b\big(Y^n_{\frac{k-1}{n}}\big)\Big(t - \frac{k-1}{n}\Big)\Big|^4\,\Big] + 8\,\mathbb{E}\big[\,|M_t - M_s|^4\,\big].$$

The second upper bound is bounded above by $9A^4\,|t-s|^2$, while the term inside $|\cdot|^4$ is bounded above by $A\,|t-s|$, where A is chosen big enough to be an upper bound for b; the upper bound (9.4) follows.

$\triangleright$

9.2. ***Complement: Regular conditional probability.*** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. As one has almost-surely

- $\mathbb{E}[\mathbf{1}_\emptyset|\mathcal{G}] = 0$ and $\mathbb{E}[\mathbf{1}_\Omega|\mathcal{G}] = 1$,
- $\mathbb{E}[\sum_{n\geqslant 0} \mathbf{1}_{A_n}|\mathcal{G}] = \sum_{n\geqslant 0} \mathbb{E}[\mathbf{1}_{A_n}|\mathcal{G}]$, for any sequence of disjoint $A_n \in \mathcal{F}$,

the map $A \in \mathcal{F} \mapsto \mathbb{P}(A|\mathcal{G}) := \mathbb{E}[\mathbf{1}_A|\mathcal{G}]$ has the properties of a probability, except that $\mathbb{P}(A|\mathcal{G})$ is a random variable (*i.e.* an equivalence class of functions, and so the above identities hold only almost-surely). It is thus natural to ask whether the random variables $\mathbb{P}(\cdot|\mathcal{G})$ can be written as $\mathbb{P}_\omega(\cdot)$ for some random probability measure $\omega \to \mathbb{P}_\omega$ (but we must be careful with measurability issues). Although such a $\mathbb{P}_\omega$ can be defined for a given *sequence* of sets $A_n \in \mathcal{F}$, the problem is that, except in trivial cases, there are *uncountably* many sequences of disjoint sets (hence meaurability problems); it is therefore not at all clear how to choose $\mathbb{P}_\omega$. And indeed, there is no such family of probabilities if no hypothesis on $(\Omega, \mathcal{F})$ is made: you can find the classical counter-example of J. Dieudonné in §43 of the book [RW00]. Yet one can construct such probabilities $\mathbb{P}_\omega$ when we are working on a Borel probability space. To stick to the previous notations I will denote a Borel space by $(S, \mathcal{S})$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be any probability space and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$(68).

Definition **87.** *A* ***regular conditional probability of*** $\mathbb{P}$ ***given*** $\mathcal{G}$ *is a family* $(\mathbb{P}_\omega)_{\omega\in\Omega}$ *of probability measures on* $(\Omega, \mathcal{F})$ *such that the function* $\omega \mapsto \mathbb{P}_\omega(A)$ *is measurable and belongs to the equivalent class of* $\mathbb{P}(A|\mathcal{G})$*, for every* $A \in \mathcal{F}$*.*

Theorem **88** (Existence and uniqueness of regular conditional probability in Borel spaces)**.** *Let* $(S, \mathcal{S}, \mathbb{P})$ *be a Borel probability space and* $\mathcal{G}$ *be a sub-*σ*-algebra of* $\mathcal{S}$*. Then there exists a regular conditional probability of* $\mathbb{P}$ *given* $\mathcal{G}$*, unique up to equivalence.*

Proof – *Existence.* We have seen in part **a)** of the proof of theorem 40 that $[0,1]$ is isomorphic to a measurable subset of $\{0,1\}^{\mathbb{N}}$; we can thus suppose without loss of generality that S is a measurable subset of $\{0,1\}^{\mathbb{N}}$, that $\mathcal{S}$ is the restriction to S of the product (or Borel) σ-algebra of $\{0,1\}^{\mathbb{N}}$, and that $\mathbb{P}$ is a probability on $\{0,1\}^{\mathbb{N}}$, with support on S. Our main ingredient to prove the existence of regular conditional probabilities will be theorem 41, stated in the

[68] The end of this section follows closely B. Tsirelson's lecture notes *Probability for mathematicians*, available at the webpage http://tau.ac.il/ tsirel/Courses/ProbMath/main.html. This proof is essentially the same as that of Ikeda-Watanabe, in [IW89].

Complement *Lebesgue measure on* $[0,1]$, in part I of the course: *Borel probability measures on* $\{0,1\}^{\mathbb{N}}$ *correspond bijectively to* additive *set functions on* $\mathcal{A}$*, equal to* 0 *on* $\emptyset$ *and* 1 *on* Ω. The generating algebra $\mathcal{A}$ being countable we can define $\mathbb{P}_\omega(A) = \mathbb{P}(A|\mathcal{G})(\omega)$ for all $A \in \mathcal{A}$ choosing a function in each equivalent class. As $\mathbb{P}(\cdot|\mathcal{G})$ is almost-surely finitely additive (on $\mathcal{S}$), $\mathbb{P}_\omega(\cdot)$ can be turned into a finitely additive set function on the countable collection $\mathcal{A}$ for all ω, by changing it adequately on a set of null probability. By theorem 41 each $\mathbb{P}_\omega(\cdot)$ has an extension to a probability measure on $(S, \mathcal{S})$; we still denote it by $\mathbb{P}_\omega(\cdot)$. Given any $B \in \mathcal{S}$, write $\mathbb{P}_\bullet(B)$ for the measurable[69] function $\omega \mapsto \mathbb{P}_\omega(B)$.

$B \in \mathcal{S}$ and $\varepsilon > 0$ being given, we shall see below that there exists $C, D \in \mathcal{S}$ such that

- $C \subset B \subset D$, $\mathbb{P}(D \backslash C) \leqslant \varepsilon$,
- C is a decreasing limit of elements of $\mathcal{A}$ and D an increasing limit of elements of $\mathcal{A}$.

As $\mathbb{P}_\bullet(A_n) = \mathbb{P}(A_n|\mathcal{G})$ almost-surely, we have almost-surely $\mathbb{P}_\bullet(C) = \mathbb{P}(C|\mathcal{G})$ and $\mathbb{P}_\bullet(D) = \mathbb{P}(D|\mathcal{G})$ by monotone convergence (for $\mathbb{P}_\bullet(\cdot)$ and $\mathbb{P}(\cdot|\mathcal{G})$). As a consequence,

$$\begin{cases} \mathbb{P}(C|\mathcal{G}) = \mathbb{P}_\bullet(C) \leqslant \mathbb{P}_\bullet(B) \leqslant \mathbb{P}_\bullet(D) = \mathbb{P}(D|\mathcal{G}), \\ \mathbb{P}(C|\mathcal{G}) \leqslant \mathbb{P}(D|\mathcal{G}) \leqslant \mathbb{P}(D|\mathcal{G}), \end{cases}$$

and so $\big|\mathbb{P}_\bullet(B) - \mathbb{P}(D|\mathcal{G})\big| \leqslant \mathbb{P}(D|\mathcal{G}) - \mathbb{P}(C|\mathcal{G})$. Also, $\mathbb{E}\big[\mathbb{P}(D|\mathcal{G}) - \mathbb{P}(C|\mathcal{G})\big] = \mathbb{P}(D\backslash C) \leqslant \varepsilon$. Taking a sequence $(\varepsilon_n)_{n\geqslant 0}$ decreasing fastly enough to 0 we get

$$\big|\mathbb{P}_\bullet(B) - \mathbb{P}(B|\mathcal{G})\big| \leqslant \inf_n \mathbb{P}(D_n|\mathcal{G}) - \mathbb{P}(C_n|\mathcal{G}) = 0 \quad \text{almost-surely.}$$

Uniqueness. As two possible regular conditional probabilities coincide almost-surely on $\mathcal{A}$ (which is countable) they must be equal on $\mathcal{S}$ by the monotone class theorem. $\triangleright$

It remains to justify the approximation result used in the existence proof; we do it for C, the argument for D being similar. Let $B \in \mathcal{S}$ be given. It comes out from the proof of Caratheodory's theorem given in section 1.2 that for any $\varepsilon > 0$ there exists an element A of $\mathcal{A}$ with the property that $\mathbb{P}(B \Delta A) \leqslant \varepsilon$. Apply this result inductively first to B and $\varepsilon = \eta$ (we get A_1), then to $B \cap A_1$ and $\varepsilon = 2^{-1}\eta$ (we get A_2), then to $B \cap (A_1 \cap A_2)$ and $\varepsilon = 2^{-2}\eta$ (we get A_3)...The set $C = \bigcap_{n \geqslant 0} A_n$ is the decreasing limit ot the $\bigcap_{k \leqslant p} A_k$ and $\mathbb{P}(B \Delta C) \leqslant 2\eta$. I let you conclude.

[69]Prove that it is indeed measurable.

Part III. Brownian motion, Lévy processes and martingales

Let us consider a physical system subject to an impredictable evolution. We model its random evolution by a filtered probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P}\big)$, where we can think of Ω as the set of all physically possible histories of the system through time, of $\mathcal{F}$ as the set of all observations one can make and of $\mathcal{F}_t$ as our information at time t of the history of the system up to that time. In that setting, (sub/super)-martingales represent quantitative informations about the system which ("increase"/"decrease") remain "constant"[70]. Sub/super-martingales have thus a universal status in the description Natural phenomena evolving randomly. In that landscape, Brownian motion plays a prominent role as we shall see that any continuous time continuous martingale can be understood as a Brownian motion run at a random speed. Being also a Markov process and a Gaussian process[71], we can say without hesitation that it is a cornerstone of modern probability theory. Section 10 is devoted to the study of some of its elementary features. Section 11 presents Lévy processes, with the help of which we shall describe the most general càdlàg continuous time martingale.

Recall that the martingale property is not an absolute property: it is related to a filtration. When unspecified, it will be implicit that we are working with the filtration generated by the process under study. Also, all filtrations will be supposed to be complete.

10. Brownian motion

10.1. Different point of views on Brownian motion.

- **a)** Lévy's constuction of Brownian motion as a series.
- **b)** **Markov process.** By its very definition, Brownian motion is a Markov process with Gaussian transition kernels.
- **c)** **Gaussian process.** In exercice 4 of example sheet 1, Brownian motion is characterized as the unique centered Gaussian process with covariance $s \wedge t$.
- **d)** **Scaling limit.** Donsker's invariance principle provides a construction based on a scaling limit of random walks,

a) Since Brownian motion has Gaussian increments, we know from Kolmogorov's regularity criterion that, for all $\alpha < \frac{1}{2}$, it has almost-surely α-Hölder-continuous paths; so its paths does not seem to be too bad. Yet, we shall prove in proposition 93 that it is almost-surely not differentiable anywhere and that it is almost-surely nowhere α-Hölder-continuous for $\alpha > \frac{1}{2}$. From this picture, it comes as a good news that Lévy's construction provides almost for free the following precise description of the local behaviour of Brownian motion.

Proposition **89** (Modulus of continuity for Brownian motion). *There exists a constant C and a positive random variable δ such that one has $\mathbb{P}$-almost surely*

$$|X_t - X_s| \leqslant C\sqrt{|t-s| \ln \frac{1}{|t-s|}}$$

for all $t, s \in [0,1]$, with $|t-s| \leqslant \delta$.

[70] In the sense that the prediction $\mathbb{E}[M_t|\mathcal{F}_s]$ of their future value equals their present value M_s.

[71] Two fundamental classes of random processes for modelization purposes.

Proof – Recall Lévy's construction of Brownian motion as a series $\sum_{n\geqslant 1}\big(B^{(n)}-B^{(n-1)}\big)$ of continuous piecewise linear functions. Given $c>\sqrt{2\log 2}$, there exists a random integer n_0 such that $\big\|B^{(n)}-B^{(n-1)}\big\|_\infty \leqslant c\sqrt{n}\,2^{-\frac{n}{2}}$ for $n\geqslant n_0$. As we have by construction

$$\big\|\big(B^{(n)}-B^{(n-1)}\big)'\big\|_\infty \leqslant 2\frac{\|B^{(n)}-B^{(n-1)}\|_\infty}{2^{-n}} \leqslant 2c\sqrt{n}\,e^{\frac{n}{2}},$$

the mean-value theorem gives us, for $t,t+h$ in $[0,1]$, and any $p\geqslant n_0$,

$$\begin{aligned}|B_{t+h}-B_t| &\leqslant \sum_{n\geqslant 1}\big|\big(B^{(n)}_{t+h}-B^{(n-1)}_{t+h}\big)-\big(B^{(n)}_t-B^{(n-1)}_t\big)\big|\\ &\leqslant \sum_{n=1}^{p} h\big\|\big(B^{(n)}-B^{(n-1)}\big)'\big\|_\infty + 2\sum_{n=p+1}^{\infty}\big\|B^{(n)}-B^{(n-1)}\big\|_\infty\\ &\leqslant \sum_{n=1}^{n_0} h\big\|\big(B^{(n)}-B^{(n-1)}\big)'\big\|_\infty + 2ch\sum_{n=n_0+1}^{p}\sqrt{n}\,2^{\frac{n}{2}} + 2c\sum_{n=p+1}^{\infty}\sqrt{n}\,2^{-\frac{n}{2}}\end{aligned}$$

As the second sum is dominated by a constant multiple of its biggest element, bound above the sum of the last two terms by $c'\big(h\sqrt{p}2^{\frac{p}{2}}+\sqrt{p}2^{-\frac{p}{2}}\big)$, for some positive constant c'. One can take $p=\lfloor \log_2\frac{1}{h}\rfloor$ for h small enough. A simple calculus gives us a constant C satisfying the inequality $2c\big(hp\sqrt{p}e^{\frac{p}{2}}+\sqrt{p}2^{-\frac{p}{2}}\big)\leqslant C\sqrt{h\ln\frac{1}{h}}$. As $C\sqrt{h\ln\frac{1}{h}}$ is also bigger than $\sum_{n=1}^{n_0} h\,\big\|\big(B^{(n)}-B^{(n-1)}\big)'\big\|_\infty$, for h small enough, this proves the statement. ▷

b) The Markovian approach to Brownian motion is extremely fruitful. Let X be a Brownian motion defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$; we write $\big(\mathcal{F}_t\big)_{t\geqslant 0}$ for its natural (completed) filtration. Denote by $\mathbb{P}_x$ the law of the Brownian motion $x+X$ starting from $x\in\mathbb{R}^d$ and set for any non-negative function f

$$T_tf(x)=\mathbb{E}_x\big[f(X_t)\big]=(2\pi t)^{-\frac{1}{2}}\int f(y)e^{-\frac{|y-x|^2}{2t}}dy.$$

Last, recall that $\mathcal{F}_{t^+}:=\bigcap_{s>t}\mathcal{F}_s$, for any $t\geqslant 0$.

Theorem **90** (Simple Markov property)**.** *Let $t\geqslant 0$ be given.*

(1) *Given any $x\in\mathbb{R}^d$, the Brownian motion $(X_{t+s}-X_t)_{s\geqslant 0}$ is independent of $\mathcal{F}_{t^+}$ under $\mathbb{P}_x$.*

(2) *Given any $x\in\mathbb{R}^d$ and $A\in\sigma(X_{t+s}\,;\,s\geqslant 0)$, we have $\mathbb{P}_x$-almost-surely*[72]

$$\mathbb{P}_x(A|\mathcal{F}_{t^+})=\mathbb{P}_{X_t}(A). \tag{10.1}$$

(3) *For any $\mathcal{C}^2$ bounded function f*

$$T_tf-f=\frac{1}{2}\int_0^t T_s(\triangle f)\,ds=\frac{1}{2}\int_0^t \triangle(T_sf)\,ds.$$

Proof – (1) First, it comes directly from the independence of the increments of Brownian motion that the Brownian motion $(X_{t+s}-X_t)_{s\geqslant 0}$ is independent of $\mathcal{F}_t$ under $\mathbb{P}$. It follows in particular that the process $(X_{t+\varepsilon+s}-X_{t+\varepsilon})_{s\geqslant 0}$ is independent of $\mathcal{F}_{t+\varepsilon}$, so of $\mathcal{F}_{t^+}$, for any $\varepsilon>0$. The vectors $\big(X_{t+s_1}-X_t,\dots,X_{t+s_n}-X_t\big)=\lim_{\varepsilon\to 0}\big(X_{t+\varepsilon+s_1}-X_{t+\varepsilon},\dots,X_{t+\varepsilon+s_n}-X_{t+\varepsilon}\big)$, are thus independent of $\mathcal{F}_{t^+}$ for any $s_1,\dots,s_n>0$; we are done as this is a Gaussian vector with the awaited covariance matrix. This means that, conditionally on $\mathcal{F}_{t^+}$, the process

[72] Note that the measurability of the map $x\mapsto\mathbb{P}_x(A)$ is trivial for elementary events A; it follows that the map is measurable for any event $A\in\mathcal{F}$.

$\big(X_{t+s}\big)_{s\geqslant 0}$ is a Brownian motion starting from X_t; call it $(X'_s)_{s\geqslant 0}$. This implies in particular that we have for any bounded function f and any $s,t\geqslant 0$

$$\begin{aligned} T_{s+t}f(x) &= \mathbb{E}_x\big[f(X_{s+t})\big] = \mathbb{E}_x\Big[\mathbb{E}_x\big[f(X_{s+t})\big|\mathcal{F}_{t^+}\big]\Big] \\ &= \mathbb{E}_x\Big[\mathbb{E}_{X_t}\big[f(X'_s)\big]\Big] = T_t\big(T_s f\big)(x) \end{aligned}$$

(2) By the monotone class theorem, it suffices to prove that we have

$$\mathbb{E}_x[\mathbf{1}_A\mathbf{1}_B] = \mathbb{E}_x\big[\mathbb{P}_{X_t}(A)\mathbf{1}_B\big]$$

for any $A\in\sigma(X_{t+s}\,;\,s\geqslant 0)$ of the elementary form $\big\{X_{t+s_1}\in A_1,\dots,X_{t+s_n}\in A_n\big\}$, with $s_i>0$, and $B\in\mathcal{F}_t$ of a similar form. But for such A and B we have by the first point

$$\mathbb{E}_x[\mathbf{1}_A\mathbf{1}_B] = \mathbb{E}_x\big[\mathbb{E}_x[\mathbf{1}_A\mathbf{1}_B\,|\,\mathcal{F}_{t^+}]\big] = \mathbb{E}_x\big[\mathbf{1}_B\mathbb{E}_x[\mathbf{1}_A\,|\,\mathcal{F}_{t^+}]\big] = \mathbb{E}_x[\mathbb{P}_{X_t}(A)\mathbf{1}_B].$$

(3) This point expresses the fact that for any $\mathcal{C}^2$ bounded function, and any x, the map $t\mapsto T_tf(x)$ is differentiable, with derivative $\frac{1}{2}T_t(\triangle f)(x)\big(=\frac{1}{2}\triangle(T_tf)(x)\big)$. This can be checked by a trivial integration by parts using the fact that the heat kernel $\varphi:(t,y)\mapsto\exp\big(-\frac{|y-x|^2}{2t}\big)$ solves the heat equation $\partial_t\varphi=\frac{1}{2}\triangle\varphi$. ▷

Corollary **91** (Blumenthal's $0-1$ law)**.** *Events of $\mathcal{F}_{0^+}$ are trivial under any $\mathbb{P}_x$: we have $\mathbb{P}_x(A)\in\{0,1\}$, for any $A\in\mathcal{F}_{0^+}, x\in\mathbb{R}^d$.*

Proof – Indeed, for $A\in\mathcal{F}_{0+}$ we have $\mathbb{P}_x$-almost-surely $\mathbf{1}_A=\mathbb{E}_x[\mathbf{1}_A\,|\,\mathcal{F}_{0+}]=\mathbb{P}_{X_0}(A)=\mathbb{P}_x(A)$. ▷

This $0-1$ law has deep and far-reaching consequences, of which the exercises provide a few examples.

Proposition **92.** *Given a one-dimensional Brownian motion X define $\tau=\inf\{t>0\,;\,X_t>0\}$ and $\tau'=\inf\{t>0\,;\,X_t<0\}$. Then almost-surely $\tau=\tau'=0$.*

Proof – One easily check that the events $\{\tau=0\}$ and $\{\tau'=0\}$ belong to $\mathcal{F}_{0+}$. As $-X$ has the same law as X we have $\mathbb{P}(\tau=0)=\mathbb{P}(\tau'=0)$. To prove that τ is almost-surely equal to 0 it suffices, by Blumenthal's law, to see that $\mathbb{P}(\tau=0)>0$. But as we have for any $t>0$, $\mathbb{P}(\tau\leqslant t)\geqslant\mathbb{P}(X_t\geqslant 0)=\frac{1}{2}$, this is straightforward. ▷

Here is another pathwise property of Brownian motion easily obtained from the Markovian point of view.

