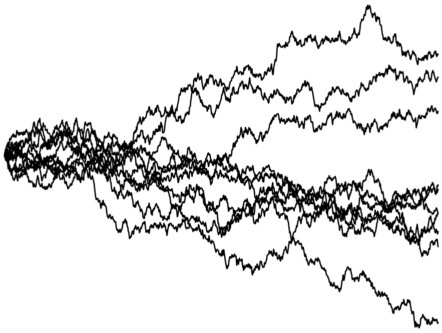
A review of probability theory foundations

Rough lecture notes



Master of mathematics
Université Paris-Dauphine – PSL

2021 – 2022
Preliminary version
Compiled September 20, 2021

Course home page on https://djalil.chafai.net/

Suggested schedule of the lectures.

We recommend that the in class or oral lectures differ from the lecture notes, ideally they should contain less details and should be focused on the essential aspects, the structure, the culture, and the intuition.

- Day 1 (2 x 1.5h)
  Chapter 1 (Preliminaries)
- Day 2 (2 x 1.5h)
  Chapter 2 (Processes, filtrations, stopping times, martingales)
- Day 3 (2 x 1.5h)
  Chapter 3 (Brownian motion)
- Day 4 (2 x 1.5h)
  Chapter 3 (Brownian motion)
- Day 5 (2 x 1.5h)
  Chapter 4 (More on martingales)

These are the lecture notes of the pre-school week, given at Université Paris-Dauphine – PSL, for second year master students in mathematics[1]. The objective of this pre-school is to recall fundamental aspects of probability theory at the level of a first year of master. This helps to prepare the students for the courses proposed in the master in particular to the course on stochastic calculus. Before the year 2021–2022, this pre-school was essentially the first part of the stochastic calculus course. What was removed (actually postponed to the stochastic calculus course) is essentially the Wiener integral, the Cameron – Martin formula, and its application to exit times, and the Kunita-Watanabe inequality.

There are many references on the subject. For probability theory at the level of the first year of master, there is for instance [10] by Allan Gut and [3] by Vivek Borkar. See also [11, 2]. Brownian motion and continuous martingales are studied for instance in the books by Daniel Revuz and Marc Yor [17], Jean-François Le Gall [14], and Fabrice Baudoin [1]. A historical reference for the foundations of probability theory and stochastic processes is the treatise by Claude Delacherie and Paul-André Meyer [7, 6, 5].

Contributors.

- 2018 – 2022 : Djalil Chafaï
- – 2018 : Halim Doss

Typos hunters.

- 2021 – 2022 : Pauline Amrouche, Faniriana Rakoto Endor
- 2020 – 2021 : Oskar Bataillon, Yi Han, Qiaoyu Luo, Gabriel Moreira-Nogueira, Diego Alejandro Murillo Taborda, Lyes Tifoun, Walid El Wahabi
- 2019 – 2020 : Oscar Cosserat, Łukasz Mądry, Alejandro Rosales Ortiz, Ziyu Zhou
- 2018 – 2019 : Clément Berenfeld

[1] MASEF (Mathématiques pour l'économie et la finance) and MATH (Mathématiques appliquées et théoriques).

Notation.

$\mathbb{R}_+$	$[0,+\infty)$
BM	Brownian motion
O, o	Landau notation
iff	if and only if
a.s.	almost surely
u.i.	uniformly integrable
w.r.t.	with respect to
$\mathbf{1}_A$	indicator of A
$x\cdot y$ or $\langle x,y\rangle$	$x_1y_1+\cdots+x_dy_d$ if $x,y\in\mathbb{R}^d$
$\lvert x\rvert$	$\sqrt{x_1^2+\cdots+x_d^2}$ if $x\in\mathbb{R}^d$
$\mathscr{B}_E$	Borel σ-algebra of E
e	exponential
d	differential element
i	the complex number $(0,1)$
d,i,j,k,m,n,ℓ	integer numbers
$p,q,r,s,t,u,v,\alpha,\beta,\varepsilon$	real numbers
$s\wedge t$ and $s\vee t$	$\min(s,t)$ and $\max(s,t)$
f is increasing	$f(y)\geq f(x)$ if $y\geq x$
$\mathrm{L}^p_{\mathbb{R}^d}(\Omega,\mathbb{P})$	$X:\Omega\to\mathbb{R}^d$ measurable with $\mathbb{E}(\lVert X\rVert^p)<\infty$
$\langle x,y\rangle_{\mathrm{H}}$	scalar product in the Hilbert space H
$\langle M,N\rangle$	angle bracket of local martingales M,N
$\langle M\rangle$	$\langle M,M\rangle$
$[M,N]$	square bracket of local martignales M,N
$[M]$	$[M,M]$
$X\sim\mu$	X has law μ

$$\Big(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geq0},\mathbb{P}\Big)$$

Unless otherwise stated, the random variables
and stochastic processes considered in this course
are defined on this enormous filtered probability space
and moreover the filtration is complete and right continuous.