Proposition **93.** *Let $\alpha>\frac{1}{2}$. Brownian motion is almost-surely nowhere α-Hölder continuous. In particular, it is almost-surely nowhere differentiable.*

Proof – Let $p\geqslant 2$ be an integer to be chosen later. Fixing some $K>0$, define an increasing sequence of events

$$A_n=\Big\{\text{for some } s\in[0,1],\ |X_t-X_s|\leqslant K|t-s|^\alpha,\ \text{whenever } |t-s|\leqslant\frac{p}{n}\Big\},\quad n\geqslant 2,$$

and name the increments $\Delta_{k,n}=\big|X_{\frac{k}{n}}-X_{\frac{k-1}{n}}\big|,\quad k=1..n$. Then

$$A_n\subset\bigcup_{k=2}^{n}\Big\{\Delta_{j,n}\leqslant 2K\frac{p^\alpha}{n^\alpha}\ \text{for each } j\in\{k-1,\dots,k+p-1\}\Big\},$$

and so

$$\mathbb{P}(A_n)\leqslant(n-2)\,\mathbb{P}\Big(\Delta_{1,n}\leqslant 2K\frac{p^\alpha}{n^\alpha}\Big)^p\leqslant n\,\mathbb{P}\Big(\big|\mathcal{N}(0,1)\big|\leqslant 2K\frac{p^\alpha}{n^{\alpha-\frac{1}{2}}}\Big)^p\leqslant c\,n^{\frac{p+2}{2}-p\alpha}$$

for some positive constant c. The upper bound converges to 0 as n goes to infinity if we choose p such that $\alpha > \frac{p+2}{2p}$. As the events A_n increase, $\mathbb{P}(A_n) = 0$ for all n. ▷

c) The Gaussian characterization provides deep insight, starting with a straightforward justification of the following facts.

PROPOSITION **94.** *Let X be an $\mathbb{R}^d$-valued Brownian motion starting from 0([73]). The following processes are also Brownian motions.*

(1) Invariance by isometries: *AX, for any linear isometry A.*
(2) *$(X_{t+s} - X_t)_{s\geqslant 0}$, for any $t \geqslant 0$,*
(3) *$\big(cX_{c^{-2}t}\big)_{t\geqslant 0}$, for any $c > 0$,*
(4) Time inversion: *$\big(tX_{1/t}\big)_{t\geqslant 0}$, with value 0 in $t = 0$,*

The time inversion property implies in particular that $\frac{X_t}{t}$ converges almost-surely to 0 as $t \to +\infty$.

PROOF – The only non-trivial point is the continuity at 0 of the process $B_t := tX_{1/t}$. Since B is almost-surely continuous on $(0,\infty)$ one can describe the event $\{B \underset{t\downarrow 0}{\to} 0\}$ in terms of conditions on the values of B at countably many points of $(0,\infty)$. But B and X being Gaussian, with the same covariance and the same value at time 1, they have the same law on $(0,\infty)$; so $\mathbb{P}\big(B_t \underset{t\downarrow 0}{\to} 0\big) = \mathbb{P}\big(X_t \underset{t\downarrow 0}{\to} 0\big) = 1$. ▷

PROPOSITION **95** (Quadratic variation of Brownian motion)**.** • *Let $t \geqslant 0$ be given. For $n \geqslant 0$ and $k \geqslant 0$, set $t_k^n = t \wedge k2^{-n}$. The quadratic variation of X over the dyadic partition $\langle X\rangle_t^n := \sum_{k\geqslant 0}\big(X_{t_k^n} - X_{t_{k-1}^n}\big)^2$ converges almost-surely and in $\mathbb{L}^2$ to t.*

PROOF – As $X_{s+h} - X_s \sim \mathcal{N}(0,h)$ we have $\mathbb{E}\big[(X_{s+h} - X_s)^2\big] = h$ and $\mathrm{VAR}\big((X_{s+h} - X_s)^2\big) = \mathrm{VAR}(N)h^2$, where N is the square of a normal random variable. Hence, for fixed t, the random variable $\langle X\rangle_t^n$ has mean t and variance $\mathrm{VAR}(N)\,2^{-n}$. The result follows. ▷

d) The scaling limit approach provides an immediate proof of the following fact.

PROPOSITION **96.** *We have almost-surely $\overline{\lim}_{t\to+\infty} X_t = -\underline{\lim}_{t\to+\infty} X_t = +\infty$,*

Much deeper insights can be gained from that picture, like the fact that some Galton-Watson branching processes and some continuous random trees are hidden in Brownian trajectories... But this is another story.

10.2. **Constructing martingales.** The use of martingales constructed from Brownian motion can provide much information about it. The following theorem provides a canonical way of constructing martingales associated with an $\mathbb{R}^n$-valued Brownian motion. You should compare it with proposition 73 characterizing the law of a Markov chain in terms of martingales.

THEOREM **97.** *Let B be a Brownian motion defined on some filtered probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P}\big)$, and let $f \in \mathcal{C}^{1,2}\big(\mathbb{R}_+ \times \mathbb{R}^d\big)$ be such that*

$$\big|f(t,x)\big| + \big|\partial_t f\big|(t,x) + \sum_{i=1..d}\big|\partial_{x_i} f\big|(t,x) + \sum_{i=1..d}\big|\partial^2_{x_i,x_j} f\big|(t,x) \leqslant Ke^{K(t+|x|)}$$

[73]The precise probability space on which it is defined is irrelevant.

for some positive constant K. Then the process

$$M_t = f(t, B_t) - f(0, B_0) - \int_0^t \Big(\frac{\partial}{\partial r} + \frac{1}{2}\triangle\Big) f(r, B_r)\, dr$$

is a martingale with respect to the Brownian filtration.

Proof – The hypotheses are designed so as to ensure the integrability of any M_t. We have to show that we have almost-surely $\mathbb{E}\big[M_{s+t} - M_s \big| \mathcal{F}_s\big] = 0$, for all $0 \leqslant s, t$. Write $\widetilde{\mathcal{F}}_t$ for $\mathcal{F}_{s+t}$ and $\widetilde{B}_t = B_{t+s} - B_s$; it is an $\widetilde{\mathcal{F}}_t$-Brownian motion independent of $\mathcal{F}_s$ conditionally on $B_s = \widetilde{B}_0$. Noting that

$$\begin{aligned} M_{s+t} - M_s &= f(s+t, B_{s+t}) - f(s, B_s) - \int_s^{s+t} \Big(\frac{\partial}{\partial r} + \frac{1}{2}\triangle\Big) f(r, B_r)\, dr \\ &= \widetilde{f}\big(t, \widetilde{B}_t + \widetilde{B}_0\big) - \widetilde{f}\big(0, \widetilde{B}_0\big) - \int_0^t \Big(\frac{\partial}{\partial r} + \frac{1}{2}\triangle\Big) \widetilde{f}\big(r, \widetilde{B}_r + \widetilde{B}_0\big)\, dr, \end{aligned}$$

where $\widetilde{f}(t,x) = f(s+t, x)$, we see that it suffices to prove that

$$\mathbb{E}\Big[\widetilde{f}\big(t, \widetilde{B}_t + \widetilde{B}_0\big) - \widetilde{f}\big(0, \widetilde{B}_0\big) - \int_0^t \Big(\frac{\partial}{\partial r} + \frac{1}{2}\triangle\Big) \widetilde{f}\big(r, \widetilde{B}_r + \widetilde{B}_0\big)\, dr \Big| \widetilde{\mathcal{F}}_0\Big] = 0,$$

or, equivalently, that $\mathbb{E}_x[M_t] = 0$ for any starting point $x \in \mathbb{R}^d$. Write $p_r(x,y) = (2\pi\, r)^{-\frac{d}{2}} \exp\Big(-\frac{|y-x|^2}{2}\Big)$ for the Gaussian kernel. Noting that we have for $0 < s < t$

$$\mathbb{E}_x[M_t - M_s] = \int_{\mathbb{R}^d} f(t,y)\,\big(p_t(x,y) - p_s(x,y)\big) dy - \int_s^t \Big(\int_{\mathbb{R}^d} p_r(x,y) \Big(\frac{\partial}{\partial r} + \frac{1}{2}\triangle\Big) f(r,y)\, dy\Big)\, dr,$$

and that $p_r(x,y)$ satisfies the heat equation $\Big(\frac{\partial}{\partial r} - \frac{1}{2}\triangle_y\Big) p_r(x,y) = 0$, for $r > 0$, an integration by parts (twice with respect to y and once with respect to r) gives

$$\begin{aligned} \int_s^t \Big(\int_{\mathbb{R}^d} p_r(x,y) \Big(\frac{\partial}{\partial r} + \frac{1}{2}\triangle\Big) f(r,y)\, dy\Big)\, dr &= \int_s^t \frac{\partial}{\partial_r} \big(p_r(x,y)\, f(r,y)\big)\, dy dr \\ &= \int_{\mathbb{R}^d} f(t,y) p_t(x,y)\, dy - \int_{\mathbb{R}^d} f(s,y) p_s(x,y)\, dy, \end{aligned}$$

from which the identity $\mathbb{E}_x[M_t - M_s] = 0$ follows. It remains to notice that $\mathbb{E}_x[M_s]$ goes to 0 as s goes to 0 to conclude. ▷

Corollary **98** (Recurrence and transience of Brownian motion). (1) *Given any starting point different from 0, the 2-dimensional Brownian motion has probability* 0 *of ever hitting* $\{0\}$*, but it hits almost-surely any neighbourhood of* 0 *at arbitrarily large times.*

(2) *In dimension bigger than* 3*, we have almost-surely* $|B_t| \to +\infty$ *as* $t \to +\infty$*.*

Proof – (1) *i)* Let $0 < a < |x| < b$. The function $\log|x|$ satisfies the identity $\triangle f = 0$ on $\mathbb{R}^2\backslash\{0\}$. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a smooth function defined on $\mathbb{R}^2$ and coinciding with $x \mapsto \log|x|$ outside the ball of radius a. As f has a sub-exponential growth, we can use it to construct the martingale M_t used in theorem 97. Let T be the hitting time $\inf\big\{t \geqslant 0\,;\, |X_t| \in \{a,b\}\big\}$; as f and $\log|x|$ coincide outside the ball of radius a we have $M_{t\wedge T} = \log|X_{t\wedge T}|$, for all times; also this stopped martingale is bounded, from the definition of T. Applying the optional stopping theorem, we then have

$$\mathbb{E}_x\big[\log|B_T|\big] = \mathbb{E}_x\big[\log|B_0|\big] = \log|x|;$$

so $\mathbb{P}_x\big(|B_T| = a\big) = \frac{\log\frac{|x|}{b}}{\log\frac{a}{b}}$. Sending a to 0 gives the first result.

ii) To deal with the case $B_0 = 0$, use the Markov property to write

$$\mathbb{P}_0\big(B_t = 0 \text{ at some time } t \geqslant \epsilon\big) = \mathbb{E}_0\Big[\mathbb{P}_{B_\epsilon}\big(B_t = 0 \text{ at some time } t \geqslant 0\big)\Big] = 0;$$

as this holds for any $\epsilon > 0$ we conclude that $\mathbb{P}_0\big(B_t = 0 \text{ at some time } t > 0\big) = 0$.
iii) Fixing now a and sending b to infinity we see that B hits $\mathbb{P}^x$-almost-surely the ball of radius a if $|x| = 1$. Proceeding as in *ii)* this fact is seen to be true for all every starting point. The result now follows from Markov property as the $\mathbb{P}_x$-probability that B hits the ball of radius a after time n equals $\mathbb{E}_x\Big[\mathbb{P}_{B_n}\big(B_t = a \text{ at some time } t \geqslant 0\big)\Big] = 1$.

(2) Use the function $|x|^{2-d}$ to prove as above that we have $\mathbb{P}_x(H_a < H_b) = \frac{|x|^{2-d} - b^{2-d}}{a^{2-d} - b^{2-d}}$ when $a < |x| < b$. Conclude as above. ▷

10.3. **Strong Markov property.** In this section we take as a framework a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which a Brownian motion X is defined, and denote by $\big(\mathcal{F}_t\big)_{t\geqslant 0}$ the (completed) filtration generated by X.

THEOREM **99** (Strong Markov property)**.** *Let T be a stopping time such that $\mathbb{P}(T < \infty) > 0$. Define the X^T by $X^T_t = X^{T+t} - X^T$ on the event $\{T < \infty\}$, and by $X^T_t \equiv 0$ on the event $\{T = \infty\}$. Then, conditionally on $\{T < \infty\}$, the process X^T is a Brownian motion independent of $\mathcal{F}_T$. This means that for any $A \in \mathcal{F}_T$, any times $0 < t_1 < \cdots < t_n$, and any bounded measurable function $F : \mathbb{R}^n \to \mathbb{R}$, we have*

$$\mathbb{E}\big[\mathbf{1}_{A\cap\{T<\infty\}}\, F\big(X^T_{t_1}, \ldots, X^T_{t_n}\big)\big] = \mathbb{P}\big(A \cap \{T < \infty\}\big)\, \mathbb{E}\big[F(X_{t_1}, \ldots, X_{t_n})\big]. \tag{10.2}$$

PROOF – By a monotone class argument, it suffices to prove (10.2) for bounded continuous functions F. Take such a function and note that due to the continuity (from the right!) of X and F the quantity $\mathbb{E}\big[\mathbf{1}_{A\cap\{T<\infty\}} F\big(X^T_{t_1}, \ldots, X^T_{t_n}\big)\big]$ is, by dominated convergence, equal to

$$\lim_{p\to+\infty} \sum_{k\geqslant 0} \mathbb{E}\big[\mathbf{1}_{A\cap\{(k-1)2^{-p}<T\leqslant k2^{-p}\}} F\big(X_{k2^{-p}+t_1} - X_{k2^{-p}}, \ldots, X_{k2^{-p}+t_n} - X_{k2^{-p}}\big)\big].$$

Note that we have approximated X_{T+t_i} by $X_{k2^{-p}+t_i}$. As the event A belongs to $\mathcal{F}_T$ the event $A \cap \{(k-1)2^{-p} < T \leqslant k2^{-p}\}$ belongs to $\mathcal{F}_{k2^{-p}}$; so the simple Markov property enables us to write the generic term of the above sum as

$$\mathbb{P}\big(A \cap \{(k-1)2^{-p} < T \leqslant k2^{-p}\}\big)\, \mathbb{E}\big[F(X_{t_1}, \ldots, X_{t_n})\big].$$

Summing over $k \geqslant 0$ and taking the limit $p \to +\infty$ gives (10.2). ▷

This fundamental property of Brownian motion has tremendously many applications, of which the following ones are remarkable.

COROLLARY **100** (Reflection principle)**.** *Let X be a Brownian motion starting from 0 and T be a finite stopping time. Set $Y_t = X_t$ for $t \leqslant T$, and $Y_t = 2X_T - X_t$ for $t \geqslant T$. Then Y is also a Brownian motion.*

PROOF – We need to prove that we have for any times $0 < t_1 < \cdots < t_n$ and any continuous bounded function $F : \mathbb{R}^n \to \mathbb{R}$

$$\mathbb{E}\big[F(Y_{t_1}, \ldots, Y_{t_n})\big] = \mathbb{E}\big[F(X_{t_1}, \ldots, X_{t_n})\big]. \tag{10.3}$$

Setting $t_0 = 0$ and $t_{n+1} = \infty$, we have

$$\begin{aligned}\mathbb{E}\big[F(Y_{t_1}, \ldots, Y_{t_n})\big] &= \sum_{i=1..n+1} \mathbb{E}\big[F(Y_{t_1}, \ldots, Y_{t_n})\mathbf{1}_{t_{i-1}\leqslant T<t_i}\big] \\ &= \sum_{i=1..n+1} \mathbb{E}\big[F\big(X_{t_1}, \ldots, X_{t_{i-1}}, X_T + (X_T - X_{t_i}), \cdots, X_T + (X_T - X_{t_n})\big)\mathbf{1}_{t_{i-1}\leqslant T<t_i}\big].\end{aligned}$$

But the generic term of the above sum equals

$$\mathbb{E}\Big[\mathbb{E}\big[F\big(X_{t_1},\dots,X_{t_{i-1}},X_T+(X_T-X_{t_i}),\cdots,X_T+(X_T-X_{t_n})\big)\mathbf{1}_{t_{i-1}\leqslant T<t_i}\big|\mathcal{F}_T\big]\Big] =$$
$$\mathbb{E}\Big[\mathbb{E}\big[F\big(X_{t_1},\dots,X_{t_{i-1}},X_{t_i},\cdots,X_{t_n}\big)\mathbf{1}_{t_{i-1}\leqslant T<t_i}\big|\mathcal{F}_T\big]\Big]$$

by the strong Markov property and because the opposite of a Brownian motion is a Brownian motion; summing these terms gives identity (10.3). ▷

Corollary **101** (Maximum process – Bachelier)**.** *Given a real-valued Brownian motion X, and $t \geqslant 0$, define $M_t^X := \max_{s\leqslant t} X_s$, for $t \geqslant 0$. Then*

$$M_t^X \overset{d}{=} M_t^X - X_t \overset{d}{=} |X_t|.$$

Proof – Denote again by Y the process defined in corollary 100. Let $a \geqslant 0$ and b be two real numbers such that $a \geqslant \max\{b,0\}$. Let $T = \inf\{s \geqslant 0\,;\, X_s = a\}$, this is an almost-surely finite stopping time (why?). The reflection principle justifies the first identity below; draw a picture to understand the third identity.

$$\begin{aligned}\mathbb{P}\big(M_t^X \geqslant a, X_t \leqslant b\big) &= \mathbb{P}\big(M_t^Y \geqslant a, Y_t \leqslant b\big) = \mathbb{P}\big(M_t^X \geqslant a, Y_t \leqslant b\big)\\ &= \mathbb{P}(X_t \geqslant 2a-b).\end{aligned}$$

This identity gives the law of the pair $\big(M_t^X, X_t\big)$, from which the result follows. ▷

10.4. **Brownian motion and the Dirichlet problem.** Let B be a bounded open set of some $\mathbb{R}^d$, with non-empty boundary ∂B, and f be a measurable real-valued function defined on ∂B. To ***solve the Dirichlet problem in*** B ***with boundary condition*** f is to find a function g defined on the closure $\overline{B}$ of B which is

(10.4)
- of class $\mathcal{C}^2$ in B, with $\triangle g = 0$ in B,
- continuous on $\overline{B}$, with restriction to ∂B equal to f.

Functions g of class $\mathcal{C}^2$ satisfying the condition $\triangle g = 0$ in B are said to be ***harmonic in*** B. You are asked to prove in exercise the following characterization of harmonic functions in terms of spheric means. For an open ball $B(x,r) \subset B$ we write $\sigma_{x,r}(dy)$ for the uniform probability on the sphere $\big\{y \in B\,;\, |y-x| = r\big\}$.

Proposition **102** (Gauss)**.** *A non-negative function g such that $g(x) = \int g(y)\sigma_{x,r}(dy)$ for any ball $B(x,r) \subset B$ is either $\equiv \infty$ or harmonic in B.*

Proof – You can also find the proof in Kallenberg's book [Kal02], lemma 24.3, p. 473. ▷

Denote by $(X_t)_{t\geqslant 0}$ the coordinate process on $\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d)$ and by $\mathbb{P}$ Wiener measure. Given any starting point $x \in B$ and any set $U \subset B$, denote by $S_U^x = \inf\{t > 0\,;\, x + X_t \notin U\}$ the exit time from U by the Brownian motion starting from x. These random times are almost-surely finite for bounded sets U (why?); note that the distribution of $X_{S^x_{B(x,r)}}$ is uniform distribution over the sphere.

Suppose the boundary condition f is non-negative and set[74]

$$H_B f(x) = \mathbb{E}\big[f(x + X_{S_B^x})\big].$$

[74]Justify that this function is measurable with respect to x.

Then by the strong Markov property, we have for any ball $B(x,r) \subset B$

$$\begin{aligned} H_B f(x) &= \mathbb{E}\Big[\mathbb{E}\big[f(x + X_{S_B^x}) \,\big|\, X_{S^x_{B(x,r)}}\big]\Big] \\ &= \int \mathbb{E}\big[f(y + X_{S_B^x})\big]\, \sigma_{x,r}(dy) = \int H_B f(y)\, \sigma_{x,r}(dy), \end{aligned} \tag{10.5}$$

so $H_B f$ is harmonic on B if, for instance, f is bounded. This simple remark gives us a good candidate for a solution to Dirichlet problem; yet something remains to be clarified as the following shows. The only harmonic functions on $B(0,1)\backslash\{0\}$ are of the form $\alpha \log |x| + \beta$ for some constant α, β (prove this): they either explode to ∞ near 0 or are constant. This fact is a hint that not only the boundary condition is important in Dirichlet problem, but also the shape of ∂B influences the issue. We give here a condition around each point $z \in \partial B$ which prevents explosion and ensures that $H_B f$ is continuous at z.[75]

Definition **103.** *A* ***boundary point*** *$z \in \partial B$ is said to be* ***regular*** *if$_{def}$ the Brownian motion starting from z almost-surely exits B immediately:*

$$\mathbb{E}[S_B^z] = 0.$$

It is said to be ***irregular*** *otherwise.*

In the above example the point 0 is irregular. By Blumenthal's $0-1$ law, z is irregular iff $\mathbb{P}(S_B^z > 0) = 1$. Also, from exercise 31, the point z is regular if it belongs to the boundary of a cone contained in the complementary of B.

Theorem **104.** *Let $B \subset \mathbb{R}^d$ be a bounded open set and $f : \partial B \to \mathbb{R}$ be a bounded Borel function. Suppose $z \in \partial B$ is regular and f is continuous at point z, then $H_B f$ is continuous at point z:*

$$\lim_{x \to z,\, x \in B} H_B f(x) = f(z).$$

Corollary **105.** *$H_B f$ solves the Dirichlet problem if B is bounded, any point of ∂B is regular and the boundary condition f continuous.*

You will prove in exercise that $H_B f$ is the unique solution to Dirichlet problem under these conditions. The proof of theorem 104 essentially rests on the following fact.

Lemma **106.** *The map $x \in \mathbb{R}^d \mapsto \mathbb{E}[S_B^x]$ is upper semi-continuous.*

Proof – Let us recall that these functions are decreasing pointwise limits of continuous functions and that they are characterized by the inequalities

$$\forall\, x \in \mathbb{R}^d, \quad \overline{\lim_{y \to x}}\, f(y) \leqslant f(x).$$

Check first the integrability of S_B^x. Choosing $R > 0$ being enough for B to be included in $B(0,R)$, the exit time S_B^x is no greater than the hitting time of the levels $\pm R$ by the first co-ordinate of X (a real-valued Brownian motion), so is integrable. For the same reason, $S_B^{x,\varepsilon} := \inf\{t\,;\, \varepsilon < t,\, x + X_t \notin B\}$ is integrable. These decreasing approximations of S_B^x converge almost-surely to S_B^x as ε decreases to 0, so we have by monotone convergence

$$\mathbb{E}[S_B^x] = \lim_{\varepsilon \downarrow 0} \downarrow \mathbb{E}[S_B^{x,\varepsilon}].$$

[75] The remainder of this section is essentially taken from K.L. Chung's excellent little book [Chu02] on Brownian motion.

But as the strong Markov property enables us to write $\mathbb{E}[S_B^{x,\varepsilon}] = \mathbb{E}\big[g(x+X_\varepsilon)\big]$, where $g(y) := \mathbb{E}[S_B^y]$ is bounded (can you see why?), $\mathbb{E}[S_B^{x,\varepsilon}]$ appears as a (smooth and so) continuous function of x. So

$$\overline{\lim_{y\to x}}\, \mathbb{E}[S_B^y] \leqslant \overline{\lim_{y\to x}}\, \mathbb{E}[S_B^{y,\varepsilon}] = \lim_{y\to x} \mathbb{E}[S_B^{y,\varepsilon}] = \mathbb{E}[S_B^{x,\varepsilon}].$$

It remains to send ε to 0 to conclude. ▷

The proof of theorem 104 is now easy.

PROOF – Let $z \in \partial B$ be a regular point. From lemma 106 we have

$$\mathbb{E}[S_B^x] \underset{x\to z,\, x\in B}{\longrightarrow} 0,$$

i.e. S_B^x converges in $\mathbb{L}^1(\mathbb{P})$ to 0. So one can extract from any sequence $\{x_n\}_{n\geqslant 0}$ converging to z a subsequence $\{x_{n(p)}\}_{p\geqslant 0}$ such that the exit times $S_B^{x_{n(p)}}$ converge almost-surely to 0. The continuity of Brownian motion ensures us that the exit points $x_{n(p)} + X_{S_B^{x_{n(p)}}}$ converge almost-surely to $z + X_0 = z$. As a consequence, if f is bounded on ∂B and continuous at z, dominated convergence justifies the convergence

$$\mathbb{E}\big[f\big(x_{n(p)} + X_{S_B^{x_{n(p)}}}\big)\big] \underset{p\to+\infty}{\longrightarrow} f(z),$$

that is

$$H_B f(x_{n(p)}) \underset{p\to+\infty}{\longrightarrow} f(z).$$

As the limit value does not depend upon the subsequence, $H_B f(x)$ converges to $f(z)$ as x tends to z. ▷

11. Lévy processes

We study in this section models of random phenomena whose properties are insensitive to time shifts. As will become clear in section 12, they are the basic objects out of which all reasonnable martingales can be described.

The definition of Lévy processes is given section 11.1, whose main result is a kind of static description of such processes through the analytic description of their Fourier transform at a fixed time. We address the construction problem of such processes in section 11.2, where we construct a general Lévy process as a limit of the sum of a Brownian motion with a drift and of (compensated) Poisson jump processes.

11.1. **Basics.**

DEFINITION **107.** *By a (real-valued)* ***Lévy process*** *we shall understand a real-valued càdlàg process starting from* 0 *and with stationary independent increments.*

Given time $0 < t_1 < \cdots < t_n$ the increments $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ are independent and the law of $X_{t_i} - X_{t_{i-1}}$ depends on the time increment $t_i - t_{i-1}$.

A Brownian motion with constant drift is a Lévy process, so are Poisson processes; we recall their definition. These are continuous time Markov processes whose dynamics is characterized by two parameters: a finite positive constant λ and a probability measure $J(\cdot)$ on $\mathbb{R}$. Denote by $(S_n)_{n\geqslant 0}$ an iid sequence of exponential random variables with parameter λ, and by $(J_n)_{n\geqslant 1}$ an iid sequence of random variables with common distribution J. The process X starts almost-surely from 0 and is constant on the interval $[0, S_1)$; it has a jump J_1 at time S_1 : $X_{S_1} = J_1$. Then it remains constant on the interval $[S_1, S_1 + S_2)$ and has a jump J_2 at time $S_1 + S_2$: $X_{S_1+S_2} = J_1 + J_2$; and so on. It is not difficult, using the memoryless property of the exponentials, to prove that this process is a Lévy process – you are asked to prove that fact in exercise. Surprisingly, Brownian motion and Poisson processes are all we need to describe the most general Lévy process, as theorem 113 will make it clear. Note that Poisson processes have that name as the number of jumps they make in a time interval of lenght t is a Poisson random variable with parameter λt. Can you prove this fact?

We start our study of Lévy processes looking at their fixed time distributions.

Lemma **108.** *Denote by $\varphi_t(\lambda)$ the characteristic function of X_t : $\varphi_t(\lambda) = \mathbb{E}\big[e^{i\lambda X_t}\big], \lambda \in \mathbb{R}, t \geqslant 0$. There exists a continuous complex-valued function $g(\lambda)$ such that $\varphi_t(\lambda) = e^{tg(\lambda)}$. This function is called the* ***characteristic exponent of the Lévy process****.*

The function g characterizes completely the finite dimensional laws of X. Given $0 = t_0 < t_1 < \cdots < t_n$, lemma 108 and the independence of increments of X imply

$$\begin{aligned}
\mathbb{E}\Big[\exp\big(i\lambda_1 X_{t_1} + \cdots + i\lambda_n X_{t_n}\big)\Big] &= \mathbb{E}\Big[e^{i\left(\sum_{\ell=1}^n \lambda_\ell\right)X_{t_1} + i\left(\sum_{\ell=2}^n \lambda_\ell\right)(X_{t_2}-X_{t_1}) + \cdots + i\lambda_n(X_{t_n}-X_{t_{n-1}})}\Big] \\
&= \prod_{k=1}^n \mathbb{E}\big[e^{i\left(\sum_{\ell=k}^n \lambda_\ell\right)(X_{t_k}-X_{t_{k-1}})}\big] \\
&= \exp\left(\sum_{k=1}^n (t_k - t_{k-1})\, g\Big(\sum_{\ell=k}^n \lambda_\ell\Big)\right)
\end{aligned}$$

Proof – Note first that since X is càdlàg and has stationary independent increments we have

$$\mathbb{P}\big(|X_t - X_s| \geqslant \epsilon\big) = \mathbb{P}(X_{|t-s|} \geqslant \epsilon) \xrightarrow[s \to t]{} 0$$

for all $\epsilon > 0$. As

$$\begin{aligned}
\big|\varphi_t(\lambda) - \varphi_s(\lambda')\big| &\leqslant \mathbb{E}\Big[\big|e^{i\lambda(X_t - X_s)} - 1\big|\Big] + \mathbb{E}\Big[\big|e^{i(\lambda-\lambda')X_s} - 1\big|\Big] \\
&\leqslant \sup_{|x| \leqslant \epsilon} \big|e^{i\lambda x} - 1\big| + 2\,\mathbb{P}\big(|X_t - X_s| \geqslant \epsilon\big) + \mathbb{E}\Big[\big|e^{i(\lambda-\lambda')X_s} - 1\big|\Big]
\end{aligned}$$

it follows (by dominated convergence) that φ is a continuous function of (t,λ). As a consequence, $\varphi_t(\lambda) \neq 0$ for t small enough, since $\varphi_0(\lambda) = 0$. Using the independence and stationarity of the increments, it follows that we have for all $t \geqslant 0, \lambda \in \mathbb{R}$

$$\varphi_t(\lambda) = \mathbb{E}\big[e^{i\lambda X_t}\big] = \prod_{j=1}^N \mathbb{E}\Big[e^{i\lambda\left(X_{\frac{j}{N}t} - X_{\frac{j-1}{N}t}\right)}\Big] = \big\{\varphi_{\frac{t}{N}}(\lambda)\big\}^N \neq 0,$$

provided N is big enough. We can thus write $\varphi_t(\lambda) = e^{a_t(\lambda) + ib_t(\lambda)}$, where a and b are continuous functions of (t,λ) and $a_0(\lambda) = b_0(\lambda) = b_t(0) = 0$. Using again the stationarity and independence of the increments, we see that $\varphi_{s+t}(\lambda) = \varphi_s(\lambda)\varphi_t(\lambda)$; as a and b are continuous, this implies that they are both linear functions of t. ▷

The general form of g was found by Lévy and Khinchin.

Theorem **109** (Lévy-Khinchin)**.** *Given $\lambda \in \mathbb{R}$ and $x \in \mathbb{R}$, set*

$$f(\lambda, x) = \big(e^{i\lambda x} - 1 - i\lambda \sin x\big)\frac{1+x^2}{x^2}$$

for $x \neq 0$ and $f(\lambda, 0) = -\frac{\lambda^2}{2}$; this formula defines a continuous function. There exists a finite non-negative Borel measure μ on $\mathbb{R}$ and a constant $b \in \mathbb{R}$ such that

$$g(\lambda) = \int f(\lambda, x)\mu(dx) + ib\lambda. \tag{11.1}$$

PROOF – We start[76] with the identity $-g(\lambda) = \lim_{t \searrow 0} \frac{1-\varphi_t(\lambda)}{t}$, $g(0) = 0$, where the limit is uniform for λ in a compact set. Denoting by ν_t the law of X_t and taking $t = \frac{1}{n}$ above, we get

$$-g(\lambda) = \lim_{n \to +\infty} \int \left(1 - e^{i\lambda x}\right) n\nu_{\frac{1}{n}}(dx), \tag{11.2}$$

and for $h > 0$ ([77])

$$-\frac{1}{2h} \int_{-h}^{h} g(\lambda)\, d\lambda = \lim_{n \to +\infty} \int \left(1 - \frac{\sin hx}{hx}\right) n\nu_{\frac{1}{n}}(dx). \tag{11.3}$$

As the function $\left(1 - e^{i\lambda x}\right)$ is continuous and bounded it is fair to try and use some compactness argument in the set of measures to write equation (11.2) under the form $\int \left(1 - e^{i\lambda x}\right) \mu(dx)$, where μ is a weak limit of the sequence $n\nu_{\frac{1}{n}}$. But as this measure has mass increasing n we need to be careful. Would its mass be bounded away from 0 and ∞, we could write it as $a_n\rho_n$, with $a_n > 0$ and ρ_n a probability measure. The tightness of the sequence of *finite measures* $(a_n\rho_n)_{n \geqslant 0}$ would then equivalent to the tightness of the sequence of *probability measures* $(\rho_n)_{n \geqslant 0}$. Choosing a subsequence along which both $(a_n)_{n \geqslant 0}$ and $(\rho_n)_{n \geqslant 0}$ converge would provides a cluster point for the sequence of measures $(a_n\rho_n)_{n \geqslant 0}$ for the weak topology. Noting that there exists a positive constant C such that $\frac{y^2}{1+y^2} \leqslant C(1 - \frac{\sin y}{y})$ for all y, equation (11.3), with $h = 1$, tells us that the sequence of measures $\left(\frac{x^2}{1+x^2}\, n\nu_{\frac{1}{n}}(dx)\right)_{n \geqslant 1}$ has mass uniformly bounded above. If the corresponding a_n converge to 0 the measures converge weakly to 0. Elsewhere, we see the tightness of this family of measures noting that since $1 - \frac{\sin hx}{hx} \geqslant 0$ is no less than $\frac{1}{2}$ for $|hx| \geqslant 2$, all the integrals $\int_{|x| \geqslant \frac{2}{h}} n\nu_{\frac{1}{n}}(dx)$ are uniformly small provided h is small enough; this is a fortiori the case for the integrals $\int_{|x| \geqslant \frac{2}{h}} \frac{x^2}{1+x^2}\, n\nu_{\frac{1}{n}}(dx)$. Choose a subsequence for which the measures $\frac{x^2}{1+x^2}\, n(p)\nu_{\frac{1}{n(p)}}(dx)$ converge weakly, say to μ. Our intuition about how to turn the limit (11.2) into a proper integral thus takes the following form.

$$\begin{aligned}
-g(\lambda) &= \lim_{n+\infty} \int \left(1 - e^{i\lambda x}\right) n\nu_{\frac{1}{n}}(dx) \\
&= \lim_{p+\infty} \left(\int \left(1 - e^{i\lambda x} + i\lambda \sin x\right) n(p)\nu_{\frac{1}{n(p)}}(dx) + i\lambda \int (\sin x)\, n(p)\nu_{\frac{1}{n(p)}}(dx) \right) \\
&= \lim_{p+\infty} \left(-\int f(\lambda, x) \frac{x^2}{1+x^2}\, n(p)\nu_{\frac{1}{n(p)}}(dx) + i\lambda \int (\sin x)\, n(p)\nu_{\frac{1}{n(p)}}(dx) \right)
\end{aligned} \tag{11.4}$$

As $f(\lambda, x)$, is a bounded continuous function of x, the first term converges to $-\int f(\lambda, x)\mu(dx)$; it follows that the integrals $\int (\sin x)\, n(p)\nu_{\frac{1}{n(p)}}(dx)$ have a limit as p goes to infinity, which defines the constant $-b$. ▷

[76]This proof of Lévy-Khinchin's representation theorem is essentially taken from N.V. Krylov's book [Kry02].

[77]You can interchange the integral with respect to λ and the limit as the terms $\int \left(1 - e^{i\lambda x}\right) n\nu_{\frac{1}{n}}(dx)$ are uniformly bounded with respect to $n \geqslant 1$ and λ in a bounded set.

Isolating a possible Dirac mass at 0 in μ, we can write

$$g(\lambda) = \int \big(e^{i\lambda x} - 1 - \lambda \sin x\big)\Lambda(dx) - \frac{\lambda^2\sigma^2}{2} + ib\lambda, \tag{11.5}$$

with $\mu'(dx) = \frac{1+x^2}{x^2}\,\mu(dx), \sigma^2 = \mu'(0)$ and $\Lambda = \mu' - \mu'(0)\,\delta_0$ has no mass on $\{0\}$. The case $\Lambda = 0$ corresponds to a Brownian motion with drift b and variance σ^2([78]).

Theorem **110** (Uniqueness)**.** *Lévy-Khinchin's decomposition for g is unique.*

Proof – It suffices to note that since we have for any $\lambda \in \mathbb{R}$ and $h > 0$

$$g(\lambda) - \frac{g(\lambda + h) + g(\lambda - h)}{2} = \int e^{i\lambda x}\frac{1 - \cos(hx)}{x^2}(1 + x^2)\mu(dx),$$

the measures $\mu_h := \frac{1-\cos(hx)}{x^2}(1 + x^2)\mu(dx)$ are uniquely determined by g, as their Fourier transforms are given by the above formula. But we have for any bounded Borel set $A \subset \big[-\frac{1}{h}, \frac{1}{h}\big]$,

$$\mu(A) = \int \mathbf{1}_A \frac{x^2}{1 - \cos(hx)}(1 + x^2)^{-1}\mu_h(dx);$$

the result follows. ▷

A triple $(b, \sigma^2\,; \Lambda)$, where Λ is a non-negative measure on $\mathbb{R}^*$ such that

$$\int (1 \wedge x^2)\Lambda(dx) < \infty$$

is called a ***Lévy triple***. Lévy-Khinchin formula gives a static description of a Lévy process in terms of a Lévy triple; it is not clear at all whether or not there corresponds a Lévy process to each such triple. This is indeed the case, and the proof given below will reveal the dynamical content of the Lévy-Khinchin formula. Theorem 113 below proves that any Lévy process has a modification which is the limit of a sum of independent processes

$$bt + \sigma B_t + P_t^{(0)} + \sum_{k=1}^{n} \widetilde{P}_t^{(k)},$$

where $bt+\sigma B_t$ is a Brownian motion with drift b and variance σ^2, the process $P^{(0)}$ is a Poisson process with intensity $\Lambda\big(\{|x| \geqslant 1\}\big)$ and jump measure $J^{(0)} = \Lambda\big(\{|x| \geqslant 1\}\big)^{-1}\mathbf{1}_{|x|\geqslant 1}\Lambda$, and the processes $\widetilde{P}^{(k)}$ Poisson processes with intensity $\Lambda\big(\{\frac{1}{k+1} \leqslant |x| < \frac{1}{k}\}\big)$, jump measure $J^{(k)} = \Lambda\big(\{\frac{1}{k+1} \leqslant |x| < \frac{1}{k}\}\big)^{-1}\mathbf{1}_{\{\frac{1}{k+1}\leqslant|x|<\frac{1}{k}\}}\Lambda$, *and a drift* $-\int x\mathbf{1}_{\{\frac{1}{k+1}\leqslant|x|<\frac{1}{k}\}}\Lambda(dx)$. We denote by $P^{(k)}$ the Poisson process without drift.

It will clarify the construction below to rewrite the characteristic exponent g of a Lévy process under the form

$$g(\lambda) = -\frac{\sigma^2\lambda^2}{2} + ib'\lambda + \int \big(e^{i\lambda x} - 1 - i\lambda x\mathbf{1}_{|x|<1}\big)\,\Lambda(dx),$$

replacing the former b by[79] $b' = b + \int \big(x\mathbf{1}_{|x|<1} - \sin x\big)\,\Lambda(dx)$. *We shall write b instead of b' below.* Note that we have

[78]Note that the sin function appearing in the above formula for the exponent $g(\lambda)$ has nothing canonical; it could equally well be replaced by any bounded continuous function which is equivalent to x near 0. This would change b accordingly.

[79]This integral converges as $\int (x^2 \wedge 1)\,\Lambda(dx) < \infty$.

$$\int \left(e^{i\lambda x} - 1 - i\lambda x \mathbf{1}_{|x|<1}\right) \Lambda(dx) = \int \left(e^{i\lambda x} - 1\right) \mathbf{1}_{|x|\geqslant 1} \Lambda(dx) + \int \left(e^{i\lambda x} - 1 - i\lambda x\right) \mathbf{1}_{|x|<1} \Lambda(dx).$$

11.2. **Construction of Lévy processes.** Denote by $\left(\mathcal{F}_t\right)_{t\geqslant 0}$ the filtration generated by $B, P^{(0)}$ and all the $\widetilde{P}^{(k)}$.