Some of the scientists related to Brownian motion and stochastic calculus

Life time	Scientist
(1975 –)	Martin Hairer
(1968 –)	Wendelin Werner
(1959 –)	Jean-François Le Gall
(1955 –)	Alain-Sol Sznitman
(1954 –)	Dominique Bakry
(1953 –)	Terry Lyons
(1951 –)	David Nualart
(1949 – 2004)	Marc Yor
(1947 –)	Shige Peng
(1947 –)	Étienne Pardoux
(1944 –)	Nicole El Karoui
(1944 –)	Jean Jacod
(1942 – 2004)	Catherine Doléans-Dade
(1940 –)	S. R. Srinivasa Varadhan
(1940 –)	Daniel W. Stroock
(1938 –)	Mark Iosifovich Freidlin
(1938 – 1995)	Fischer Black
(1935 –)	Shinzo Watanabe
(1934 –)	Albert Shiryaev
(1934 – 2003)	Paul-André Meyer
(1930 –)	Henry McKean
(1930 – 2011)	Anatoliy Skorokhod
(1930 – 1997)	Ruslan Stratonovich
(1927 – 2013)	Donald Burkholder
(1925 – 2010)	Paul Malliavin
(1924 – 2014)	Eugene Dynkin
(1923 – 2020)	Freeman Dyson
(1916 – 2008)	Gilbert Hunt
(1915 – 2008)	Kiyosi Itô
(1915 – 1940)	Wolfgang Doeblin
(1914 – 1984)	Mark Kac
(1911 – 2004)	Shizuo Kakutani
(1910 – 2004)	Joseph Leo Doob
(1908 – 1989)	Robert Horton Cameron
(1906 – 1970)	William Feller
(1903 – 1987)	Andrey Kolmogorov
(1900 – 1988)	George Uhlenbeck
(1896 – 1971)	Paul Lévy
(1894 – 1964)	Nobert Wiener
(1879 – 1955)	Albert Einstein
(1875 – 1941)	Henri Lebesgue
(1872 – 1946)	Paul Langevin
(1872 – 1917)	Marian Smoluchowski
(1871 – 1956)	Émile Borel
(1870 – 1942)	Jean Baptiste Perrin
(1870 – 1946)	Louis Bachelier
(1856 – 1922)	Andrey Markov
(1856 – 1894)	Thomas Joannes Stieltjes
(1773 – 1858)	Robert Brown

Uhlenbeck's attitude to Wiener's work was brutally pragmatic and it is summarized at the end of footnote 9 in his paper (written jointly with Ming Chen Wang) "On the Theory of Brownian Motion II" (1945): the authors are aware of the fact that in the mathematical literature, especially in papers by N. Wiener, J. L Doob, and others [cf. for instance Doob (Annals of Mathematics 43, 351 1942) also for further references], the notion of a random (or stochastic) process has been defined in a much more refined way. This allows [us], for instance, to determine in certain cases the probability that the random function y(t) is of bounded variation or continuous or differentiable, etc. However it seems to us that these investigations have not helped in the solution of problems of direct physical interest and we will therefore not try to give an account of them.

Stanisław Ulam (1909 – 1984) about George Uhlenbeck
in *Adventures of a Mathematician* (1984).
This was before the completion of the theory of stochastic processes and stochastic calculus,
its numerical applications, and the rise of nowadays mathematical finance which is based on it.
About Brownian motion across physics and mathematics, the reader may take a look at [12, 18, 8, 15, 4].

"... Ainsi l'intégrale et les processus d'Itô, lointains descendants de la théorie de la spéculation de Bachelier, retournent à la spéculation financière. Ils méritent à tous égards d'être intégrés dans la culture générale des mathématiciens."

Jean-Pierre Kahane, *Le mouvement brownien.*
Un essai sur les origines de la théorie mathématique
Société Mathématique de France, 1998.

Contents

Chapter 1

Preliminaries

We refer to [9] and to [10] for the essential basic notions of probability theory (and more).

1.0 Sigma-algebras, random variables, and probabilities

A σ-field or σ-algebra – tribu in French – on a set Ω is a collection of subsets $\mathscr{A} \subset \mathscr{P}(\Omega)$ such that

- $\Omega \in \mathscr{A}$
- for all $A \in \mathscr{A}$, we have $A^c \in \mathscr{A}$
- for all at most countable family $(A_n)_n$ in $\mathscr{A}$, we have $\cap_n A_n \in \mathscr{A}$

where $A^c = \Omega \setminus A$. By combining these properties we also get $\varnothing \in \mathscr{A}$ and $\cup_n A_n \in \mathscr{A}$. We say that the couple $(\Omega, \mathscr{A})$ is a measurable space. Extreme examples of σ-algebras are $\mathscr{P}(\Omega)$ and $\{\varnothing, \Omega\}$.

- The intersection of an arbitrary family of σ-algebras is a σ-algebra.
- The σ-algebra generated by a subset of $\mathscr{P}(\Omega)$ is the $\cap$ of all the σ-algebras containing the subset.
- If Ω is equipped with a topology $\mathscr{T}$, the σ-algebra generated by $\mathscr{T}$ is called the Borel σ-algebra $\mathscr{B}$[1].

A map $f : \Omega \to E$ where $(\Omega, \mathscr{F})$ and $(E, \mathscr{E})$ are measurable is measurable when $f^{-1}(B) \in \mathscr{F}$ for all $B \in \mathscr{E}$.

A (positive) measure on a measurable space $(\Omega, \mathscr{A})$ is a map $\mu : \mathscr{A} \to [0, +\infty]$ such that

- $\mu(\varnothing) = 0$
- for all at most countable family $(A_n)_n$ of parwise disjoint elements of $\mathscr{A}$, we have $\mu(\cup_n A_n) = \sum_n \mu(A_n)$.

The triplet $(\Omega, \mathscr{A}, \mu)$ is a measured space. The measure μ is a probability measure when $\mu(\Omega) = 1$, and in this case the triplet $(\Omega, \mathscr{A}, \mu)$ is then a probability space.

A random variable X taking values in a measurable space $(E, \mathscr{E})$ is a measurable map defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$. By default we always assume that there is an underlying probability space $(\Omega, \mathscr{A}, \mathbb{P})$.

1.1 Expectation and law

If $X = \mathbf{1}_A$ then $\mathbb{E}(X) = \mathbb{P}(A)$, by linearity and monotone convergence this allows to define $\mathbb{E}(X) \in [0, +\infty]$ when X takes its values in $[0, +\infty]$. Next L^1 is the set of random variables such that $\mathbb{E}(|X|) < \infty$. If $X = X_+ - X_-$ then $|X| = X_+ + X_- \in \mathrm{L}^1$ if and only if $X_\pm \in \mathrm{L}^1$, and then we have $\mathbb{E}(|X|) = X_+ - X_-$ and $\mathbb{E}(X) = \mathbb{E}(X_+) - \mathbb{E}(X_-)$.

The law $\mathbb{P}_X$ of a real random variable X is characterized by

- distribution: $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$ for all $B \in \mathscr{B}_{\mathbb{R}}$
- cumulative distribution function: $F_X(x) = \mathbb{P}_X((-\infty, x])) = \mathbb{P}(X \le x)$ for all $x \in \mathbb{R}$
- characteristic function: $\varphi_X(t) = \mathbb{E}_{\mathbb{P}_X}(\mathrm{e}^{\mathrm{i}t\bullet}) = \mathbb{E}(\mathrm{e}^{\mathrm{i}tX})$ for all $t \in \mathbb{R}$,

[1] Further reading: `https://djalil.chafai.net/blog/2016/03/21/integration-alpha-et-omega/`

- Laplace transform (when $X \geq 0$): $L_X(t) = \mathbb{E}_{\mathbb{P}_X}(\mathrm{e}^{-t\bullet}) = \mathbb{E}(\mathrm{e}^{-tX})$ for all $t \geq 0$.

More generally, for a random variable $X : (\Omega, \mathscr{A}) \to (E, \mathscr{B})$, the law $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ of X is a probability measure on $(E, \mathscr{B})$. This infinite dimensional dual functional object is characterized by considering its values on a sufficiently large family of test functions such as, when $(E, \mathscr{B}) = (\mathbb{R}, \mathscr{B})$, $\mathbf{1}_{(-\infty,x]}$, $x \in \mathbb{R}$, or $\mathrm{e}^{\mathrm{i}t\bullet}$, $t \in \mathbb{R}$, etc.

1.2 Independence

1. A family $(\mathscr{A}_i)_{i\in I}$ of sub-σ-algebras of $\mathscr{A}$ is independent when for all finite $J \subset I$ and all $A_i \in \mathscr{A}_i$ we have

$$\mathbb{P}(\cap_{i\in J} A_i) = \prod_{i\in J} \mathbb{P}(A_i).$$

2. We say that a family $(X_i)_{i\in I}$ of random variables is independent, $X_i : (\Omega, \mathscr{A}) \mapsto (E_i, \mathscr{B}_i)$, when the family of sub-σ-algebras $(\sigma(X_i))_{i\in I}$ is independent, where

$$\sigma(X_i) = \{X_i^{-1}(B) : B \in \mathscr{B}_i\}$$

is the σ-algebra generated by X_i. Thus $(X_i)_{i\in I}$ is independent iff for all $J \subset I$ finite,

$$\mathbb{P}_{X_i : i\in J} = \otimes_{i\in J} \mathbb{P}_{X_i} \quad \text{on} \quad (\prod_{i\in J} E_i, \otimes_{i\in J} \mathscr{B}_i).$$

It follows that if $X_1, X_2, \ldots, X_n$ are real random variables integrable and independent then

$$\prod_{i=1}^n X_i \in \mathrm{L}^1 \quad \text{and} \quad \mathbb{E}\big(\prod_{i=1}^n X_i\big) = \prod_{i=1}^n \mathbb{E}(X_i).$$

1.3 Markov, Cauchy – Schwarz, Hölder, Jensen, convergence, Borel – Cantelli, LLN, LIL, CLT, ...

Markov inequality. If $U(X) \geq 0$ for a non-decreasing function U then for all $r > 0$,

$$\mathbb{P}(X \geq r) \leq \frac{\mathbb{E}(U(X))}{U(r)}.$$

This allows to control tails with moments. Conversely, we can control moments by tails via

$$\mathbb{E}(U(|X|)) = U(0) + \int_0^\infty U'(t)\mathbb{P}(|X| \geq t)\mathrm{d}t.$$

Cauchy – Schwarz inequality. In $[0, +\infty]$, with equality if and only if X and Y are colinear,

$$\mathbb{E}(XY) \leq \mathbb{E}(|X|^2)^{1/2}\mathbb{E}(|Y|^2)^{1/2}.$$

Hölder inequality. If $p \in [1, \infty]$ and $q = 1/(1-1/p) = (p-1)/p$ then, in $[0, +\infty]$,

$$\mathbb{E}(|XY|) \leq \mathbb{E}(|X|^p)^{1/p}\mathbb{E}(|Y|^q)^{1/q}.$$

Jensen inequality. If $U : \mathbb{R}^d \to \mathbb{R}$ is convex and $X \in \mathrm{L}^1$ with $U(X) \in \mathrm{L}^1$ then

$$U(\mathbb{E}(X)) \leq \mathbb{E}(U(X)),$$

moreover when U is strictly convex then equality is achieved only if X is (almost surely) constant. Useful examples include $U(x) = x^p$, $p \geq 1$, $U(x) = \mathrm{e}^{cx}$, $c \in \mathbb{R}$, $U(x) = +\infty\mathbf{1}_{x<0} + x\log(x)\mathbf{1}_{x\geq 0}$.

Convergences. Below $(X_n)_{n\geq 1}$, $(Y_n)_{n\geq 1}$, X, Y are real random variables on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$, of law μ_n, ν_n, μ, ν and cumulative distribution function F_n, G_n, F, G respectively.

Almost sure convergence. We say that $X_n \xrightarrow{\text{a.s.}} X$ when

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$$

in other words $\mathbb{P}(\{\omega\in\Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$. This is the notion of convergence in the SLLN.