LEMMA **111.** (1) *The Lévy process $P^{(0)}$ has characteristic exponent $\int \left(e^{i\lambda x} - 1\right) \mathbf{1}_{|x|\geqslant 1} \Lambda(dx)$.*

(2) *Each Lévy process $\widetilde{P}^{(k)}$ is a càdlàg $\left(\mathcal{F}_t\right)_{t\geqslant 0}$-martingale with characteristic exponent*

$$\int \left(e^{i\lambda x} - 1 - i\lambda x\right) \mathbf{1}_{\frac{1}{k+1}\leqslant |x| < \frac{1}{k}} \Lambda(dx).$$

(3) *The process $\left(\widetilde{P}_t^{(k)}\right)^2 - t \int x^2 \mathbf{1}_{\frac{1}{k+1}\leqslant |x| < \frac{1}{k}} \Lambda(dx)$ is a càdlàg $\left(\mathcal{F}_t\right)_{t\geqslant 0}$-martingale.*

It follows from this lemma that the characteristic exponent of our approximating Lévy process $\sigma B_t + bt + P_t^{(0)} + \sum_{k=1}^n \widetilde{P}_t^{(k)}$ is

$$-\frac{\sigma^2\lambda^2}{2} + ib\lambda + \int \left(e^{i\lambda x} - 1\right) \mathbf{1}_{|x|\geqslant 1} \Lambda(dx) + \sum_{k=1}^n \int \left(e^{i\lambda x} - 1 - i\lambda x\right) \mathbf{1}_{\frac{1}{k+1}\leqslant |x| < \frac{1}{k}} \Lambda(dx),$$

that is

$$-\frac{\sigma^2\lambda^2}{2} + ib\lambda + \int \left(e^{i\lambda x} - 1 - i\lambda x \mathbf{1}_{\frac{1}{n+1}\leqslant |x| < 1}\right) \Lambda(dx). \tag{11.6}$$

PROOF – (1) The computations of the characteristic functions of $P_t^{(0)}$ and $\widetilde{P}_t^{(k)}$ are done in the same way and use the following elementary fact. The distribution of the number N_t of jumps by time t of a Poisson process of intensity ρ is a Poisson random variable with parameter ρt. Writing ρ for $\Lambda\big(\{|x| \geqslant 1\}\big)$ we thus have

$$\begin{aligned}\mathbb{E}\big[e^{i\lambda P_t^{(0)}}\big] &= \sum_{n\geqslant 0} \mathbb{E}\big[e^{i\lambda P_t^{(0)}} \big| \sigma(N_t)\big]\, \mathbb{P}(N_t = n) = \sum_{n\geqslant 0} \Big(\frac{1}{\rho} \int e^{i\lambda x} \mathbf{1}_{|x|\geqslant 1} \Lambda(dx)\Big)^n e^{-\rho} \frac{\rho^n}{n!} \\ &= \exp\Big(\int \big(e^{i\lambda x} - 1\big) \mathbf{1}_{|x|\geqslant 1} \Lambda(dx)\Big).\end{aligned}$$

(2) As the non-drifted process $P^{(k)}$ is a Poisson process (hence a Lévy process) it has independent increments. It follows that

$$\mathbb{E}\big[P_t^{(k)} - P_s^{(k)} \big| \mathcal{F}_s\big] = \mathbb{E}\big[P_t^{(k)} - P_s^{(k)}\big] = (t-s) \int x \mathbf{1}_{\frac{1}{k+1}\leqslant |x| < \frac{1}{k}} \Lambda(dx),$$

which proves that $\widetilde{P}^{(k)}$ is an $\left(\mathcal{F}_t\right)_{t\geqslant 0}$-martingale.

(3) I leave you to justify the fact that $\widetilde{P}_t^{(k)} \in \mathbb{L}^2$. As the process $\widetilde{P}^{(k)}$ has independent centered increments it suffices to see that $\mathbb{E}\Big[\big(\widetilde{P}_t^{(k)}\big)^2\Big] = t \int x^2 \mathbf{1}_{\frac{1}{k+1}\leqslant |x| < \frac{1}{k}} \Lambda(dx)$ for each $t > 0$. Writing $\widetilde{P}_t^{(k)} = P_t^{(k)} - bt$, with $b = \int x \mathbf{1}_{\frac{1}{k+1}\leqslant |x| < \frac{1}{k}} \Lambda(dx)$, it amounts to proving that $\mathbb{E}\Big[\big(P_t^{(k)}\big)^2\Big] = b^2t^2 + t \int x^2 \mathbf{1}_{\frac{1}{k+1}\leqslant |x| < \frac{1}{k}} \Lambda(dx)$. This is done by a direct computation, conditionning on the number of jumps of $P^{(k)}$ by time t, which is a Poisson random variable with parameter $t\Lambda\big(\big\{\frac{1}{k+1} \leqslant |x| < \frac{1}{k}\big\}\big)$. ▷

The next proposition provides a good functional framework where to take the limit of our approximations.

PROPOSITION **112.** *Denote by $\mathcal{H}$ the space of $\big(\mathcal{F}_t\big)_{t\geqslant 0}$-martingales bounded in $\mathbb{L}^2$ and define on $\mathcal{H}$ the metric $d(X,Y) = \Big(\mathbb{E}\big[\sup_{0\leqslant s\leqslant 1}\big|X_s - Y_s\big|^2\big]\Big)^{\frac{1}{2}}$. Then the metric space $(\mathcal{H}, d)$ is complete.*

PROOF – Let $(Y^n)_{n\geqslant 0}$ be a Cauchy sequence of elements of $\mathcal{H}$. Each $(Y^n_t)_{n\geqslant 0}$ is then a Cauchy sequence in $\mathbb{L}^2$, so converges to some random variable $Y_t \in \mathbb{L}^2$. We get the martingale property of Y by passing to the limit in the corresponding identity for Y^n. It suffices then to apply Doob's $\mathbb{L}^2$-inequality to get

$$\mathbb{E}\big[\sup_{0\leqslant s\leqslant 1}|Y^n_s - Y_s|^2\big] \leqslant 2\sup_{0\leqslant s\leqslant 1}\mathbb{E}\big[|Y^n_s - Y_s|^2\big] \leqslant 2\,\mathbb{E}\big[|Y^n_1 - Y_1|\big] \to 0,$$

which proves that Y^n converges to Y in $(\mathcal{H}, d)$. ▷

THEOREM **113** (Construction of Lévy processes)**.** *To any Lévy triple there corresponds a Lévy process with characteristic exponent given by formula* (11.5).

PROOF – Set $Y^{(n)}_t = \sum_{k=1}^n \widetilde{P}^{(k)}_t$, for $t \in [0,1]$. Using the fact that $\mathbb{E}\big[P^{(k)}_1 P^{(\ell)}_1\big] = 0$ for $k \neq \ell$, and Doob's $\mathbb{L}^2$-inequality we have, for $m \geqslant n$,

$$\begin{aligned}\mathbb{E}\Big[\sup_{0\leqslant s\leqslant 1}\big|Y^m_s - Y^n_s\big|^2\Big] &= \mathbb{E}\Big[\sup_{0\leqslant s\leqslant 1}\Big|\sum_{k=n+1}^m \widetilde{P}^{(k)}_s\Big|^2\Big] \\ &\leqslant 4\,\mathbb{E}\Big[\Big|\sum_{k=n+1}^m \widetilde{P}^{(k)}_1\Big|^2\Big] = 4\sum_{k=n+1}^m \mathbb{E}\Big[\big|\widetilde{P}^{(k)}_1\big|^2\Big] \\ &\leqslant 4\int x^2\,\mathbf{1}_{\frac{1}{m+1}\leqslant|x|<\frac{1}{n+1}}\Lambda(dx).\end{aligned}$$

The last inequality comes from point (3) of lemma 111. Since the integral $\int x^2\mathbf{1}_{|x|\leqslant 1}\Lambda(dx)$ converges, the above quantity is arbitrarily small provided m and n are big enough. This proves that $Y^{(n)}$ is a Cauchy sequence in the complete space $(\mathcal{H}, d)$; denote by Y its limit. I leave you to check that Y has independent stationary increments, simply by passing to the limit in the corresponding identities. Also, as $\sup_{0\leqslant s\leqslant 1}\big|Y_s - Y^{(n)}_s\big|$ converges in $\mathbb{L}^2$ to 0, a subsequence converges almost-surely to 0; this makes the process Y appear as a uniform limit of càdlàg paths, so Y is itself càdlàg, and hence is a Lévy process.

Recall the expression of the characteristic exponent of $X^{(n)}_t = bt + \sigma B_t + P^{(0)}_t + Y^{(n)}_t$ given in equation (11.6); set $X_t = bt + \sigma B_t + P^{(0)}_t + Y_t$. Using the almost-sure convergence of a subsequence of $Y^{(n)}_t$, the estimate $\big|\big(e^{i\lambda x} - 1 - \lambda x\big)\mathbf{1}_{|x|<1}\big| \leqslant Cx^2\mathbf{1}_{|x|<1}$ (for some constant $C > 0$) and dominated convergence[80], we obtain

$$\mathbb{E}\big[e^{i\lambda X_t}\big] = \exp\Big(-\frac{\sigma^2\lambda^2}{2}t + i\lambda bt + t\int\big(e^{i\lambda t} - 1 - \lambda y\mathbf{1}_{|x|<1}\big)\Lambda(dx)\Big),$$

this proves that the Lévy process X has Lévy triple $(b, \sigma\,; \Lambda)$. ▷

12. (...) AND MARTINGALES

12.1. **Representation of continuous martingales.** We prove in this section that any continuous martingale can be seen as a time change of a Brownian motion. This will happen to be a beautiful application of the strong Markov property. To set notations, write $(\Omega_0, \mathcal{F}_0)$ for $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$, equipped with its Borel σ-algebra, X for the coordinate

[80] We have $\int x^2\,\mathbf{1}_{|x|\leqslant 1}\Lambda(dx) < \infty$.

process and $\mathbb{P}_0$ for Wiener measure on $(\Omega_0, \mathcal{F}_0)$. Write also $\mathcal{R}$ for the Borel σ-algebra of $\mathbb{R}_+$.

THEOREM **114** (Brownian motion as the father of all continuous martingales)**.** *Let $(M_t)_{t\geqslant 0}$ be a continuous martingale defined on some probability space $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$ and starting from 0. Then one can define on $\big(\Omega_0\times\Omega, \mathcal{F}_0\otimes\mathcal{F}, \mathbb{P}_0\otimes\mathbb{P}\big)$ a Brownian motion B and a measurable time change $\phi : (t, \omega_0, \omega) \mapsto \phi_t(\omega_0, \omega) \in \mathbb{R}_+$ on $\big(\mathbb{R}_+ \times \Omega_0 \times \Omega, \mathcal{R} \otimes \mathcal{F}' \otimes \mathcal{F}\big)$ such that for each $t \geqslant 0$ we have $\mathbb{P}_0 \otimes \mathbb{P}$-almost-surely*

$$M_t(\omega) = B_{\phi_t(\omega_0,\omega)}(\omega_0, \omega).$$

In this sense, M appears as a random time-change of a Brownian motion.

It will be useful to introduce the following notations where ϵ is any positive constant, and where we use the convention $\inf \emptyset = +\infty$. Given a continuous function $x = (x_t)_{t\in[0,T]}$ defined on some interval $[0, T]$ and starting from 0, define by induction

$$S^x_1(\epsilon) = \inf\{t \in [0, T] \,;\, x_t = \epsilon\}, \quad S^x_{n+1}(\epsilon) = \inf\{t \in \big[S^x_n(\epsilon), T\big] \,;\, \big|x_t - x_{S^x_n(\epsilon)}\big| = \epsilon\}.$$

Denote by $N^x(\epsilon)$ the biggest $n \geqslant 1$ for which $S^x_n(\epsilon) < \infty$ and set $\mathcal{T}^x(\epsilon) = \big\{S^x_n(\epsilon) \,;\, n = 1..N^x(\epsilon)\big\}$. Note the inclusions

$$\mathcal{T}^x\Big(\frac{\epsilon}{2}\Big) \subset \mathcal{T}^x(\epsilon). \tag{12.1}$$

PROOF – *1)* The proof rests on the following simple observation. Let $x = (x_t)_{t\geqslant 0}$ be a continuous real-valued path for which *all the $S_n(\epsilon)$ are finite, whatever $n \geqslant 1$ and $\epsilon > 0$*. Let $y = (y_t)_{t\in[0,T]}$ be a non-constant continuous function defined on some interval $[0, T]$; provided $\epsilon_0 > 0$ is small enough, the collection of times $\mathcal{T}^y(\epsilon)$ is non-empty for $0 < \epsilon \leqslant \epsilon_0$.

LEMMA **115.** *One can construct a function x^y and a continuous non-decreasing time change ϕ from $[0, T]$ to $\big[0, \phi(T)\big]$ such that*

$$x^y(s) = y_t, \quad \textit{if } s = \phi(t),$$

and $\phi\big(S^y_n(\epsilon)\big) = S^x_n(\epsilon)$, for ϵ small enough and $n \in \big\{1, .., N^y(\epsilon)\big\}$.

PROOF – Taking ϵ of the form 2^{-p}, there exists a unique continuous piecewise linear map ϕ_p for $[0, T]$ to $\mathbb{R}_+$ such that
- $\phi_p(0) = 0$, $\phi_p\big(S^y_n(2^{-p})\big) = S^x_n(2^{-p})$, and
- $\phi_p(t) = c + t$, for some constant c and $t \geqslant S^y_{N^y(2^{-p})}(2^{-p})$.

We define a continuous function x^p on $\big[0, \phi_p(T)\big]$ setting

$$x^p(s) = y_t, \quad \text{if } s = \phi_p(t).$$

Note that due to the inclusion (12.1), for each p_0 and $n \in \{1, .., N^y(2^{-p_0})\}$, the sequence of times $\big\{\phi_p\big(S^y_n(2^{-p_0})\big)\big\}_{p\geqslant 0}$ is constant for $p \geqslant p_0$. Using the continuity of x and y, it is then a simple thing to prove that the sequence of time changes $(\phi_p)_{p\geqslant 0}$ converges uniformly to some non-decreasing time change $\phi : [0, T] \to \big[0, \phi(T)\big]$. Check that the function x^y defined on $\big[0, \phi(T)\big]$ by the formula

$$x^y(s) = y_t, \quad \text{if } s = \phi(t)$$

has the desired properties. ⊙

2) Recall we denote by X the coordinate process on Ω_0 and that it is a Brownian motion under Wiener measure $\mathbb{P}_0$. So almost-all paths $X(\omega_0)$ have all their $S_n(\epsilon)$ finite. Applying lemma 115 to $x = X(\omega_0)$ and $y = \big(M_t(\omega)\big)_{t\in[0,T]}$, we get a time change $\phi : [0,T] \rightarrow \big[0,\phi(T)\big]$ and a path $\big(X^M_s\big)_{s\in[0,\phi(T)]}$; this random path is defined on the probability space $\big(\Omega_0\times\Omega, \mathcal{F}_0\otimes\mathcal{F}, \mathbb{P}_0\otimes\mathbb{P}\big)$.

LEMMA **116.** *The process $\big(X^M_s\big)_{s\in[0,\phi(T)]}$ is a Brownian motion (defined on a random interval).*

PROOF – Notice first that, for 2^{-p} small enough and $n \in \big\{0,..,N^M(2^{-p})\big\}$, we have by the martingale property of M

$$\mathbb{P}\big(M_{S^M_{n+1}(2^{-p})} = M_{S^M_n(2^{-p})} \pm \epsilon \big| \mathcal{F}_{S^M_n(2^{-p})}\big) = \frac{1}{2}.$$

Denote the above $\{\pm 1\}$-valued random variable by ϵ^p_n and define a new continuous path $\widehat{X}^p$ requiring that

$$\widehat{X}^p_\cdot - \widehat{X}^p_{S^X_n(2^{-p})} = \epsilon^p_n\big(X_\cdot - X^p_{S^X_n(2^{-p})}\big)$$

on the interval $\big[S^X_n(2^{-p}), S^X_{n+1}(2^{-p})\big]$. The process $\widehat{X}^p$ is by the strong Markov property a Brownian motion. Note that $\widehat{X}^p_s = X^M_s$ at all times s of the form $S^X_n(2^{-p})$. It follows from this fact that $\mathbb{P}_0\otimes\mathbb{P}$-almost-surely the functions $\widehat{X}^p$ converge uniformly to X^M on the interval $\big[0,\phi(T)\big]$. As each of them is a Brownian motion, the process X^M is also a Brownian motion. ⊙

Lemmas 115 and 116 together prove the representation theorem, up to the measurability statements. These can be proved examining the above construction, and are not really important for us. ▷

You will see an improved (and more sophisticated) version of that result in the course on stochastic calculus: *there exists an $(\mathcal{F}^M_t)_{t\geqslant 0}$-adapted random time change $\langle M\rangle_t$ and a Brownian motion B (with respect to some other filtration) such that $M_t = B_{\langle M\rangle_t}$ for all $t \geqslant 0$.*

12.2. **Representation of general martingales.** Although getting a proper description of the structure of the most general martingales would require the introduction of new concepts, we have all the tools needed to understand this structure perfectly. In the same way as a $\mathcal{C}^1$ function from $\mathbb{R}$ to $\mathbb{R}$ is infinitesimally well-approximated by its tangent line (so well that we can recover the function from the family of its tangents: $f(t) = f(0)+\int_0^t f'(s)\,ds$), any càdlàg martingale M is infinitesimally well-approximated by a Lévy process. Roughly speaking, at each time t there exists a random Lévy triple $(0,\sigma_t^2\,;\,\Lambda_t)$, measurable with respect to $\mathcal{F}_t$, such that the martingale M is δt-close to the corresponding Lévy process over the time interval $[t,t+\delta t]$. To get the martingale property at time $t+\delta t$ we ask the measure Λ_t to be symmetric.

As we have seen, Lévy processes with Lévy triples $(b,\sigma^2\,;\,\Lambda)$ are characterized by the identity $\mathbb{E}[e^{i\lambda X_t}] = e^{g_t(\lambda)}$, where

$$g_t(\lambda) = i\lambda b t - \frac{\lambda^2\sigma^2 t}{2} + \int \big(e^{i\lambda x} - 1 - ix\mathbf{1}_{|x|\leqslant 1}\big)\, t\Lambda(dx);$$

by the independence of the increments, this holds iff $\exp(i\lambda X_t)/\exp\big(g_t(\lambda)\big)$ is a martingale. The above "infinitesimal" euristics gets a proper rephrasing in the following statement.

THEOREM **117** (cf. [JS03], Chap. II, §2)**.** *Given any càdlàg martingale* $(M_t)_{t\geqslant 0}$ *there exists an adapted process* $(\sigma_t^2)_{t\geqslant 0}$ *and an adapted random measure-valued process* $(\Lambda_t)_{t\geqslant 0}$ *such that* $(0, \sigma_t^2\,;\,\Lambda_t)$ *is a Lévy triple for all* $t \geqslant 0$*, the measures* Λ_t *are symmetric, and the process*

$$\exp(i\lambda M_t)/\exp\big(\psi_t(\lambda)\big)$$

is a martingale, where $\psi_t(\lambda) = -\frac{\lambda^2\sigma_t^2}{2} + \int \big(e^{i\lambda x} - 1 - ix\mathbf{1}_{|x|\leqslant 1}\big)\,\Lambda_t(dx)$*. There is only one such process* $(0, \sigma_t^2\,;\,\Lambda_t), t \geqslant 0$*, which is previsible.*

In short, a continuous time process $(t,\omega) \in \mathbb{R}_+ \times \Omega \mapsto Y_t(\omega)$ is said to be ***previsible*** if$_{\text{def}}$ it is measurable with respect to the σ-algebra on $\mathbb{R}_+ \times \Omega$ generated by the adapted continuous processes. Allowing general (previsible) Lévy triples $(b_t, \sigma_t^2\,;\,\Lambda_t)$ in the above description leads to the class of semi-martingales, which is the good class of processes to consider when constructing the theory of stochastic integration. You will certainly encounter it under a different costume: $(Y_t)_{t\geqslant 0}$ is a semi-martingale if$_{\text{def}}$ one can find an adapted process A with finite variation, an increasing sequence of finite stopping times T^n, and a sequence $(M^n)_{0\leqslant t\leqslant T^n}$ of closed martingales such that

$$\forall\, n \geqslant 0,\ \forall\, t \leqslant T^n, \quad Y_t = M_t^n + A_t.$$

But this is the beginning of another story...

13. Comments and exercises

References. The book [Chu02] of Kai Lai Chung will give you a nice view on Brownian motion. Rogers and Williams' book, [RW00], as always, is recommended.

You will find interesting material on Lévy processes in Krylov's book [Kry02]. You will also find in the first chapter of Sato's book [Sat99] useful and basic informations on Lévy processes.