Convergence in probability. We say that $X_n \xrightarrow{\mathbb{P}} X$ when

$$\forall\varepsilon>0,\quad \lim_{n\to\infty}\mathbb{P}(|X_n - X|\geq\varepsilon) = 0$$

which means that $\forall\varepsilon>0$, $\lim_{n\to\infty}\mathbb{P}(\{\omega\in\Omega : |X_n(\omega)-X(\omega)|\geq\varepsilon\}) = 0$. This is used in the weak LLN.

Mean convergence. For all $p\in[1,\infty)$, we say that $X_n \xrightarrow{\mathrm{L}^p} X$ when

$$X\in\mathrm{L}^p \quad\text{and}\quad \lim_{n\to\infty}\mathbb{E}(|X_n - X|^p) = 0.$$

The most useful cases are $p\in\{1,2,4\}$.

Convergence in law. The following properties are equivalent and we say then that $X_n \xrightarrow{\text{law}} X$, or $X_n \xrightarrow{\text{d}} \mu$ (convergence in distribution), or $\mu_n \xrightarrow{\text{nar.}} \mu$ (narrow convergence). This is used in the CLT.

1. $\lim_{n\to\infty}\mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ for all bounded and continuous $f:\mathbb{R}\to\mathbb{R}$
2. $\lim_{n\to\infty}\mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ for all $\mathscr{C}^\infty$ and compactly supported $f:\mathbb{R}\to\mathbb{R}$
3. *Cumulative distribution function.* $\lim_{n\to\infty}\mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ for all $f=\mathbf{1}_{(-\infty,x]}$ with x continuity point of $\mathbb{P}(X\leq\bullet)$, in other words $F_n(x)=\mathbb{P}(X_n\leq x)\to F(x)=\mathbb{P}(X\leq x)$ as soon as F is continuous at x
4. *Fourier transform or characteristic function.* $\lim_{n\to\infty}\mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ for all $f=\mathrm{e}^{\mathrm{i}t\bullet}$, $t\in\mathbb{R}$
5. *Laplace transform.* **(on $\mathbb{R}_+$)** $\lim_{n\to\infty}\mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ for all $f=\mathrm{e}^{-t\bullet}$, $t\geq 0$.

Contrary to the other modes of convergence, the convergence in law does not depend on the law of the couple (X_n, X) and uses only marginal laws. The Fourier and Laplace transforms convert sums of independent random variables into products, for which the expectation is the product of expectations.

Apart the convergence in law, the other modes of convergence are stable by finite linear combinations. The almost sure convergence, the convergence in probability, and the convergence in law are stable by composition with continuous functions, and this is referred to sometimes as the continuous mapping theorem.

The notions of convergence extend naturally to random vectors by using a distance/norm/scalar product, for instance for the characteristic function by replacing $\mathrm{i}tX$ by $\mathrm{i}\langle t, X\rangle$.

$$\begin{array}{ccccc} & & \boxed{\mathrm{L}^p\text{ CV}} & & \\ & & \Downarrow & & \\ & & \boxed{\mathrm{L}^1\text{ CV}} & & \\ & & \Downarrow & & \\ \boxed{\text{a.s. CV}} & \Rightarrow & \boxed{\text{CV in }\mathbb{P}} & \Rightarrow & \boxed{\text{CV in law}} \end{array}$$

If X is constant then the convergence in law implies the convergence in probability. The convergence in L^1 is equivalent to uniform integrability and convergence in probability.

Monotone convergence theorem. If $(X_n)_{n\geq1}$ takes its values in $[0,+\infty]$ and $\nearrow$ then

$$\mathbb{E}(\lim_{n\to\infty} X_n) = \lim_{n\to\infty}\mathbb{E}(X_n)\in[0,+\infty].$$

Fatou lemma. If $(X_n)_{n\geq1}$ takes its values in $[0,+\infty]$ then

$$\mathbb{E}(\varliminf_{n\to\infty} X_n) \leq \varliminf_{n\to\infty}\mathbb{E}(X_n)\in[0,+\infty].$$

Dominated convergence theorem. If $X_n \xrightarrow{\text{a.s.}} X$ and $\sup_n|X_n|\leq Y$, $\mathbb{E}(Y)<\infty$, then

$$\lim_{n\to\infty}\mathbb{E}(X_n) = \mathbb{E}(\lim_{n\to\infty} X_n) = \mathbb{E}(X).$$

The dominated convergence is an easy to check criterion of uniform integrability.

Scheffé lemma. If $X_n, X\in\mathrm{L}^1$, $X_n \xrightarrow{\text{a.s.}} X$ then $X_n \xrightarrow{\mathrm{L}^1} X$ iff $\mathbb{E}(|X_n|)\to\mathbb{E}(|X|)$.

Slutsky lemma. If $X_n \xrightarrow{\text{law}} X$ and $Y_n \xrightarrow{\text{law}} Y$ and Y is constant then $(X_n, Y_n) \xrightarrow{\text{law}} (X, Y)$. In particular $X_n Y_n \xrightarrow{\text{law}} XY$, $X_n + Y_n \xrightarrow{\text{law}} X + Y$, $X_n / Y_n \xrightarrow{\text{law}} X/Y$ if $Y \neq 0$.

Fubini–Tonelli theorem. Let $(\Omega_1, \mathscr{A}_1, \mu_1)$ and $(\Omega_2, \mathscr{A}_2, \mu_2)$ two measurable spaces, and let $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be a measurable function. If $f \geq 0$ or if $f \in \mathrm{L}^1(\mu_1 \otimes \mu_2)$ then

$$\int f(x,y)\mathrm{d}(\mu_1 \otimes \mu_2)(x,y) = \int \Big(\int f(x,y)\mathrm{d}\mu_1(x)\Big)\mathrm{d}\mu_2(y).$$

Borel–Cantelli lemma. Let $(A_n)_n$ be events in a probability space $(\Omega, \mathscr{A}, \mathbb{P})$. We define

$$\begin{cases} \underline{\lim}_n A_n & = \cup_n \cap_{m\geq n} A_m = \{\omega \in \Omega : \omega \in A_n \text{ for } n \text{ large enough}\}, \\ \overline{\lim}_n A_n & = \cap_n \cup_{m\geq n} A_m = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many values of } n\}. \end{cases}$$

We have $(\underline{\lim}_n A_n^c)^c = \overline{\lim}_n A_n$, and $\overline{\lim}_n \mathbf{1}_{A_n} = \mathbf{1}_{\overline{\lim}_n A_n}$ and $\underline{\lim}_n \mathbf{1}_{A_n} = \mathbf{1}_{\underline{\lim}_n A_n}$.

1. (Cantelli) if $\sum_n \mathbb{P}(A_n) < \infty$ then $\mathbb{P}(\overline{\lim}_n A_n) = 0$
2. (Borel zero-one law) if $\sum_n \mathbb{P}(A_n) = \infty$ and the $(A_n)_n$ are independent then $\mathbb{P}(\overline{\lim}\, A_n) = 1$.

The Borel–Cantelli lemma is a great provider of almost sure convergence. Note that if X takes its values in $[0, +\infty]$ then $\mathbb{E}(X) < \infty$ implies $\mathbb{P}(X < \infty) = 1$, and this allows to prove the Cantelli part:

$$\sum_n \mathbb{P}(A_n) = \sum_n \mathbb{E}\mathbf{1}_{A_n} = \mathbb{E}\sum_n \mathbf{1}_{A_n} \quad \text{and} \quad \Big\{\sum_n \mathbf{1}_{A_n} = \infty\Big\} = \overline{\lim_n}\, A_n.$$

Strong Law of Large Numbers (SLLN). If $X \in \mathrm{L}^1$ and $X_1, X_2, \ldots$ are i.i.d.[2] copies of X then, with $m = \mathbb{E}(X)$,

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\text{a.s.}} m \quad \text{and} \quad \frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\mathrm{L}^1} m.$$

Central limit theorem (CLT). If moreover $X \in \mathrm{L}^2$, then with $\sigma^2 = \mathrm{Var}(X) = \mathbb{E}((X-m)^2) = \mathbb{E}(X^2) - m^2$,

$$\frac{\sqrt{n}}{\sigma}\Big(\frac{X_1 + \cdots + X_n}{n} - m\Big) = \frac{X_1 - m + \cdots + X_n - m}{\sqrt{n}\sigma} \xrightarrow[n\to\infty]{\text{law}} \mathscr{N}(0,1).$$

Law of iterated logarithm (LIL). Under the assumptions and with the notation of the CLT, almost surely

$$\overline{\lim_{n\to\infty}}\Big(\frac{\sqrt{n}}{\sigma\sqrt{2\log\log(n)}}\Big(\frac{X_1 + \cdots + X_n}{n} - m\Big)\Big) = \overline{\lim_{n\to\infty}}\Big(\frac{X_1 - m + \cdots + X_n - m}{\sqrt{2n\log\log(n)}\sigma}\Big) = 1$$

and

$$\underline{\lim_{n\to\infty}}\Big(\frac{\sqrt{n}}{\sigma\sqrt{2\log\log(n)}}\Big(\frac{X_1 + \cdots + X_n}{n} - m\Big)\Big) = \underline{\lim_{n\to\infty}}\Big(\frac{X_1 - m + \cdots + X_n - m}{\sqrt{2n\log\log(n)}\sigma}\Big) = -1$$

Note that the CLT gives $\frac{X_1+\cdots+X_n}{n} \xrightarrow[n\to\infty]{\mathbb{P}} m$, which is a weak form of LLN.

1.4 Uniform integrability

For any family $(X_i)_{i\in I} \subset \mathrm{L}^1$, the following three properties are equivalent[3]. When one (and thus all) of these properties holds true, we say that the family $(X_i)_{i\in I}$ is uniformly integrable (u.i.) or equi-integrable[4]. The first property can be seen as a natural definition of uniform integrability.

1. *(definition of uniform integrability)* $\lim_{r\to+\infty} \sup_{i\in I} \mathbb{E}(|X_i|\mathbf{1}_{|X_i|\geq r}) = 0$

[2]Independent and identically distributed, in French "indépendantes et identiquement distribuées".
[3]Further reading: `https://djalil.chafai.net/blog/2014/03/09/de-la-vallee-poussin-on-uniform-integrability/`
[4]The terminology comes from the fact that by dominated convergence, we have $X \in \mathrm{L}^1$ if and only if $\lim_{r\to\infty} \mathbb{E}(|X|\mathbf{1}_{|X|\geq r}) = 0$.

2. *(epsilon-delta criterion)* the family is bounded in L^1 in the sense that

$$\sup_{i\in I}\mathbb{E}(|X_i|)<\infty$$

and moreover $\forall\varepsilon>0$, $\exists\delta>0$, $\forall A\in\mathcal{F}$, $\mathbb{P}(A)\le\delta\Rightarrow\sup_{i\in I}\mathbb{E}(|X_i|\mathbf{1}_A)\le\varepsilon$

3. *(de la Vallée Poussin[5] boundedness in* L^U *criterion)* there exists a non-decreasing convex $U:\mathbb{R}_+\to\mathbb{R}_+$ such that $\lim_{x\to+\infty}U(x)/x=+\infty$ and such that $(U(|X_i|))_{i\in I}$ is bounded in L^1, namely

$$\sup_{i\in I}\mathbb{E}(U(|X_i|))<\infty.$$

Note that this implies boundedness in L^1, and is implied by boundedness in L^p with $p>1$.