To be written: Comments on the "Poisson random measure" approach to Lévy processes.

13.1. **Exercises.** *B denotes a real-valued or $\mathbb{R}^d$-valued Brownian motion constructed on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$; the distribution of $x + X$ is denoted by $\mathbb{P}_x$.*

1. Kolmogorov's $0-1$ law. *This exercise is the companion to exercise 8 of example sheet 2.* Let us work in $\mathbb{R}^d$. Define the *tail σ-algebra* : $\mathcal{T} = \bigcap_{t \geqslant 0} \sigma(B_{s+t}\,;\, s \geqslant 0)$. Using the inversion property of Brownian motion and Blumenthal's $0-1$ law, prove that all the events of $\mathcal{T}$ are trivial under $\mathbb{P}$.

2. Let A be an open subset of the $(d-1)$-dimensional sphere and U the cone $\{ta\,;\, a \in A, 0 \leqslant t \leqslant \varepsilon\}$ of vertex 0 (for some $\varepsilon > 0$). Prove that the hitting time $\tau_U = \inf\{t > 0\,;\, B_t \in U\}$ of U for a Brownian motion starting from 0 is almost-surely equal to 0. *This result is useful to solve Dirichlet problem by the probabilistic method in concrete cases as it ensures that all points of the boundary of an open set O are regular if any point of ∂O is the vertex of a cone contained in O^c.*

3. Using the martingale property of Brownian motion, prove that we have for any positive a, b

$$\mathbb{P}(H_{-a} < H_b) = \frac{b}{b+a} \quad \text{and} \quad \mathbb{E}[H_{-a} \wedge H_b] = ab.$$

4. Let B be a real-valued Brownian motion and $\sigma \in \mathbb{R}$.

a) Show that the process $\left(e^{\sigma B_t - \frac{\sigma^2}{2} t}\right)_{t \geqslant 0}$ is a martingale with respect to the filtation of B.

b) Deduce, by differentiating with respect to σ, that the following processes are also martingales: $\left(B_t^2 - t\right)_{t \geqslant 0}$, $\left(B_t^3 - 3tB_t\right)_{t \geqslant 0}$, $\left(B_t^4 - 6tB_t^2 + 3t^2\right)_{t \geqslant 0}$.

5. Given $c \in \mathbb{R}$, the process $B_t^c = B_t + ct$, is called the *Brownian motion with drift c*. For fixed $x > 0$ and $-a < 0 < b$, set $H_x^c = \inf\{t \geqslant 0\,;\, B_t^c = a\}$.

a) Fix $\lambda > 0$. Under which conditions on $\theta \in \mathbb{R}$ is the process $\exp\left(\theta B_t^c - \lambda t\right)$ a martingale?

b) Supposing θ chosen appropriately, deduce from **a)** that

$$\mathbb{E}\left[e^{-\lambda H_x^c}\right] = \exp\left(-x\sqrt{c^2 + 2\lambda} - c\right),$$

and so, that the distribution of H_x^c has density $\frac{x}{\sqrt{2\pi t^3}} \exp\left(-\frac{(x-ct)^2}{2t}\right)$. Is it surprising?

c) Conclude that

$$\mathbb{P}(H_x^c < \infty) = 1 \text{ if } c \geqslant 0, \text{ and } e^{-2|c|x} \text{ if } c < 0.$$

6. a) Given $a > 0$, set $H_a = \inf\{s \geqslant 0\,;\, B_s = a\}$. Prove that the distribution of H_a has a density with respect to Lebesgue measure on $\mathbb{R}_+$, equal to $\frac{a}{(2\pi t^3)^{1/2}} \exp\left(-\frac{a^2}{2t}\right)$.

b) Prove that the process of hitting times $(T_a)_{a \geqslant 0}$ has stationnary independent increments. Is it a Lévy process?

7. Given any $a \geqslant 0$, set $S_a = \inf\{t \geqslant 0\,;\, B_t > a\}$ and $T_a = \inf\{t \geqslant 0\,;\, B_t \geqslant a\}$.

a) Prove that S_b and T_b are almost-surely equal.

b) Let L be a non-negative random time independent of the filtration generated by B. Prove that the event $\{T_L \neq S_L\}$ is measurable and $\mathbb{P}(T_L \neq S_L) = 0$.

c) Find a random time L for which $\mathbb{P}(T_L = S_L) = 0$.

8. Occupation time. Let D be an open ball of $\mathbb{R}^d$ and x be any starting point for Brownian motion.

a) Prove that $\mathbb{P}_x\big(\int_0^\infty \mathbf{1}_D(B_t)\,dt = \infty\big) = 1$, if $d = 1$ or 2.

b) Prove that $\mathbb{E}_x\big[\int_0^\infty \mathbf{1}_D(B_t)\,dt\big] < \infty$, for $d \geqslant 3$.

9. Let $B = (B^1, B^2)$ be a 2-dimensional Brownian motion starting from the point with coordinates $(1, 0)$. Setting $T = \inf\{t \geqslant 0\,;\, B^2_t = 0\}$, what is the law of B^1_T?

10. Let B be here an $\mathbb{R}^d$-valued Brownian motion, $r > 0$ and $x \in \mathbb{R}^d$ with $\|x\| < r$. Set $H = \inf\{s \geqslant 0\,;\, \|B_s\| = r\}$. Prove that $\mathbb{E}_x[T] = \frac{r^2 - \|x\|^2}{d}$.

11. Uniqueness in Dirichlet problem. Let O be a bounded open set and g be a solution to Dirichlet problem, with continuous boundary condition f. Prove that

$$\max_{x\in O} g(x) = \max_{y\in\partial O} g(y) \ \big(= \max_{y\in\partial O} f(y)\big).$$

Conclude that the Dirichlet problem has at most one solution.

12. Let N be a Poisson process of intensity λ. Prove that the number of jumps of N by time $t > 0$ is a Poisson random variable with parameter λt.

13. Prove that a Poisson process is a Lévy process.

14. A Poisson process of rate λ is observed by someone who believes that the first holding time if longer than all the other holding times. How long on average will it take before the observer is proved wrong?

15. Let N be a Poisson process of intensity λ. Given any time $t > 0$, denote by $T_t = \inf\{s \geqslant t\,;\, N_s \neq N_t\}$ the next jump time after time t.

a) Prove that we have almost-surely $T_t > t$.

b) Prove that $T_t - t$ is exponentially distributed, with parameter λ. *This is surprising as the interval $[t, T_t - t]$ is contained in one of the intervals between jumps, all of which are exponentially distributed, with parameter λ(!).* Can you explain that paradox?

16. Is the sum of two Lévy processes always a Lévy process?

17. Can a process with stationnary and independent increments not be a Lévy process?

18. Given a Lévy process X, set $\Delta X_t := X_t - X_{t-}$. Prove that we have almost-surely $\Delta X_t = 0$ for any fixed $t > 0$, so Lévy processes do not have jumps at fixed times. *This result generalizes the corresonding result for Poisson processes proved in question **a)** in the exercise 15.*

19. Let X be a Lévy process with jump measure Λ_X of finite mass.

a) Prove that X has almost-surely finitely many jumps in any bounded interval of time.

b) Denote by $(\Delta X)_n$ the n^{th} jump of X, and let $(\epsilon_n)_{n\geqslant 1}$ a collection of independent Bernoulli random variables , with parameter $p \in (0, 1)$, independent of X. Let Y be the process obtained from X by removing from X all the jumps of X for which $\epsilon_n = 0$, at the time when they occur: If X has made n_t jumps by time t we have $Y_t = X_t - \sum_{j=1}^{n_t}(1 - \epsilon_j)(\Delta X)_j$. The process Y is càdlàg . Prove that Y is a Lévy process and find its jump measure Λ_Y.

20. Using the same method as was used for Brownian motion in the course, state and prove the strong Markov property for a Lévy process.

21. Using the same method as in exercise 19 in example sheet 2, prove that the filtration generated by a Lévy process, completed with null sets, is continuous on the right.

22. **a)** Prove that a Lévy process can always be written as the sum of two independent Lévy processes.

b) Deduce from **a)** and exercise 17 that a Lévy process is almost-surely continuous iff it is a Brownian motion with drift.

14. Complement to Part III

14.1. ***Complement: Infinite sums of infinitesimal independent random variables.*** As is clear from the definition of a Lévy process X, the random variable X_1 can be decomposed for all $n \geqslant 1$ as a sum of n iid random variables: $X_1 = \sum_{k=1}^{n} \big(X_{\frac{k}{n}} - X_{\frac{k-1}{n}} \big)$. Random variables which have this property are called ***infinitely divisible***. The proof of Lévy-Khinchin's formula can be copied word by word to prove that any infinitely divisible random variable has a characteristic function of the form $e^{g(\lambda)}$ for a Lévy-Khinchin function g. Rather than using the measure Λ with support in $\mathbb{R}^*$ we shall use the measure μ and the "drift" b obtained initially in formula (11.1), out of which Λ was derived by isolating the mass at 0. With this formalism, Lévy triples become Lévy pairs $(b\,;\mu)$. The following stability property is worth being noted.

Lemma **118.** *If a sequence of infinitely divisible random variables converges weakly then its weak limit is infinitely divisible.*

Proof – Let φ be the characteristic function of the weak limit of a sequence of infinitely divisible random variables, with characteristic functions $\varphi^{(k)}$. As each $\varphi^{(k)} = \big\{ \varphi_n^{(k)} \big\}^n$, for some charcateristic function $\varphi_n^{(k)}$, the functions $\big| \varphi^{(k)} \big|^{\frac{2}{n}}$ are characteristic functions[81]. Since they converge to the continuous function $|\varphi|^{\frac{2}{n}}$ as $k \to \infty$, the latter is a characteristic function, by Lévy's continuity theorem; so $|\varphi|^2$ is infinitely divisible, as n is arbitrary. As such, it cannot vanish, and φ cannot either. All the functions $\varphi^{\frac{1}{n}} = \lim_{k\to\infty} \big\{ \varphi_n^{(k)} \big\}^{\frac{1}{n}}$ are thus well-defined characteristic functions (as they are continuous at 0), which proves the claim. $\triangleright$

Definition **119.** *By a* ***triangular array*** *we mean a sequence of finite collections* $\big\{ X_{nk}\,;\, 1 \leqslant k \leqslant k(n) \big\}$ *of independent random variables.*

Set $S_n = X_{n1} + \cdots + X_{nk(n)}$. We are going to prove that S_n converges to an infinite divisible random variable under quite general conditions.

Assumptions. • *All the random variables X_{nk} are in $\mathbb{L}^2$,*

- $\sup_{1\leqslant k \leqslant k(n)} \mathrm{Var}(X_{nk}) \underset{n\to\infty}{\longrightarrow} 0$,
- $\sum_{k=1}^{k(n)} \mathrm{Var}(X_{nk})$ *is bounded above by a constant independent of n, say C.*

It will be convenient to denote by μ_{nk} the law of X_{nk} and by $\overline{\mu}_{nk}$ the law of the recentered random variable $X_{nk} - \mathbb{E}[X_{nk}]$.

[81]Denote by $X_n^{(k)}$ a random variable whose distribution has characteristic function $\varphi_n^{(k)}$, and let $\widehat{X}_n^{(k)}$ be an independent copy of $-X_n^{(k)}$. Then $\big| \varphi^{(k)} \big|^{\frac{2}{n}}$ is the characteristic function of $X_n^{(k)} + \widehat{X}_n^{(k)}$.

PROPOSITION **120.** *The random variables S_n converge weakly iff the sequence of infinite divisible laws with exponent*

$$\psi_n(\lambda) = \sum_{k=1}^{k(n)} \Big(i\lambda \mathbb{E}[X_{nk}] + \int (e^{i\lambda x} - 1)\,\overline{\mu}_{nk}(dx) \Big)$$

converges, in which case the two limits are equal.

PROOF – Write $\overline{\varphi}_{nk}(\lambda)$ for the characteristic function of $X_{nk} - \mathbb{E}[X_{nk}]$, and set $a_{nk}(\lambda) = \overline{\varphi}_{nk}(\lambda) - 1 = \int (e^{i\lambda x} - 1)\,\overline{\mu}_{nk}(dx)$. We have

$$\varphi_{S_n}(\lambda) = e^{i\lambda \sum_{k=1}^{k(n)} \mathbb{E}[X_{nk}]} \prod_{k=1}^{k(n)} \overline{\varphi}_{nk}(\lambda) = e^{i\lambda \sum_{k=1}^{k(n)} \mathbb{E}[X_{nk}]} \prod_{k=1}^{k(n)} \big(1 + a_{nk}(\lambda)\big).$$

Note that since $\int x\,\overline{\mu}_{nk}(dx) = 0$, one can write $a_{nk}(\lambda) = \int \big(e^{i\lambda x} - 1 - i\lambda x\big)\,\overline{\mu}_{nk}(dx)$. As the absolute value of the integrand is bounded above by $\frac{\lambda^2 x^2}{2}$, the estimate $\big|a_{nk}(\lambda)\big| \leqslant \frac{\lambda^2}{2}\mathrm{VAR}(X_{nk})$ follows and shows that $a_{nk}(\lambda)$ converges to 0, uniformly for λ in a compact. The statement is then obtained directly from the following inequalities and our assumptions.

$$\begin{aligned}
\Big|\log \varphi_{S_n}(\lambda) - \sum_{k=1}^{k(n)} \big(i\lambda \mathbb{E}[X_{nk}] + a_{nk(\lambda)}\big)\Big| &= \Big|\sum_{k=1}^{k(n)} \big(\log \overline{\varphi}_{nk}(\lambda) - a_{nk(\lambda)}\big)\Big| \\
&\leqslant \sum_{k=1}^{k(n)} \sum_{k \geqslant 2} \frac{\big|a_{nk}(\lambda)\big|^p}{p} \leqslant \frac{1}{2} \sum_{k=1}^{k(n)} \frac{\big|a_{nk}(\lambda)\big|^2}{1 - \big|a_{nk}(\lambda)\big|} \\
&\leqslant \max_{k=1..k(n)} \big|a_{nk}(\lambda)\big| \sum_{k=1}^{k(n)} \big|a_{nk}(\lambda)\big| \leqslant \frac{\lambda^2}{2} C \max_{k=1..k(n)} \big|a_{nk}(\lambda)\big|.
\end{aligned}$$

▷

This statement brings back the study of the behaviour of S_n to the study of a sequence of infinite divisible random laws. Denote by $(b_n\,;\nu_n)$ the Lévy pair associated to the exponent ψ_n constructed in proposition 120. We shall write $\mathrm{ID}(b\,;\nu)$ for a generic infinitely divisible random variable with Lévy pair $(b\,;\nu)$.

THEOREM **121.** *The random variables S_n converge weakly to some infinite divisible random variable with Lévy pair $(b\,;\nu)$ iff*

(1) *the measures ν_n converge weakly to ν,*
(2) *b_n converges to b.*

Before proving this statement let us single out the following two important practical cases.

COROLLARY **122** (Convergence to normal and Poisson laws). • *The random variables S_n converge weakly to a normal random variable iff $b = 0$ and for all $\epsilon > 0$ we have $\sum_{k=1}^{k(n)} \int_{|x|>\epsilon} x^2\,\overline{\mu}_{nk}(dx) \underset{n\to\infty}{\longrightarrow} 0$ and $\sum_{k=1}^{k(n)} \int_{|x|<\epsilon} x^2\,\overline{\mu}_{nk}(dx) \underset{n\to\infty}{\longrightarrow} 1$,*

• *Suppose $\sum_{k=1}^{k(n)} \mathbb{E}[X_{nk}] \underset{n\to\infty}{\longrightarrow} \lambda$ and $\sum_{k=1}^{k(n)} \mathrm{VAR}(X_{nk}) \underset{n\to\infty}{\longrightarrow} \lambda$. Then S_n converge weakly to a Poisson random variable iff for all $\epsilon > 0$ we have $\sum_{k=1}^{k(n)} \int_{|x-1|>\epsilon} x^2\,\overline{\mu}_{nk}(dx) \underset{n\to\infty}{\longrightarrow} 0$.*

PROOF – According to proposition 120, everything amounts to prove that $\mathrm{ID}(b_n\,;\nu_n)$ converges weakly to $\mathrm{ID}(b\,;\nu)$ iff ν_n converges weakly to ν and b_n converges to b. Denote by $\psi(\lambda)$ the characteristic exponent of $\mathrm{ID}(b\,;\nu)$. The implication $\Leftarrow$ is obvious since the characteristic function of $\mathrm{ID}(b_n\,;\nu_n)$ converges to the characteristic function of $\mathrm{ID}(b\,;\nu)$ in that case[82], so Lévy theorem on characteristic functions applies.

To prove the converse implication, note that since the weak convergence of $\mathrm{ID}(b_n\,;\nu_n)$ to $\mathrm{ID}(b\,;\nu)$ implies the convergence of their characteristic functions, uniformly on bounded intervals, ψ_n converges to ψ in that sense. We now play the same game as in the proof of the uniqueness of Lévy-Khinchin's representation. Namely, define

$$\rho(\lambda) = \int_{-1}^{1} \big(\psi(\lambda) - \psi(\lambda+s)\big)\,ds = 2\int e^{i\lambda x}\Big(1-\frac{\sin x}{x}\Big)\frac{1+x^2}{x^2}\,\nu(dx),$$

and ρ_n by a similar formula. As ρ_n converges uniformly on compacts to ρ the measures $\big(1-\frac{\sin x}{x}\big)\frac{1+x^2}{x^2}\,\nu_n(dx)$ converge weakly to the measure $\big(1-\frac{\sin x}{x}\big)\frac{1+x^2}{x^2}\,\nu(dx)$, by Lévy's theorem. Since the integrand is continuous and bounded away from 0 the measures ν_n themselves converge weakly to ν. The convergence of b_n to b follows. $\triangleright$

15. Solutions to the exercises

15.1. **Exercises on part I. 4. a)** Suppose there exists a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a collection of real-valued random variables $X=(X_t)_{t\in T}$ defined on $(\Omega,\mathcal{F})$ which form a Gaussian process. Their distribution $\mathbb{Q}$ is a probability measure on the product space $\mathbb{R}^T$, equipped with its product σ-algebra. As the class of elementary events $\{x\in\mathbb{R}^T\,;\,x_{t_1}\in A_1,\dots,x_{t_n}\in A_n\}$, for some $1\leqslant n<\infty$, $t_1,\dots,t_n\in T$ and $A_1,\dots,A_n$ Borel sets of $\mathbb{R}$, is a π-system generating the product σ-algebra, $\mathbb{Q}$ is entirely determined by its values on these elementary sets. Fixing n and $t_1,\dots,t_n$,

$$\mathbb{Q}(x\in\mathbb{R}^T\,;\,x_{t_1}\in A_1,\dots,x_{t_n}\in A_n) = \mathbb{P}(\omega\in\Omega\,;\,X_{t_1}(\omega)\in A_1,\dots,X_{t_n}(\omega)\in A_n).$$

Now, the distribution of the $\mathbb{R}^n$-valued random variable $(X_{t_1},\dots,X_{t_n})$ is characterized by its Fourier transform, so if we know $\mathbb{E}\big[e^{i\sum_{k=1..n}c_kX_{t_k}}\big]$ for all $c_i\in\mathbb{R}$, we (formally) know $\mathbb{P}(X_{t_1}\in A_1,\dots,X_{t_n}\in A_n)$ for all $A_1,\dots,A_n$. This is precisely the case as $\mathbb{E}\big[e^{i\sum_{k=1..n}c_kX_{t_k}}\big] = e^{im(c_1,\dots,c_n)-\frac{\sigma^2(c_1,\dots,c_n)}{2}}$ is determined by the mean and covariance functions $m(\cdot)$ and $\sigma^2(\cdot)$ respectively. As $n\geqslant 1$ and $t_1,\dots,t_n$ are arbitrary we are done.

b) Let denote by $(\Omega,\mathcal{F},\mathbb{P})$ the product space $\mathbb{R}^{\mathbb{N}}$ with the product probability $\mathcal{N}(0,1)^{\otimes\mathbb{N}}$.