Here are examples for uniformly integrable families:

- every finite subset of L^1 is uniformly integrable. In particular if $X\in\mathrm{L}^1$ then there exists a non-decreasing convex and super-linear U such that $U(|X|)\in\mathrm{L}^1$, but beware that this U depends on X.
- if $(X_i)_{i\in I}$ is bounded in L^p with $p>1$ then it is u.i.
- if $\sup_{i\in I}|X_i|\in\mathrm{L}^1$ (domination: $|X_i|\le X\in\mathrm{L}^1$ for all $i\in I$) then $(X_i)_{i\in I}$ is u.i.
- if $\mathcal{T}\in\{\mathbb{N},\mathbb{R}_+\}$ and $X_t\xrightarrow[t\to\infty]{\mathrm{L}^1}X\in\mathrm{L}^1$ then $(X_t)_{t\in\mathcal{T}}$, $(X_t)_{t\in\mathcal{T}}\cup\{X\}$, and $(X_t-X)_{t\in\mathcal{T}}$ are u.i.
- if $X\in\mathrm{L}^1$ and $X_i=\mathbb{E}(X\mid\mathcal{F}_i)$ for all $i\in I$ for σ-algebras $(\mathcal{F}_i)_{i\in I}$ then $(X_i)_{i\in I}$ is uniformly integrable.

The notion of uniform integrability leads to a stronger version of the dominated convergence theorem: for any $p\ge1$, and for any random variables X and $(X_t)_{t\in\mathcal{T}}$, $\mathcal{T}\in\{\mathbb{N},\mathbb{R}_+\}$, we have

$$X_t,X\in\mathrm{L}^p\text{ and }X_t\xrightarrow[t\to\infty]{\mathrm{L}^p}X\quad\text{if and only if}\quad(|X_t|^p)_{t\in\mathcal{T}}\text{ is u.i. and }X_t\xrightarrow[t\to\infty]{\mathbb{P}}X$$

In particular the convergence in probability together with u.i. implies $X\in\mathrm{L}^1$, which is remarkable!
The dominated convergence theorem corresponds to the special case $\sup_{t\in\mathcal{T}}|X_t|\in\mathrm{L}^1$.

1.5 Conditioning

1. Orthogonal projection in a Hilbert space. Let H be a Hilbert space and $F\subset H$ be a closed sub-space. For all $x\in H$ there exists a unique $y\in F$, called the orthogonal projection of x on F, which satisfies one (and thus all) the following equivalent properties:
   - (orthogonality) for all $z\in F$, $x-y\perp z$ namely $\langle x,z\rangle=\langle y,z\rangle$
   - (variational: least squares) for all $z\in F$, $\|x-y\|\le\|x-z\|$ namely $\|x-y\|=\min_{z\in F}\|x-z\|$.

2. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space and $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{A}$. Let us consider the Hilbert space $H=\mathrm{L}^2(\Omega,\mathcal{A},\mathbb{P})$. The set $F=\mathrm{L}^2(\Omega,\mathcal{F},\mathbb{P})$ is a closed sub-space of H. If $X\in H$, it is natural to consider the best (least squares) approximation of X by an element of F, denoted Y. The random variable Y is the orthogonal projection of X on F, characterized by the following:

$$Y\in\mathrm{L}^2(\Omega,\mathcal{F},\mathbb{P})\text{ and, for all }Z\in\mathrm{L}^2(\Omega,\mathcal{F},\mathbb{P}),\ \mathbb{E}(|X-Y|^2)\le\mathbb{E}(|X-Z|^2).$$

   Using the relation to scalar product, the second property is equivalent to

   - for all $Z\in\mathrm{L}^2(\Omega,\mathcal{F},\mathbb{P})$, $\mathbb{E}(XZ)=\mathbb{E}(YZ)$, or even for all $B\in\mathcal{B}$, $\mathbb{E}(X\mathbf{1}_B)=\mathbb{E}(Y\mathbf{1}_B)$.

   We denote $Y=\mathbb{E}(X\mid\mathcal{F})$ and we call it the *conditional expectation of* Y *given* $\mathcal{F}$. It is the best approximation in L^2 (in a sense least squares) of X by an $\mathcal{F}$-measurable square integrable random variable.

[5]After Charles-Jean Étienne Gustave Nicolas de la Vallée Poussin (1866 – 1962), Belgian mathematician.

3. If now $X \in \mathrm{L}^1(\Omega, \mathscr{A}, \mathbb{P})$, we define by extension $Y = \mathbb{E}(X \mid \mathscr{F})$, a real random variable characterized by
   (a) $Y \in \mathrm{L}^1(\Omega, \mathscr{F}, \mathbb{P})$
   (b) for all Z bounded and $\mathscr{F}$ measurable, $\mathbb{E}(XZ) = \mathbb{E}(YZ)$, or for all $B \in \mathscr{F}$, $\mathbb{E}(X\mathbf{1}_B) = \mathbb{E}(Y\mathbf{1}_B)$.