Note first that as the random variables G_n are independent we have for any $1\leqslant p<q<\infty$

$$\Big\|\sum_{n=p}^{q} h^nG_n\Big\|_2 = \sum_{n=p}^{q}(h^n)^2,$$

the sequence $\Big(\sum_{n=0}^{q} h^nG_n\Big)_{q\geqslant 0}$ converges in $\mathbb{L}^2(\mathbb{P})$, so the random variable $X_h=\sum_{n\geqslant 0}h^nG_n$ is well-defined in $\mathbb{L}^2(\mathbb{P})$ and almost-surely, and has null mean. From the independence of the G_n: $\mathbb{E}[X_hX_{h'}] = \sum_{n\geqslant 0}h^n(h')^n = (h,h')$.

c) (i) Pick $0\leqslant s_1<s_2<\cdots<s_n$. As the random vector $(B_{s_1},B_{s_2}-B_{s_1},\cdots,B_{s_n}-B_{s_{n-1}}) = \big(X_{\mathbf{1}_{[0,s_1]}},X_{\mathbf{1}_{(s_1,s_2]}},\cdots,X_{\mathbf{1}_{(s_{n-1},s_n]}}\big)$ is a Gaussian vector, its components are independent iff it has diagonal covariance matrix, i.e. iff

$$\mathbb{E}\big[X_{\mathbf{1}_{(s_{i-1},s_i]}}X_{\mathbf{1}_{(s_{j-1},s_j]}}\big] = 0$$

for $i\neq j$, which holds since the expectation equals $\int \mathbf{1}_{(s_{i-1},s_i]}(x)\mathbf{1}_{(s_{j-1},s_j]}(x)\,dx = 0$.

[82]Recall the function $f(\lambda,x)$ appearing in Lévy-Khinchin's formula is continuous and bounded.

(ii) As $\mathbb{E}\big[|B_t - B_s|^4\big] = |t-s|^2$, Kolmogorov's regularity criterion applies.

(iii) This modification retains the finite dimensional properties of the original process, so it has covariance $\mathbb{E}[B_s B_t] = \min(s,t)$. Question **a)** shows that this property characterizes Brownian motion amongst the Gaussian processes.

(iv) Check that X is centered, Gaussian, with the above covariance function. The only non-trivial point is the continuity at 0 of the process $B_t := tX_{1/t}$. Since B is almost-surely continuous on $(0,\infty)$ one can describe the event $\{B \underset{t\downarrow 0}{\to} 0\}$ in terms of conditions on the values of B at countably many points of $(0,\infty)$. But B and X being Gaussian, with the same covariance and the same value at time 1, they have the same law on $(0,\infty)$; so $\mathbb{P}\big(B_t \underset{t\downarrow 0}{\to} 0\big) = \mathbb{P}\big(X_t \underset{t\downarrow 0}{\to} 0\big) = 1$.

5. a) As the complementary set of an open set if a closed set, the collection $\mathcal{C}$ is stable by complementation; it contains $[0,1]$. Let $\epsilon > 0$ be given, $(B_n)_{n\geqslant 0}$ be a sequence of disjoint elements of $\mathcal{C}$ and, for each $n \geqslant 0$, let O_n (resp. C_n) be an open (resp. closed) set containing (resp. contained in) B_n, with $\mathbb{P}(O_n \backslash B_n) \leqslant \epsilon\, 2^{-n}$ and $\mathbb{P}(B_n \backslash C_n) \leqslant \epsilon\, 2^{-n}$. Pick N large enough to have $\mathbb{P}\big(\bigcup_{n\geqslant 0} B_n \backslash \bigcup_{n=0}^N B_n\big) = \mathbb{P}\big(\bigcup_{n\geqslant N+1} B_n\big) \leqslant \epsilon$. The set $\bigcup_{n=0}^N C_n$ is closed and

$$\begin{aligned}\mathbb{P}\Big(\bigcup_{n\geqslant 0} B_n \backslash \bigcup_{n=0}^N C_n\Big) &\leqslant \mathbb{P}\Big(\bigcup_{n\geqslant N+1} B_n\Big) + \mathbb{P}\Big(\bigcup_{n=0}^N B_n \backslash \bigcup_{n=0}^N C_n\Big) \leqslant \epsilon + \mathbb{P}\Big(\bigcup_{n=0}^N (B_n \backslash C_n)\Big)\\ &\leqslant \epsilon + \sum_{n=0}^N \epsilon\, 2^{-n} \leqslant 3\,\epsilon.\end{aligned}$$

Also, the set $\bigcup_{n\geqslant 0} O_n$ is open and

$$\mathbb{P}\Big(\bigcup_{n\geqslant 0} O_n \backslash \bigcup_{n\geqslant 0} B_n\Big) \leqslant \mathbb{P}\Big(\bigcup_{n\geqslant 0} (O_n \backslash B_n)\Big) \leqslant \sum_{n\geqslant 0} \mathbb{P}(O_n \backslash B_n) \leqslant 2\,\epsilon.$$

As $\epsilon > 0$ is arbitrary, this proves that $\bigcup_{n\geqslant 0} B_n \in \mathcal{C}$, from which it follows that $\mathcal{C}$ is a σ-algebra.

b) Trivially, intervals are in $\mathcal{C}$, so the σ-algebra they genearate is included in $\mathcal{C}$. This σ-algebra is Bor, which proves the inner and outer regularity of $\mathbb{P}$.

6. a) Recall that $(X_n)_{n\geqslant 0}$ converges weakly to X iff $\mathbb{E}\big[f(X_n)\big] \to \mathbb{E}\big[f(X)\big]$ for any bounded uniformly continuous function f (2nd statement of Alexandrov's characterization). For such an f we have for each $\epsilon > 0$

$$\begin{aligned}\big|\mathbb{E}\big[f(X_n) - f(X)\big]\big| &\leqslant \mathbb{E}\big[2\|f\|_\infty \mathbf{1}_{|X_n - X|\geqslant\epsilon}\big] + \mathbb{E}\big[\big|f(X_n) - f(X)\big|\mathbf{1}_{|X_n-X|<\epsilon}\big]\\ &\leqslant 2\|f\|_\infty\, \mathbb{P}\big(|X_n - X| \geqslant \epsilon\big) + o_\epsilon(1),\end{aligned}$$

by the uniform continuity of f. The upper bound converges to $o_\epsilon(1)$ as n goes to infinity, which can be made arbitrarily small by choosing small ϵ.

b) Let X be equal to 0 or 1 with equal probability, and $X_n = X$ for all $n \geqslant 1$. Then X_n has the same distribution as $1 - X$ but does not converge to $1 - X$ in probability.

7. A sequence $(\mu_n)_{n\geqslant 0}$ converging to μ in $\mathcal{B}_b(\mathbb{R})^*$ also converges to μ in $\mathcal{C}_b(\mathbb{R})^*$. The converse does not hold: $\delta_{\frac{1}{n}}$ converges to δ_0 in $\mathcal{C}_b(\mathbb{R})^*$ but not in $\mathcal{B}_b(\mathbb{R})^*$ since we have $0 = \big(\mathbf{1}_{[-1,0]}, \delta_{[\frac{1}{n}]}\big) \neq \big(\mathbf{1}_{[-1,0]}, \delta_0\big) = 1$, for all $n \geqslant 0$.

8. Again, we use here as in exercise 6 the fact that $(\mu_n)_{n\geqslant 0}$ converges weakly to μ iff the integrals (f, μ_n) converge to (f, μ) for all bounded uniformly continuous functions f.

Suppose this convergence holds a priori only for all continuous functions with compact support, and let f be a bounded uniformly continuous functions. Pick $\epsilon > 0$ and let $0 \leqslant \phi \leqslant 1$ be a function with compact support, equal to 1 in an interval $[-M, M]$, big enough so that we have $(\phi, \mu) \geqslant 1 - \epsilon$, and so $\big((1-\phi), \mu\big) \leqslant \epsilon$. Then

$$\big|(f, \mu_n) - (f, \mu)\big| \leqslant \big|(f\phi, \mu_n) - (f\phi, \mu)\big| + \big|\big((1-\phi)f, \mu_n\big) - \big((1-\phi)f, \mu\big)\big|$$

The first term on the rhs converges to 0 since $f\phi$ has compact support. The second term is bounded above by

$$\|f\|_\infty\big((1-\phi, \mu_n) + (1-\phi, \mu)\big) \leqslant \|f\|_\infty\big(1 + \epsilon - (\phi, \mu_n)\big).$$

As $(\phi, \mu_n) \to (\phi, \mu) \geqslant 1 - \epsilon$, the upper bound is smaller than $2\epsilon\,\|f\|_\infty$ for n big enough.

9. Using the almost-sure representation of weak convergence, one can write $\phi_n(\lambda) = \mathbb{E}\big[e^{i\lambda X_n}\big]$ and $\phi(\lambda) = \mathbb{E}\big[e^{i\lambda X}\big]$, for some random variables X_n with law μ_n and X with law μ, defined on some probability space ($[0,1]$, actually!), with X_n converging almost-surely to X. Given $M > 0$ and $\epsilon > 0, \eta > 0$, we have $|X_n - X| \leqslant \frac{\epsilon}{M}$ on a set of probability bigger than $1 - \eta$, for $n \geqslant N(\epsilon, \eta)$. So

$$\begin{aligned}
\sup_{\lambda\in[-M,M]} |\phi_n(\lambda) - \phi(\lambda)| &= \sup_{\lambda\in[-M,M]} \big|\mathbb{E}\big[e^{i\lambda X_n} - e^{i\lambda X}\big]\big| \\
&\leqslant \sup_{\lambda\in[-M,M]} \mathbb{E}\big[\big|e^{i\lambda X_n} - e^{i\lambda X}\big|\mathbf{1}_{|\lambda X_n - \lambda X|\leqslant\epsilon}\big] + 2\eta \\
&\leqslant 2\sin\frac{\epsilon}{2} + 2\eta,
\end{aligned}$$

for $n \geqslant N(\epsilon, \eta)$. The result follows as $\epsilon > 0$ and $\eta > 0$ are arbitrary.

10. Suppose the family $(\mu_n)_{n\geqslant 0}$ tight and associate to any $\epsilon > 0$ an $M_\epsilon > 0$ such that $\mu_n\big([-M_\epsilon, M_\epsilon]\big) \geqslant 1 - \epsilon$, for all $n \geqslant 0$. Then

$$\big|\phi_n(\lambda) - 1\big| \leqslant \Big|\mathbb{E}\Big[\big(e^{i\lambda X_n} - 1\big)\mathbf{1}_{\big[-M_\epsilon, M_\epsilon\big]}\Big]\Big| + 2\epsilon.$$

For $\lambda \leqslant \frac{\eta}{M_\epsilon}$, we have $\big(e^{i\lambda X_n} - 1\big)\mathbf{1}_{\big[-M_\epsilon, M_\epsilon\big]} \leqslant 2\sin\frac{\eta}{2}$, from which the result follows.

Reciprocally, if the ϕ_n's are equicontinuous at 0, use formula just before the proof of theorem 29 to conclude that the family $(\mu_n)_{n\geqslant 0}$ of probabilities is tight.

11. We proceed in steps, proving first the statement for an iid sequence $(U_n)_{n\geqslant 0}$ of uniformly distributed random variables. Given $t \in [0,1]$, the random variables $\mathbf{1}_{X_n\leqslant t}$ are iid. The SLLN gives in that case the almost-sure convergence $\widehat{F}_n(t) \to \mathbb{E}\big[\mathbf{1}_{U_0\leqslant t}\big] = t$. As a finite intersection of events of probability 1 has probability 1, we have almost-surely

$$\sup_{t\in F} \big|\widehat{F_n}(t) - t\big| \to 0$$

for any finite family F of elements of $[0,1]$. Now, by monotonocity of $\widehat{F_n}$, and given some times $0 = t_0 < t_1 < \cdots < t_p = 1$,

$$\sup_{t\in[0,1]} \big|\widehat{F_n}(t) - t\big| \leqslant \max_{k\in\{0,\cdots,p\}} \big|\widehat{F_n}(t_k) - t_k\big| + \max_{k\in\{0,\cdots,p-1\}} \big|t_{k+1-t_k}\big|.$$

Sending n to infinity and refining the partition, we get the result in that case.

To deal with the general case, we use the representation of a random variable as the image of a uniformly distributed random variable. Let G denote the distribution function of the common law of the X_n's. Set $g(t) = \sup\{y\,;\, G(y) < t\}$, so that $g(U_n) \leqslant x$ iff $U_n \leqslant G(x)$, that is, the sequence $\big(g(U_n)\big)_{n\geqslant 0}$ has the same law as $(X_n)_{n\geqslant 0}$. We are thus brought back to prove that we have almost-surely

$$\sup_{x\in\mathbb{R}} \Big|\frac{1}{n+1}\sum_{k=0..n} \mathbf{1}_{g(U_n)\leqslant x} - G(x)\Big| \to 0$$

As the lhs equals

$$\sup_{t\in[0,1]} \big|\widehat{F_n}(t) - t\big|$$

by a change of variable, this is clear.

12. b) Using the almost-sure representation theorem for weakly convergent sequences, one can write almost-surely by Taylor's theorem for $\mathcal{C}^1$ functions

$$\sqrt{n}\big(f(X_n) - f(m)\big) = f'(m)\sqrt{n}(X_n - m) + \sqrt{n}\,o\big(|X_n - m|\big) = f'(m)\sqrt{n}(X_n - m) + o\big(\sqrt{n}|X_n - m|\big).$$

As $\sqrt{n}(X_n - m)$ is almost-surely converging to some random variable Y the rhs above converges almost-surely to $f'(m)\,Y$, hence the statement.

14. b) The application $\phi : x \in \mathcal{C}\big([0,1],\mathbb{R}\big) \to \max_{t\in[0,1]} x_t$ is continuous. Denote by μ_n the law of X_n under $\mathbb{P}$. All the μ_n's have support in the set $\big\{x \in \mathcal{C}\big([0,1],\mathbb{R}\big)\,;\, \max_{t\in[0,1]} x_t = 1\big\}$, so the image measure of μ_n by ϕ is the Dirac mass at 1. The image measure of the law of X by ϕ is the Dirac mass at 0, so $(X_n)_{n\geqslant 0}$ cannot converge wakly to X by **a)**.

15. a) The vector $\big(X^0_{t_1},\ldots,X^0_{t_n},X_1\big)$ is Gaussian under $\mathbb{P}$, with $X_1 \sim \mathcal{N}(0,1)$. We check by a direct computation that its covariance matrix has the form $\begin{pmatrix} A & (0) \\ (0) & 1 \end{pmatrix}$, for some symmetric $n\times n$ matrix A. It follows that X_1 is independent under $\mathbb{P}$ of the $\mathbb{R}^n$-valued Gaussian random vector $\big(X^0_{t_1},\ldots,X^0_{t_n}\big)$; in particular

$$\mathbb{P}\big(X^0_{t_1}\in A_1,\ldots,X^0_{t_n}\in A_n \big| 0 \leqslant X_1 \leqslant \varepsilon\big) = \mathbb{P}\big(X^0_{t_1}\in A_1,\ldots,X^0_{t_n}\in A_n\big),$$

so X^0 has under $\mathbb{P}_\varepsilon$ the same *finite dimensional laws* as X^0 under $\mathbb{P}$, that is $\mathbb{P}_0$. As the finite dimensional distributions characterize uniquely the distribution it follows that the distribution of X^0 under $\mathbb{P}_\varepsilon$ is independent of ε and equal to $\mathbb{P}_0$.

b) Let now F be a closed set of $\Big(\mathcal{C}\big([0,1],\mathbb{R}\big),\|\cdot\|_\infty\Big)$, and $F^\epsilon = \big\{x \in \mathcal{C}\big([0,1],\mathbb{R}\big)\,;\, d(x,F) \leqslant \epsilon\big\}$ be the ϵ-beighbourhood of F; this is a closed set, and $\bigcap_{\epsilon>0} F^\epsilon = F$. As we have almost-surely $|X^0 - X| \leqslant \epsilon$ under $\mathbb{P}_\epsilon$, the random path $X^0(\omega)$ is $\mathbb{P}_\epsilon$-almost-surely in F^ϵ if $X(\omega)$ is in F. So, fixing η and taking $0 < \epsilon < \eta$, we have

$$\mathbb{P}_\epsilon(X\in F) \leqslant \mathbb{P}_\epsilon(X^0\in F^\epsilon) \leqslant \mathbb{P}_\epsilon(X^0\in F^\eta) = \mathbb{P}(X^0\in F^\eta).$$

Send first ϵ to 0 to get
$$\varlimsup_{\epsilon\searrow 0}\mathbb{P}_\epsilon(X\in F)\leqslant \mathbb{P}(X^0\in F^\eta),$$
then send η to 0 (using monotone convergence)
$$\varlimsup_{\epsilon\searrow 0}\mathbb{P}_\epsilon(X\in F)\leqslant \mathbb{P}\big(X^0\in \bigcap_{\eta\searrow 0} F^\eta\big)=\mathbb{P}(X^0\in F).$$

15.2. **Exercises on part II.** This section was contributed by Bati Sengul; thanks to him for his work.

1. a) We need to prove that we have $\mathbb{E}\big[h(V)\mathbf{1}_A\big]=\mathbb{E}[g(U)\mathbf{1}_A]$, for each $A\in\sigma(U)$; any such event is by definition of the form $\mathbf{1}_B(U)$, for some measurable subset B of $\mathbb{R}$. Using Fubini's theorem, we have
$$\begin{aligned}\mathbb{E}\big[h(V)\mathbf{1}_B(U)\big] &= \int_{\mathbb{R}}\int_{\mathbb{R}} f_{U,V}(u,v)h(v)\mathbf{1}_B(u)\,dudv=\int_{\mathbb{R}}\mathbf{1}_B(u)\int_{\mathbb{R}} f_{U,V}(u,v)h(v)\,dvdu\\ &= \int_{\mathbb{R}}\mathbf{1}_B(u)g(u)f_U(u)\,du=\mathbb{E}\big[g(U)\mathbf{1}_B(u)\big].\end{aligned}$$

b) Consider
$$X:=V-\frac{Cov(U,V)}{Var(U)}U$$
then X is a centred Gaussian random variable, moreover
$$\mathbb{E}[XU]=\mathbb{E}[UV]-\frac{Cov(U,V)}{Var(U)}\mathbb{E}\big[U^2\big]=\mathbb{E}[UV]-Cov(U,V)=0$$
hence X is independent of U, so $\mathbb{E}[X|\sigma(U)]=0$. Now
$$\begin{aligned}\mathbb{E}\big[V|\sigma(U)\big] &= \mathbb{E}\Big[\frac{Cov(U,V)}{Var(U)}U+X\Big|\sigma(U)\Big]\\ &= \frac{Cov(U,V)}{Var(U)}U.\end{aligned}$$

c) Let us prove more generally that any $\sigma(U)$-measurable almost-surely finite random variable X is of the form $f(U)$ for some measurable function $f:\mathbb{R}\to\mathbb{R}$.

Suppose first that X takes only finitely many values $x_1,\dots,x_n$. As each set $A_i=X^{-1}\big(\{x_i\}\big)$ belongs to $\sigma(U)$, it is of the form $U^{-1}(B_i)$ for some measurable $B_i\subset\mathbb{R}$; the B_i's are disjoint. Set $f(x)=x_i$ if $x\in B_i$ for some i, and $f(x)=0$ elsewhere. We check directly that $f(U)=X$.

For $X\geqslant 0$, we define a $\sigma(U)$-measurable random variable setting
$$X_n=\sum_{j=0}^{n2^n}\frac{j}{2^n}\mathbf{1}_{X\in(j2^{-n},(j+1)2^{-n}]}.$$
As it takes only finitely many values, it is of the form $f_n(U)$. Note that $X_n\uparrow X$ almost-surely. Set $\overline{f}=\varlimsup f_n$ and $f=\overline{f}\mathbf{1}_{\overline{f}<\infty}$ and check that $f(U)=X$ as X is almost-surely finite.