   Proof. Let μ be the bounded measure on $(\Omega, \mathscr{F})$ defined by $\mu(B) = \mathbb{E}(X\mathbf{1}_B)$, $B \in \mathscr{F}$. Set $\nu = \mathbb{P}_{\mathscr{F}}$. For all $B \in \mathscr{F}$, if $\nu(B) = 0$ then $\mu(B) = 0$. From the Radon–Nikodym theorem, there exists a unique $Y \in \mathrm{L}^1(\Omega, \mathscr{F}, \nu)$ such that $\int_B Y d\nu = \mu(B)$, for all $B \in \mathscr{F}$ in other words $\mathbb{E}(Y\mathbf{1}_B) = \mathbb{E}(X\mathbf{1}_B)$, for all $B \in \mathscr{F}$. ■

The expectation and the variance of square integrable random variables have a variational interpretation. Namely if $X \in \mathrm{L}^2$ then $\mathrm{var}(X)$ is the square distance in L^2 of X to the sub-space of constants r.v. namely

$$\mathrm{var}(X) = \inf_{c \in \mathbb{R}} \mathbb{E}((X-c)^2) = \inf_{c \in \mathbb{R}} (\mathbb{E}(X^2) - 2c\mathbb{E}(X) + c^2).$$

This infinimum is a minimum, achieved for $c = \mathbb{E}(X)$, which is therefore the orthogonal projection of X in L^2 on the sub-space of constant random variables, and

$$\mathrm{var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}(X)) + (\mathbb{E}(X)^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

which follows in fact from the Pythagoras theorem in L^2. More generally we have

$$\begin{aligned}
\mathrm{var}(X) &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \\
&= \mathbb{E}(X^2) - \mathbb{E}((\mathbb{E}(X \mid \mathscr{F}))^2) + \mathbb{E}((\mathbb{E}(X \mid \mathscr{F}))^2) - (\mathbb{E}(X))^2 \\
&= \mathbb{E}(\mathrm{var}(X \mid \mathscr{F})) + \mathrm{var}(\mathbb{E}(X \mid \mathscr{F}))
\end{aligned}$$

where $\mathrm{var}(X \mid \mathscr{F}) = \mathbb{E}(X^2 \mid \mathscr{F}) - (\mathbb{E}(X \mid \mathscr{F}))^2$. Note that by definition of $\mathbb{E}(X \mid \mathscr{F})$,

$$\begin{aligned}
\inf_{Y:\sigma(Y) \subset \mathscr{F}} \mathbb{E}((X-Y)^2) &= \mathbb{E}((X - \mathbb{E}(X \mid \mathscr{F}))^2) \\
&= \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}(X \mid \mathscr{F})) + \mathbb{E}((\mathbb{E}(X \mid \mathscr{F}))^2) \\
&= \mathbb{E}(X^2) - \mathbb{E}((\mathbb{E}(X \mid \mathscr{F}))^2) \\
&= \mathbb{E}(\mathrm{var}(X \mid \mathscr{F})).
\end{aligned}$$

Note that $\mathbb{E} = \mathbb{E}(\cdot \mid \mathscr{T})$ where $\mathscr{T} = \{\varnothing, \Omega\}$. The conditional expectation generalizes the expectation and has all the properties of an expectation, and more. Namely, for all sub-σ-algebra $\mathscr{F}$ of $\mathscr{A}$:

- **Linearity.** for all $\alpha, \beta \in \mathbb{R}$ and $X, Y \in \mathrm{L}^1$, $\mathbb{E}(\alpha X + \beta Y \mid \mathscr{F}) = \alpha\mathbb{E}(X \mid \mathscr{F}) + \beta\mathbb{E}(Y \mid \mathscr{F})$
- **Independence.** If X is independent of $\mathscr{F}$ (always the case when X is constant) then $\mathbb{E}(X \mid \mathscr{F}) = \mathbb{E}(X)$
- **Factorization.** If X is $\mathscr{F}$-measurable, $Y \in \mathrm{L}^1$, $XY \in \mathrm{L}^1$, then $\mathbb{E}(XY \mid \mathscr{F}) = X\mathbb{E}(Y \mid \mathscr{F})$, in particular we recover the "projection property" $\mathbb{E}(X \mid \mathscr{F}) = X$ if $X \in \mathrm{L}^1(\Omega, \mathscr{F}, \mathbb{P})$ which is the case when X is constant
- **Composed "projections" or "tower property".** For all sub-σ-algebras $\mathscr{F}, \mathscr{G}$ with $\mathscr{G} \subset \mathscr{F}$ and all $X \in \mathrm{L}^1$,

  $$\mathbb{E}(\mathbb{E}(X \mid \mathscr{F}) \mid \mathscr{G}) = \mathbb{E}(\mathbb{E}(X \mid \mathscr{G}) \mid \mathscr{F}) = \mathbb{E}(X \mid \mathscr{G}),$$

  and in particular for all $X \in \mathrm{L}^1$, $\mathbb{E}(\mathbb{E}(X \mid \mathscr{F})) = \mathbb{E}(X)$, and if X is constant then $\mathbb{E}(X \mid \mathscr{F}) = X$.
- **Normalization.** $\mathbb{E}(\mathbf{1}_\Omega \mid \mathscr{F}) = \mathbf{1}_\Omega$ (follows from some of the properties above)
- **Positivity or monotonicity.** For all $X, Y \in \mathrm{L}^1$, if $X \le Y$ then $\mathbb{E}(X \mid \mathscr{F}) \le \mathbb{E}(Y \mid \mathscr{F})$, or equivalently for all $X \in \mathrm{L}^1$ if $X \ge 0$ then $\mathbb{E}(X \mid \mathscr{F}) \ge 0$. In particular for all $X \in \mathrm{L}^1$,

  $$|\mathbb{E}(X \mid \mathscr{F})| \le \mathbb{E}(|X| \mid \mathscr{F})$$

- **Convexity.** Jensen inequality: for all non-negative convex $U : \mathbb{R}^d \to \mathbb{R}$ and all $X \in \mathrm{L}^1$,

$$U(\mathbb{E}(X \mid \mathscr{F})) \le \mathbb{E}(U(X) \mid \mathscr{F}).$$

  In particular, for all $p \in [1,\infty)$, $|\mathbb{E}(X \mid \mathscr{F})|^p \le \mathbb{E}(|X|^p \mid \mathscr{F})$. Moreover for all $X \in \mathrm{L}^p$ and $Y \in \mathrm{L}^q$ with $1 \le p, q < \infty$, $1/p + 1/q = 1$ ($q = p/(p-1)$), we have the Hölder inequality

$$|\mathbb{E}(XY \mid \mathscr{F})| \le (\mathbb{E}(|X|^p \mid \mathscr{F})^{1/p} \mathbb{E}(|Y|^q \mid \mathscr{F})^{1/q}.$$