2. The trivial case $k=1$ is obvious. So suppose that the statement holds for k, i.e.
$$\mathbb{P}(T>kN)\leqslant(1-\epsilon)^k.$$

Then by using $\mathbb{P}(T > n+N|\mathcal{F}_n) \leqslant 1-\epsilon$ and the fact $\{T > (k+1)N\} \subset \{T > kN\}$ we have that[83]

$$\begin{aligned}
\mathbb{E}[\mathbf{1}_{T>(k+1)N}] &= \mathbb{E}[\mathbb{E}[\mathbf{1}_{T>(k+1)N}|\mathcal{F}_{kN}]] \\
&= \mathbb{E}[\mathbb{E}[\mathbf{1}_{T>(k+1)N}\mathbf{1}_{T>kN}|\mathcal{F}_{kN}]] \\
&= \mathbb{E}[\mathbf{1}_{T>kN}\mathbb{E}[\mathbf{1}_{T>(k+1)N}|\mathcal{F}_{kN}]] \\
&\leqslant \mathbb{E}[\mathbf{1}_{T>kN}(1-\epsilon)] \leqslant (1-\epsilon)^k(1-\epsilon).
\end{aligned}$$

3. a) Work with $\Omega = \mathcal{C}\big(\mathbb{R}_+, \mathbb{R}\big)$, the coordinate process and its filtration $(\mathcal{F}_t)_{t\geqslant 0}$, and set $\mathcal{G}_t = \{\emptyset, \Omega\}$ for $t \leqslant 1$, and $\mathcal{G}_t = \mathcal{F}_{t-1}$ for $t \geqslant 1$. Look at the hitting time of some level.

b) For any $b \leqslant a$ we have by the continuity of ω

$$\{\gamma_a \leqslant b\} = \{\omega \in \Omega\,;\, \omega_t > 0 \text{ for all } t \in (b,a]\} = \bigcap_{t\in(b,a]\cap\mathbb{Q}} \{\omega \in \Omega\,;\, \omega_t > 0\} \in \mathcal{F}_a.$$

Next we show that $\{\gamma_a < t\} \notin \mathcal{F}_t$ for $t < a$. Intuitively this fails ultimately because at time $t < a$ we cannot deduce if γ_a has happened or not, given the path up to time t. More rigorously

$$\{\gamma_a < t\} = \{\omega_s \neq 0\, \forall s \in [t,a]\} = \{\omega_s \neq 0\, \forall s \in [t,a]\cap\mathbb{Q}\} = \bigcap_{s\in[t,a]\cap\mathbb{Q}} \{\omega_s \neq 0\}$$

where we have used the continuity in the second equality. Now the last part is not in $\mathcal{F}_t$ and hence $\{\gamma_a < t\} \notin \mathcal{F}_t$.

4. b) Obviously we have that $\mathcal{F}_{S\wedge T} \subset \sigma(\mathcal{F}_S, \mathcal{F}_T)$. For the converse notice first that $\sigma(F_S, F_T)$ is generated by events in $\mathcal{F}_S$ and $\mathcal{F}_T$, hence by the monotone class theorem, it suffices to check that $\mathcal{F}_S$ and $\mathcal{F}_T$ are included in $\mathcal{F}_{S\wedge T}$. Let $A \in \mathcal{F}_S$, then it suffices to show that $A \cap \{S\wedge T > t\} \in \mathcal{F}_t$ for each $t \geqslant 0$. Notice that

$$A \cap \{S \wedge T > t\} = A \cap \{S > t\} \cap \{T > t\}.$$

Now $B := A \cap \{S > t\} \in \mathcal{F}_t$, by definition of $\mathcal{F}_S$, and as T is a stopping time $B \cap \{T > t\} \in \mathcal{F}_t$ and hence $A \in \mathcal{F}_{S\wedge T}$. Similarly for $A \in \mathcal{F}_T$.

5. Suppose that $X_n \to X$ in L^1. Then by Markov's inequality $X_n \to X$ in probability:

$$\mathbb{P}\big(|X_n - X| > \epsilon\big) \leqslant \epsilon^{-1}\mathbb{E}\big[|X_n - X|\big] \to 0 \quad \text{as } n \to \infty.$$

Fix $\epsilon > 0$, then there exists an $N \in \mathbb{N}$ such that $\mathbb{E}\big[|X_n - X|\big] < \epsilon$ for $n \geqslant N$. The sequence $X, X_1, ..., X_N$ is finite and hence uniformly integrable, so there exists a $K > 0$ such that $\mathbb{E}\big[|X_n|\mathbf{1}_{|X_n|>K}\big] < \epsilon$ for all $n \leqslant N$ and $\mathbb{E}\big[|X|\mathbf{1}_{[X|>K}\big] < \epsilon$. For $n > N$ we have

$$\mathbb{E}\big[|X_n|\mathbf{1}_{|X_n|>K}\big] \leqslant \mathbb{E}\big[|X_n - X|\mathbf{1}_{|X_n|>K}\big] + \mathbb{E}\big[|X|\mathbf{1}_{|X_n|>K}\big] < \epsilon + \mathbb{E}\big[|X|\mathbf{1}_{|X_n|>K}\big].$$

Then the second term is small if $\mathbb{P}\big(|X_n| > K\big)$ is small, uniformly in n. But then by Markov's inequality and the fact that $\mathbb{E}[|X_n|] \leqslant \mathbb{E}[|X|] + \epsilon$ we have

$$\mathbb{P}\big(|X_n| > K\big) \leqslant K^{-1}\mathbb{E}\big[|X_n|\big] \leqslant K^{-1}(\mathbb{E}\big[|X|\big] + \epsilon)$$

which can be made small by choosing K large.

[83] Here I will use the fact that $\mathbf{1}_{T>(k+1)N}\mathbf{1}_{T>kN} = \mathbf{1}_{T>(k+1)N}$.

For the converse suppose that X_n is UI and $X_n \to X$ in probability. Consider the following approximation

$$\mathbb{E}\big[|X_n - X|\big] \leqslant \mathbb{E}\big[|X_n - X|\mathbf{1}_{|X_n-X|\leqslant K}\big] + \mathbb{E}\big[|X_n - X|\mathbf{1}_{|X_n-X|>K}\big] \leqslant K + \mathbb{E}\big[|X_n - X|\mathbf{1}_{|X_n-X|>K}\big].$$

Now by the uniform integrability the term on the RHS can be made small given that $\mathbb{P}\big(|X_n - X| > K\big)$ is small. Pick $K = \epsilon$ small and let $n > N$ be sufficiently large such that $\mathbb{E}\big[|X_n - X|\mathbf{1}_{|X_n-X|>K}\big] < \epsilon$.

6. Suppose that $\mathbb{P} << \mathbb{Q}$ then by the Radon-Nikodym theorem we have that $\mathbb{P}(A) = \mathbb{E}_{\mathbb{Q}}[X\mathbf{1}_A]$ where $X \in L^1(\mathbb{Q})$ and $0 \leqslant X \leqslant 1$, so in particular $\mathbb{P}(A) \leqslant \mathbb{Q}(A)$.

Conversely let $\mathbb{Q}(A) = 0$, then for each epsilon $\mathbb{P}(A) < \epsilon$, i.e. $\mathbb{P}(A) = 0$.

7. Suppose first that $\mathbb{Q} << \mathbb{P}$, then by the Radon-Nikodym theorem $\mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}[X\mathbf{1}_A]$ for all $A \in \mathcal{F}$ with $X \in L^1(\mathbb{P})$ and $0 \leqslant X \leqslant 1$. In particular we have that $M_n = \mathbb{E}[X|\mathcal{F}_n]$ and hence M_n is uniformly integrable.

Suppose on the other hand that M_n is uniformly integrable (with respect to $\mathbb{P}$), then M_n converges in $L^1(\mathbb{P})$ and a.s. to some M_∞, so that $M_n = \mathbb{E}[M_\infty|\mathcal{F}_n]$.

8. The idea is to prove that $\mathcal{T}$ is independent of itself. To that end define $\mathcal{F}_n := \sigma(X_k : k \leqslant n)$, then $\mathcal{F}_n$ is independent of $\sigma(X_k : k > n)$ (as the random variables are independent), and in particular independent from $\mathcal{T}$. This holds for all $n \in \mathbb{N}$ and hence $\mathcal{T}$ is independent of $\mathcal{F}_\infty := \bigvee_{n\geqslant 1} \mathcal{F}_n$. However $\mathcal{T} \subset \mathcal{F}_\infty$ and hence $\mathcal{T}$ is independent of itself. Now for any $A \in \mathcal{T}$, we have that $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A)$ so $\mathbb{P}(A)$ is either 0 or 1.

The trivial counterexample to when X_i are not independent is by considering $X_i = X$ for some non-trivial random variable, then $\mathcal{T} = \sigma(X)$ which is non-trivial.

9. a) Let $X \in \mathcal{F}_\infty$ be bounded, then $X_n := \mathbb{E}[X|\mathcal{F}_n]$ makes sense and is bounded by the same bound. Then by the martingale convergence $X_n \to X$ in L^1 and hence the result.

b) By part a), the bounded elements of $L^1(\mathcal{F}_\infty)$ are limit points of $\mathbb{E}[\cdot|\mathcal{F}_n] \in \bigcup_{k\geqslant 0} \mathcal{F}_k$. Now if $X \in L^1(\mathcal{F}_\infty)$ is not bounded, then it can be approximated by bounded functions.

c) Kolmogorov's 0-1 Law: We have that for any $A \in \mathcal{T}$

$$\mathbb{E}[\mathbf{1}_A|\mathcal{F}_n] \to \mathbf{1}_A$$

so as before A is independent of $\mathcal{F}_n$, hence $\mathbb{E}[\mathbf{1}_A|\mathcal{F}_n] = \mathbb{P}(A)$.

d) In the case $\mathcal{F}_n$ is finite, then

$$\mathbb{E}[X|\mathcal{F}_n] = \sum \frac{\mathbb{E}[X\mathbf{1}_{A_n}]}{\mathbb{P}(A_n)}\mathbf{1}_{A_n}$$

which is computable. So then the limits also may be computed explicitly.

(i) Suppose that the measure space is separable. First note that $L^1(\mathcal{F}_n)$ has countably many simple functions with rational coefficients and they are dense. Now $\bigcup L^1(\mathcal{F}_n)$ has a countable dense subset. By using double approximation, this set is also dense in $L^1(\mathcal{F}_\infty)$.

10. a) Notice that S_n is a submartingale and $S_n^{T_{ab}}$ is bounded and hence by the optional stopping theorem

$$\mathbb{E}[S_0] = 0 \leqslant \mathbb{E}[S_{T_{ab}}] = a\,\mathbb{P}(T_a \leqslant T_b) + b\,\mathbb{P}(T_b < T_a).$$

The equation above gives a lower bound

$$\mathbb{P}(T_b < T_a) \geqslant \frac{-a}{b-a}.$$

Now as $a \to -\infty$, $T_a \to \infty$ and the right hand side converges to 1, which gives that $\mathbb{P}(T_b < \infty) \geqslant 1$. From this it follows that both T_b and T_{ab} are finite.

(i) Direct computation shows that

$$\mathbb{E}\left[\left(\frac{q}{p}\right)^{S_n} - \left(\frac{q}{p}\right)^{S_{n-1}} |\mathcal{F}_{n-1}\right] = \mathbb{E}\left[\left(\frac{q}{p}\right)^{S_{n-1}}\left(\left(\frac{q}{p}\right)^{X_n} - 1\right)|\mathcal{F}_{n-1}\right] = \left(\frac{q}{p}\right)^{S_{n-1}}\left(\mathbb{E}\left(\frac{q}{p}\right)^{X_n} - 1\right).$$

It suffices to check that $\mathbb{E}[(q/p)^{X_n}] = 1$:

$$\mathbb{E}[(q/p)^{X_n}] = p\frac{q}{p} + q\frac{p}{q} = p + q = 1.$$

The martingale $X_n := (q/p)^{S_n}$ is bounded by 1, hence we may apply the optional stopping theorem to obtain

$$\mathbb{E}[X_0] = 1 = \mathbb{E}[X_{T_{ab}}] = (q/p)^a\mathbb{P}(T_a < T_b) + (q/p)^b\mathbb{P}(T_b < T_a).$$

Rearranging the above gives that

$$\mathbb{P}(S_{T_{ab}} = a) = \mathbb{P}(T_a < T_b) = \frac{1-(q/p)^b}{(q/p)^a - (q/p)^b}.$$

(ii) Let $X_n : S_n - n(p-q)$, then X_n is a martingale. Notice that $X^{T_{ab}}$ is bounded by $-a \vee b$, so by the optional stopping theorem

$$\mathbb{E}[X_0] = 0 = \mathbb{E}[X_{T_{ab}}] = \mathbb{E}[S_{T_{ab}}] - (p-q)\mathbb{E}[T_{ab}].$$

11. a) Let X_i^n be i.i.d Bernoulli $\{0,2\}$ with equal probability and $\mathcal{F}_n := \sigma(X_i^k : i \geqslant 1, k \leqslant n)$ then $Z_n := \sum_{i=1}^{Z_{n-1}} X_i^n$. Then $\mathbb{E}[Z_n|\mathcal{F}_{n-1}] = Z_{n-1}\mathbb{E}[X_1^n] = Z_{n-1}$ so Z_n is a martingale. The martingale Z_n is positive so by the martingale convergence theorem it converges to some Z_∞ a.s. Now we show that the limit must be 0. For any $k > 0$ we have that $\mathbb{P}(Z_{n+1} = k|Z_n = k) = 1/2$ and so

$$\mathbb{P}(Z_n = k; ...; Z_{n+N} = k) \leqslant 2^{-N}.$$

But now $\mathbb{P}(\lim Z_n = k) \leqslant 2^{-N}$ for each $N \in \mathbb{N}$.

(ii) The convergence again follows from non-negative martingale convergence. First consider the case $\mu < 1$. Then we have that $\mathbb{E}[Z_n] = \mu^n$ and so

$$\mathbb{P}(Z_n > 0) = \sum_{k\geqslant 1}\mathbb{P}(Z_n = k) \leqslant \sum_{k\geqslant 1}k\mathbb{P}(Z_n = k) = \mu^n.$$

By taking $n \to \infty$ we see that $Z_n = 0$ a.s.

Now for the case $\mu = 1$ we ignore the case $\mathbb{P}(Z_1 = 1) = 1$ otherwise the result does not hold, nor do we have any interesting activities. So then $p := \mathbb{P}(Z_1 = 0) > 0$ as the expectation is 1. Following the idea as above, let $k > 0$, then we have instead

$$\mathbb{P}(Z_n = k; ...; Z_{n+N} = k) \leqslant (1-p)^N$$

and hence $\mathbb{P}(\lim Z_n = k)$ which is the union of events of the form $\{\forall n \geqslant N, Z_n = k\}$ is zero.

(iii) Let $p = \mathbb{P}(M_\infty = 0)$. There are a possible number of cases to consider. If $p = 0$ or 1, then the result follows easily. If $0 < p < 1$ then on the set $\{M_\infty = 0\}$, $Z_n \to 0$ and hence $p^{Z_n} \to 1$. On the set $\{M_\infty > 0\}$ we have that as $\mu > 1$, $Z_n \to \infty$, now as $p < 1$ this implies that $p^{Z_n} \to 0$. Thus $p^{Z_n} \to \mathbf{1}_{M_\infty = 0}$. Z_n roughly behaves like $M_\infty \mu^n$ asymptotically.

c) First by the tower law

$$Var[Z_n] = \mathbb{E}\big[Var[Z_n^2|Z_{n-1}]\big] = \mathbb{E}[Z_{n-1}\sigma^2] = \mu^{n-1}\sigma^2$$

so then $Var(M_n) = \mu^{-n-1}\sigma^2$ which shows the bound in L^2. An application of Cauchy-Schwartz gives that $\mathbb{E}[Z_n \mathbf{1}_{Z_n>0}]^2 \leqslant \mathbb{P}(Z_n > 0)\mathbb{E}[Z_n^2]$ so that

$$\mathbb{P}(Z_n > 0) \geqslant \frac{\mathbb{E}[Z_n]^2}{\mathbb{E}[Z_n^2]} = \frac{\mu^{2n}}{\mu^{n-1}\sigma^2 + \mu^{2n}} \geqslant \frac{1}{1+\sigma^2} > 0.$$

Thus the probability of survival is strictly positive.

12. Take $\{X_i\}_{i\geqslant 1}$ to be i.i.d. Bernoulli $\{0, 2\}$, i.e. $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 2) = 1/2$. Consider $M_n := \prod_{i=1}^n X_i$, then we have that

$$\mathbb{E}[M_n - M_{n-1}|\mathcal{F}_{n-1}] = \mathbb{E}\big[(X_n - 1)M_{n-1}|\mathcal{F}_n\big] = \big(\mathbb{E}[X_n] - 1\big)M_{n-1} = 0$$

so M_n is a martingale. Next notice that $\mathbb{E}[M_n] = 1$ by the independence of the X_i, so that M_n cannot converge in L^1 to 0. On the other hand observe that

$$\mathbb{P}(M_n \neq 0) = \prod_{i=1}^n \mathbb{P}(X_i \neq 0) = \frac{1}{2^n}$$

and hence by Borel-Cantelli $M_n \to 0$ a.s.

13. First as M is bounded in L^1, it converges in L^1 to M_∞. On the event $\{T = \infty\}$, $M_T = M_\infty \in L^1$, and on the event $\{T < \infty\}$, by dominated convergence $\mathbb{E}\big[|M_T|\big] = \lim_{t\to\infty} \mathbb{E}\big[|M_{T\wedge t}|\big]$ and as $|M|$ is a submartingale $\mathbb{E}\big[|M_{T\wedge t}|\big] \leqslant \mathbb{E}\big[|M_t|\big] \leqslant \mathbb{E}\big[|M_\infty|\big]$.

For the counterexample take $M_t = B_t$ a standard Brownian motion and $T := \inf\{t \geqslant 0 : B_t = 1\}$, then $\mathbb{E}[B_T] = 1 \neq 0 = \mathbb{E}[B_0]$.

14. Let $\mathcal{F}_n := \sigma([a, b) : a, b \in \mathbb{D}_n)$, then F_n increases to $\mathcal{F}_\infty$ which is the Borel sigma-algebra. With this set up M_n is nothing but the projection of f' on to $\mathcal{F}_n$, i.e. $M_n = \mathbb{E}[f'|\mathcal{F}_n]$. Indeed for any $[a, b)$ being a basic set in $\mathcal{F}_n$, we have that $\int_a^b f_n'(x)dx = \int_a^b f'(x)dx$. So now by Lévy's upward theorem $M_n \to \mathbb{E}[f'|\mathcal{F}_\infty] = f'$ a.s and in L^1 as f' is continuous and hence Borel measurable.

15. Let e_n be an orthonormal basis of H, we wish to show that $\sum_{k=1}^n h_k G_k \to X_h$ in L^2 and a.s., where G_k are i.i.d. normal and $h = \sum h_n e_n$. We have seen before that $X_h \in L^2$. Let $\mathcal{F}_n := \sigma(G_1, ..., G_n)$, then consider the martingale $M_n := \mathbb{E}[X_h|\mathcal{F}_n] = \sum_{k=1}^n h_k G_k$. Now by theorem 72, the convergence is a.s. as well.

16. The Borel σ-algebra of $\mathcal{C}\big([0, 1], \mathbb{R}\big)$ is generated by the coordinate process, with elementary events $\{X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n\}$, for $0 \leqslant t_1 < \cdots < t_n \leqslant 1$ and B_i measurable subsets of $\mathbb{R}$. It is also generated by the events of the form $A = \{X_{t_1} - X_0 \in C_1\, X_{t_2} - X_{t_1} \in C_2, \ldots, X_{t_n} - X_{t_{n-1}} \in C_n\}$, for C_i measurable subsets of $\mathbb{R}$. Can you prove it? To prove that $\mathbb{P}^1$ is absolutely continuous with respect to $\mathbb{P}$ and find its Radon-Nikodym derivative, it suffices then to compare $\mathbb{P}^1(A)$ and $\mathbb{P}(A)$. Using the independence of the increments and their Gaussian nature, you can easily see that $\mathbb{P}^1(A) =$

$\mathbb{E}\big[e^{-aX_{t_n}-\frac{a^2t_n^2}{2}}\mathbf{1}_A\big] = \mathbb{E}\big[e^{-aX_1-\frac{a^2}{2}}\mathbf{1}_A\big]$, since the process $\big(e^{-aX_t-\frac{a^2t^2}{2}}\big)_{0\leqslant t\leqslant 1}$ is a martingale. It follows that $\frac{d\mathbb{P}^1}{d\mathbb{P}} = e^{-aX_1-\frac{a^2}{2}}$.

17. This is pretty much the same argument as Corollary 79. Let $\mathbb{P}$ be the uniform measure on $\mathfrak{S}_n$, X be the coordinate map and $\mathcal{F}_k := \sigma(X_1, ..., X_k)$. Then we are done if we can estimate $\big|\mathbb{E}[f|\mathcal{F}_{k+1}] - \mathbb{E}[f|\mathcal{F}_k]\big|$. Notice that $\mathbb{E}[f|\mathcal{F}_k]$ is the average of f on $\{\sigma : \sigma_i = x_i, i \leqslant k\}$ so that moving between the two averages, the function f could then at most differ by one change of coordinate, and hence $\big|\mathbb{E}[f|\mathcal{F}_{k+1}] - \mathbb{E}[f|\mathcal{F}_k]\big| \leqslant 1$ as f is a contraction. Hence by the Theorem 78, the result follows.

18. The idea is to construct a set which can be determined by $f(t)$ for any $t > 0$. So take $\{f : \inf\{t \geqslant 0 : f(t) \neq f(0)\} = 0\}$. Notice that

$$\{f : \inf\big\{t \geqslant 0 : f(t) \neq f(0)\big\} = 0\} = \bigcap_{t>0}\big\{f : f(t) \neq f(0)\big\} \in \mathcal{F}_t$$

for any $t > 0$. However this set cannot be in $\mathcal{F}_0$ as this cannot be determined by sets of the form $\big\{f : f(0) \in A\big\}$.

19. a) Take $A \in \cap_{n\geqslant 1}\sigma(\mathcal{G}, \mathcal{G}_n, ...)$ and consider $X := \mathbf{1}_A - \mathbb{E}[\mathbf{1}_A|\mathcal{G}]$. Now as $\mathbb{E}[X|\mathcal{G}] = 0$, X is independent of $\mathcal{G}$. By definition $X \in \sigma(\mathcal{G}, \mathcal{G}_n, ...)$ for each $n \geqslant 1$ and hence $X \in \sigma(\mathcal{G}_n, ...)$, therefore $X \in \cap_{n\geqslant 1}\sigma(\mathcal{G}_n, ...)$.[84]

Then Kolmogorov's 0-1 law gives that X is constant, but $\mathbb{E}[X] = 0$, so $X = 0$ a.s. In other words $\mathbf{1}_A = \mathbb{E}[\mathbf{1}_A|\mathcal{G}]$, i.e. there exists a set $B \in \mathcal{G}$ such that $\mathbf{1}_A = \mathbf{1}_B$ a.s.

b) By the independence of the increments $\mathcal{T}_n := \sigma\big(B_{t+1/n} - B_{t+\frac{1}{n+1}}\big)$ are independent from each other and from $\mathcal{G}_t$. Then as $\mathcal{G}_{t+} = \cap_{n\geqslant 1}\sigma\big(\mathcal{G}_t, \mathcal{T}_n, ...\big)$ from the previous part we have that $\mathcal{G}_t$ and $\mathcal{G}_{t+}$ coincide up to null events. The result now follows as they both contain all the null events.

15.3. **Exercises on part III. 1.** Set $\widetilde{B}_t = tB_{1/t}$ for $t > 0$ and $\widetilde{B}_0 = 0$. We know from proposition 94 that $\widetilde{B}$ is a Brownian motion. Also, as $\mathcal{F}^{\widetilde{B}_{0+}} = \mathcal{T}$, Blumenthal's 0-1 law applied to $\widetilde{B}$ shows that $\mathcal{T}$ is made up of trivial events *for* $\mathbb{P}$. (They might be non-trivial for a different probability!)

2. We proceed as in the proof of proposition 92, denoting by $\mathcal{C}$ the cone. As the event $\{\tau_U = 0\} \in \mathcal{F}_{0^+}$, it suffices to prove that $\mathbb{P}(\tau_U = 0) \geqslant c$ for some positive constant c to prove that it has probability 1, by Blumenthal's 0-1 law. Let $\epsilon > 0$ be given. As $\mathbb{P}(\tau_U \leqslant \epsilon) \geqslant \mathbb{P}(B_\epsilon \in \mathcal{C})$ and the law of B is invariant by rotations, we have $\mathbb{P}(B_\epsilon \in \mathcal{C}) = |A|\Big(\int \frac{e^{-r^2/2\epsilon}}{(2\pi\epsilon)^{d/2}}\mathbf{1}_{r\leqslant a}\, r^{d-1}dr\Big) = |A|\Big(\int \mathbf{1}_{u\leqslant a\epsilon^{-1/2}}\frac{2^{-u^2/2}}{(2\pi)^{d/2}}\, u^{d-1}du\Big) \geqslant c$, where $|A|$ is the surface of $A \subset \mathbb{S}^{d-1}$. Sending ϵ to 0 gives the conclusion.

3. We make the same reasonning as in exercise 10 in example sheet 2. Set $T = \min\{H_{-a}, H_b\}$ and write p for $\mathbb{P}(H_{-a} < H_b)$. As the *stopped processes* $(B_t)_{t\geqslant T}$ is a bounded martingale, the optionnal stopping theorem gives us: $0 = p(-a) + (1-p)b$, hence the value of p. Use the martingale $B_t^2 - t$ to compute $\mathbb{E}[T]$.

4. b) Given $0 \leqslant s < t$ and $A \in \mathcal{F}_s$, we have $\mathbb{E}\big[e^{\sigma B_t-\frac{\sigma^2}{2}t}\mathbf{1}_A\big] = \mathbb{E}\big[e^{\sigma B_s-\frac{\sigma^2}{2}s}\mathbf{1}_A\big]$ for all σ. Expanding the exponential on both sides in power series of σ, use the fact that

[84] This is true as the L^2 is the sum of orthogonal components.

$\mathbb{E}[B^{2k}] = \prod_{p=0}^{k-1}(2k-2p-1)$ (induction) to justify the interchange of $\mathbb{E}$ and $\sum_k$. The term $\mathbb{E}\big[(B_t^2 - t)\mathbf{1}_A\big]$ appears as the coefficient of σ on the lhs and the term $\mathbb{E}\big[(B_s^2 - s)\mathbf{1}_A\big]$ as the coefficient of σ on the rhs. Their identification gives the martingale property of the first process as we can take any $0 \leqslant s < t$ and $A \in \mathcal{F}_s$. Look at the coefficients of σ^2 and σ^3 to obtain the martingale property of the two other processes.[85]

5. a) We need θ to satisfy $\lambda - \theta c = \frac{\theta^2}{2}$, that is $\theta = \sqrt{c^2 + 2\lambda} - c$, since it is positive.

b) As the *stopped martingale* $\left(e^{\theta B_t^c - \lambda t}\right)_{0 \leqslant t \leqslant T}$ is bounded, the optionnal stopping theorem implies: $1 = \mathbb{E}\big[e^{\theta x - \lambda H_x^c}\big]$, hence the formula. We check that this function of $\lambda > 0$ coincides with the Laplace transform of the given density. As the Laplace transform characterizes the distribution H_x^c has the mentionned density/

c) It suffices to let λ decrease to 0.

6. b) Recall the strong Markov property: *Given any finite stopping time T, the process $(B_{T+t} - B_T)_{t \geqslant 0}$ is a Brownian motion independent of $\mathcal{F}_T$.* Apply it to T_a for some $a \geqslant 0$. The fact that it is a Brownian motion says that if $b > a$ then $T_b - T_a$ is distributed as T_{b-a}, giving the stationnarity of the process $(T_a)_{a \geqslant 0}$. The independence of the increments comes from the second piece of information provided by the strong Markov property: the independence of $(B_{T+t} - B_T)_{t \geqslant 0}$ with respect to $\mathcal{F}_T$. Given $a_1 < a_2 < \cdots < a_n$, a straightforward induction enables to prove that the increments $T_{a_2} - T_{a_1}, \ldots, T_{a_n} - T_{a_{n-1}}$ are independent. It is not a Lévy process though, as it is not càdlàg but continuous on the *left* with right limits. Prove it!

7. a) We have $T_a \leqslant S_a$ and $S_a = T_a + \inf\{t \geqslant 0\,;\, B_{t+T_a} - B_{T_a} > 0\}$. The strong Markov property gives $S_a = T_a$, almost-surely.

c) Take for L the time in $[0,1]$ where B_t is maximum. Prove that it is almost-surely < 1. It follows that we have almost-surely $S_L \geqslant 1$ and $S_L \neq T_L$.

8. a) Set $T_0 = 0$ and define inductively $S_n = \inf\{t \geqslant T_{n-1}\,;\, B_t \in D\}$ and $T_n = \inf\{t \geqslant S_n\,;\, B_t \notin B(0,2r)\}$. By the strong Markov property and the invariance of the law of Brownian motion by rotations, the random variables $\int_{S_k}^{T_k} \mathbf{1}_D(B_s)\,ds$ are iid. As they have positive mean, the strong law of large numbers gives $\int_0^\infty \mathbf{1}_D(B_s)\,ds \geqslant \sum_{n=0}^\infty \int_{S_k}^{T_k} \mathbf{1}_D(B_s)\,ds = \infty$, almost-surely.

b) Denote by $p_t(x,y)$ the transition kernel of Brownian motion. By Fubini's theorem, we have $\mathbb{E}_x\big[\int_0^\infty f(B_t)\,dt\big] = \int \Big(\int_0^\infty p_t(x,y)\Big) f(y)\,dy$, for any non-negative function f. The time integral equals $|y-x|^{2-d}$ up to a multiplicative constant C. (Do the computation! We see why we need $d \geqslant 3$.) This function of y is locally integrable with respect to y.[86]

9. We know from exercise 7 the distribution of T. As it is independent of B^1, we have

$$\mathbb{E}\big[f(B_T^1)\big] = \int_0^\infty \frac{2^{-\frac{1}{2t}}}{\sqrt{2\pi t^3}} f(B_t^1)\,dt = \int f(x)\Big(\int_0^\infty \frac{2^{-\frac{1}{2t}}}{\sqrt{2\pi t^3}} \frac{2^{-\frac{x^2}{2t}}}{\sqrt{2\pi t}}\,dt\Big)\,dx = \int f(x) \frac{dx}{\pi(1+x^2)},$$

[85] Note that I have not tried to work directly with the conditionnal expectation identity $\mathbb{E}\big[e^{\sigma B_t - \frac{\sigma^2}{2}t} \big| \mathcal{F}_s\big] = e^{\sigma B_s - \frac{\sigma^2}{2}s}$ as this identity involves random variables defined only almost-surely, so it is not obvious how to differentiate with respect to σ in a mathematically neat way.

[86] If x is not in the domain of integration, no problem; otherwise, use polar coordinates near x.

for any bounded measurable function $f : \mathbb{R} \to \mathbb{R}$. We read the distribution of B_T^1 above: it is a Cauchy random variable.

10. The process $M_t = |B_t|^2 - t\,d$ $\big(= \sum_{i=1}^d |B_t^i|^2 - t, \text{sum of independent martingales}\big)$ is a martingale. We would like to use the optionnal stopping theorem to the stopped martingale $(M_t)_{t\leqslant T}$; yet this process is not bounded, so it is convenient ot replace first T by $T\wedge n$ (rather than proving for instance that $(M_t)_{t\leqslant T}$ is uniformly integrable, which can be done). The new stopped martingale is bounded. So we have $|x|^2 = \mathbb{E}\big[|B_{T\wedge n}|^2 - d(T\wedge n)\big]$, that is $\mathbb{E}\big[T\wedge n\big] = \frac{\mathbb{E}\left[|B_{T\wedge n}|^2\right]-|x|^2}{d}$. Use monotone convergence on the lhs, and dominated convergence on the rhs, to conclude by sending n to infinity.

11. Suppose g has a maximum M at a point x_0 inside O. As it has the mean value property, g needs to be equal to M near x_0; this shows that the closed set where g attains its maximum is also open. As O is connected, g is constant, equal to its maximum, on the whole of O.

Would a given Dirichlet problem have two solutions, their difference would be a solution to the Dirichlet problem with null boundary condition, so would have a null maximum. As the opposite of this difference is also a solution, it would also have a null maximum, leading to the equality of the two functions.

13. Let denote by $(N_t)_{t\geqslant 0}$ a Poisson process of intensity λ and jump measure J. Can you see why it suffices to consider the case where $J(\cdot) = \delta_1(\cdot)$? In that case, we need to prove that given any $n \geqslant 1$, any times $t_1 < \cdots < t-n$ and any integers $i_1, \dots, i_n$, we have

$$\mathbb{P}\big(N_{t_2} - N_{t_1} = i_1, \dots, N_{t_n} - N_{t_{n-1}} = i_{n-1}\big) = \prod_{k=1}^{n-1} \frac{\big(\lambda(t_k - t_{k-1})\big)^{i_k}}{i_k!} e^{-\lambda(t_k - t_{k-1})}.$$

We proceed by induction on $n \geqslant 1$. The case $n = 1$ is treated in exercise 12. To make the induction step, it suffices to prove that

(15.1)

$$\mathbb{P}\Big(N_{t_{n+1}} - N_{t_n} = i_n \Big| N_{t_k} - N_{t_{k-1}} = i_{k-1},\ \text{for } k = 1..n\Big) = \frac{\big(\lambda(t_n - t_{n-1})\big)^{i_n}}{i_!} e^{-\lambda(t_{n+1} - t_n)}$$

Set $i = i_1 + \cdots + i_{n-1}$ and denote by H_i the hitting time of $\{i\}$ by the process N. Then, conditionally on the event $\{H_i < t_{n-1} < H_i + S_i\}$, time $H_i + S_i - t_{n-1}$ to wait after t_{n-1} before the next jump is exponentially distributed, with parameter λ, by the memoryless property of S_i. Identity (15.1) follows as $\mathbb{P}\big(N_{t_{n+1}} - N_{t_n} = i_n \big| N_{t_k} - N_{t_{k-1}} = i_{k-1},\ \text{for } k = 1..n\big) = \mathbb{P}\big(N_{t_{n+1}} - N_{t_n} = i_n \big| H_i < t_{n-1} < H_i + S_i\big)$, by the strong Markov property of the Markov chain $(N_t)_{t\geqslant 0}$.

14. Denote by S_1 the first holding time. The obvserver is proved wrong if at some time t he observes that $\{N_t = N_{t-S_1}\}$. Given $s > 0$, let define the stopping time $T_s = \inf\{t \geqslant s\,;\, N_t = N_{t-s}\}$ – with respect to which filtration? Then, conditionning on the first jump, the strong Markov property gives

$$\mathbb{E}\big[T_s\big] = se^{-\lambda s} + \int_0^s \big(a + \mathbb{E}[T_s]\big)\lambda e^{-\lambda a}\,da,$$

so $\mathbb{E}\big[T_s\big] = \frac{e^{\lambda S_1 - 1}}{\lambda}$. The mean time until one sees a holding time bigger than S_1 is thus

$$\int_0^{\infty} \big(s + \mathbb{E}[T_s]\big)\, \lambda e^{-\lambda s}\, ds = \infty.$$

15. a) It suffices to prove, for all $n \geqslant 1$, that $S_1 + \cdots + S_n$ is almost-surely different from t. (Can you see why?) This follows from the fact that the random variable $S_1 + \cdots + S_n$ has a density with respect to Lebesgue measure on $\mathbb{R}_+$.

b) Note that

$$\begin{aligned}\mathbb{P}(T_t > t+s) &= \sum_{k=1}^{\infty} \mathbb{P}(N_t = k, N_{t+s} = k) = \sum_{k=1}^{\infty} \mathbb{P}(N_t = k, N_{t+s} - N_t = 0) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(N_t = k)\, \mathbb{P}(, N_{t+s} - N_t = 0) = \sum_{k=1}^{\infty} \mathbb{P}(N_t = k)\, e^{-\lambda s} = e^{-\lambda s}.\end{aligned} \tag{15.2}$$

So $T_t - t$ is exponentially distributed, with parameter λ.

16. It's even worse! The sum of two Brownian motions can be non-Brownian! To see that, let us work on the subset Ω of $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^2)$ made up of paths starting from 0, equipped with its Borel σ-algebra. Let X be the coordinate process $X_t(\omega) = \omega(t) = (\omega_1(t), \omega_2(t)) \in \mathbb{R}^2$, for $\omega \in \Omega$, and let $\mathbb{P}$ be Wiener measure. Let $\mathbb{P}'$ be the measure on $(\Omega, \mathcal{F})$ under which X is a Wiener measure with correlation -1. Let $\mathbb{Q} = \frac{\mathbb{P}+\mathbb{P}'}{2}$. I let you prove that the processes ω_1 and ω_2 are Wiener processes under $\mathbb{Q}$. Can you prove by a simple calculation that the process $\omega + \omega_2$ is not Gaussian? As Brownian motion with dift (a Gaussian process!) is the only continuous Lévy process (see exercise 21), this proves the claim.

17. Let[87] Ω be an arbitrary space and $\mathcal{F}$ be the trivial σ-algebra over it. (We work with deterministic processes!). Let also $(x_\alpha)_\alpha$ be a Hamel basis of Haar over the rational numbers. For every $t \geqslant 0$, let X_t be the sum of the coordinates of t in the Hamel basis. As $X_{t+s} = X_t + X_s$, the process X has stationnary independent increments. As X is highly discontinuous (it takes values in $\mathbb{Q}$!), it does not have a modification which is càdlàg .

18. For $s < t$, we have $\mathbb{E}\big[e^{i\lambda(X_t - X_s)}\big] = e^{(t-s)g(\lambda)}$. Senging $s \uparrow t$ we conclude that $\mathbb{E}\big[e^{i\lambda(\Delta X_t)}\big] = 1$, so ΔX_t has the same fistribution as the constant 0, that is ΔX_t is almost-surely null.

19. a) We know, from the general construction of Lévy processes given in the course, that X has the same law as the sum of a difted Brownian motion, an independent Poisson process with finite intensity, and a infinite sum of independent compensated Poisson processes. (This sum takes care of the fact that the jump measure can have an infinite mass.) In the case of a finite jump measure, only the first two termsare needed; as Poisson processes have almost-surely finitely many jumps in any finite time interval, we are done.

b) We can forget the continuous part (drifted Brownian motion) and work only with the Poisson process. Let S_i, J_i be the successive holding and jump times of the process; they are all independent. By construction, the process Y is constructed out of the sequence of jump times $\big((S_1 + \cdots + S_i)\mathbf{1}_{\epsilon_i = 1}\big)_{i \geqslant 1}$ and the corresponding jumps. The time between two jumps will have the same law as $S_1 + \cdots + S_N$, where N is a geometrical random variable with parameter p. A straightforward computation shows that this random sum

[87] This solution is taken from the excellent book [Med07] by P. Medvegyev.

with exponentially distributed, with parameter $p\lambda$. So Y is a Lévy process with jump measure $p\Lambda_X$.

20 - 21. Copy word by word what has been done previously elsewhere.

22. See for instance theorem (28.12), p. 76, in [RW00]

References

[BMP02] P. Baldi, L. Mazliak, and P. Priouret. *Martingales and Markov chains.* Chapman & Hall/CRC, Boca Raton, FL, 2002. Solved exercises and elements of theory, Translated from the 1998 French original.

[Chu02] Kai Lai Chung. *Green, Brown, and probability & Brownian motion on the line.* World Scientific Publishing Co. Inc., River Edge, NJ, 2002.

[Doo94] J. L. Doob. *Measure theory*, volume 143 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1994.

[Dud02] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[DY79] E. B. Dynkin and A. A. Yushkevich. *Controlled Markov processes*, volume 235 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 1979. Translated from the Russian original by J. M. Danskin and C. Holland.

[GS04] I. I. Gikhman and A. V. Skorokhod. *The theory of stochastic processes. I.* Classics in Mathematics. Springer-Verlag, Berlin, 2004. Translated from the Russian by S. Kotz, Reprint of the 1974 edition.

[IW89] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library.* North-Holland Publishing Co., Amsterdam, second edition, 1989.

[JS03] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, second edition, 2003.

[Kal02] O. Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[Kry02] N. V. Krylov. *Introduction to the theory of random processes*, volume 43 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2002.

[Med07] P. Medvegyev. *Stochastic integration theory*, volume 14 of *Oxford Graduate Texts in Mathematics.* Oxford University Press, Oxford, 2007.

[RW00] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 1.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Foundations, Reprint of the second (1994) edition.

[Sat99] K. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 1999. Translated from the 1990 Japanese original, Revised by the author.

[Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, second edition, 1996. Translated from the first (1980) Russian edition by R. P. Boas.

[Sto87] J. M. Stoyanov. *Counterexamples in probability.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester, 1987.

[Wil91] D. Williams. *Probability with martingales.* Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.

Index

0-1 law, 50

Adapted process, 33

Backward martingale, 38
Borel σ-algebra , 6
Borel space, 9

Characteristic exponent of a Lévy process, 57
Closed martingale, 36
Conditional expectation, 30

Dirichlet problem, 54

Entrance and hitting times, 34
Equivalence of measures, 42

Filtration, 33
Finite dimensional convergence, 22
Finite-dimensional laws, 7

Harmonic function, 54

Index set, 5
Indistinguishability, 14
Isomorphism, 9

Lévy process, 57
Lévy-Khinchin theorem, 58

Martingale, 35
Modification, 14

Previsible process, 33
Process, 5
Product σ-algebra , 6
Projective sequence of proability measures, 12

Quadratic variation of Brownian motion, 51

Reflection principle, 54
Regular conditional probability, 47
Regular/irregular boundary point, 55

Simple Markov property, 50
Stopped process, 34
Stopping time, 33
Strong Markov property, 53

Tightness, 20

Uniform integrability, 32
Usual conditions for a filtration, 46

Weak topology, 16
Wiener measure, 11