  The Cauchy–Schwarz inequality corresponds to the special case $p = q = 1/2$

- **Monotone convergence.** If $X_n \ge 0$, $X_n \nearrow X$, $X \in \mathrm{L}^1$, then $\mathbb{E}(X_n \mid \mathscr{F}) \nearrow \mathbb{E}(X \mid \mathscr{F})$. This allows to define $\mathbb{E}(X \mid \mathscr{F})$ for all non-negative random variable X taking values in $[0, +\infty]$.

Theorem 1.5.1. Transfer or the meaning of being measurable.

If $T : \Omega \to (F, \mathscr{F})$ are $Y : \Omega \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ and random variables then Y is $\sigma(T)$ measurable if and only if there exists $g : (F, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ measurable such that $Y = g \circ T$.

Proof. If $Y = \mathbf{1}_A$ for $A \in \sigma(T)$, then $A = T^{-1}(B)$ for some $B \in \mathscr{F}$, and therefore $Y = \mathbf{1}_B \circ T$. If $Y = \sum_{i \in I} a_i \mathbf{1}_{A_i}$ with I finite and $A_i = T^{-1}(B_i)$, $B_i \in \mathscr{F}$, then $Y = (\sum_{i \in I} a_i \mathbf{1}_{B_i}) \circ T$. The property is thus satisfied when Y is a step function. Now, if Y is non-negative and $\sigma(T)$ measurable, then there exists a sequence $(Y_n)_n$ of step functions, $\sigma(T)$ measurable, such that $Y_n \nearrow Y$, and $Y_n = g_n \circ T$. By setting $g = \overline{\lim}\, g_n$, we get $Y = g \circ T$. Finally, if Y is just $\sigma(T)$ measurable, then it suffices to write $Y = Y_+ - Y_-$. ■

Let $X \in \mathrm{L}^1(\Omega, \mathscr{A}, \mathbb{P})$ and let $T : (\Omega, \mathscr{A}) \to (F, \mathscr{F})$ be a random variable. The conditional expectation of X given T, denoted $\mathbb{E}(X \mid T)$, is defined by $\mathbb{E}(X \mid T) = \mathbb{E}(X \mid \sigma(T))$. It is characterized by the following properties:

1. There exists $g : (F, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ with $\mathbb{E}(X \mid T) = g(T)$ and $g(T) \in \mathrm{L}^1$
2. For all $h : (F, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ measurable and bounded,

$$\mathbb{E}(Xh(T)) = \mathbb{E}(g(T)h(T)).$$

If $X \in \mathrm{L}^2$ then, thanks to the transfer theorem (Theorem 1.5.1), the conditional expectation $\mathbb{E}(X \mid T)$ is the best approximation in L^2 (least squares!) of X by a measurable function of T.

For a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, an event $A \in \mathscr{F}$, and a sub-σ-algebra $\mathscr{A} \subset \mathscr{F}$, the quantity $\mathbb{P}(A \mid \mathscr{A}) = \mathbb{E}(\mathbf{1}_A \mid \mathscr{A})$ is a random variable taking its values in $[0,1]$. Similarly, conditioning with respect to an event makes sense in the sense that $\mathbb{E}(X \mid A) = \mathbb{E}(X \mid \mathbf{1}_A = 1)$, and

$$\begin{aligned}\mathbb{E}(X \mid \mathbf{1}_A) &= \frac{\mathbb{E}(X\mathbf{1}_A)}{\mathbb{P}(A)}\mathbf{1}_A + \frac{\mathbb{E}(X\mathbf{1}_{A^c})}{\mathbb{P}(A^c)}\mathbf{1}_{A^c}\\ &= \mathbb{E}(X \mid \mathbf{1}_A = 1)\mathbf{1}_A + \mathbb{E}(X \mid \mathbf{1}_A = 0)\mathbf{1}_{A^c}.\end{aligned}$$

Finally, when X and Y take their values in an at most countable set then

$$\mathbb{E}(X \mid Y) = F(Y) \quad \text{where} \quad F(y) = \mathbb{E}(X \mid Y = y) = \sum_x x\mathbb{P}(X = x \mid Y = y).$$

Remark 1.5.2. Conditional expectation as averaging of residual randomness.

Let X and Y be random variables defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and let $\mathscr{A}$ be a sub-σ-algebra of $\mathscr{F}$. If X is independent of $\mathscr{A}$ and if Y is $\mathscr{A}$-measurable, then, using the monotone class theorem, for all $\mathscr{F}$-measurable and bounded or positive $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, we get

$$\mathbb{E}(f(X,Y) \mid \mathscr{A}) = \mathbb{E}(f(X,Y) \mid Y) = g(Y) \quad \text{where} \quad g(y) = \mathbb{E}(f(X,y)).$$

This suggests to interpret intuitively the conditional expectation as an averaging of residual randomness, and not only as the best approximation in the sense of least squares.

Let X and Y be two random variables taking values in the measurable spaces $(E,\mathscr{E})$ and $(F,\mathscr{F})$ respectively. The conditional law of X given Y is a family $(N(y,\cdot))_{y\in F}$ of probability measures on $(E,\mathscr{E})$, in other words a transition kernel, such that for all $A\in\mathscr{E}$, the map $y\in F\mapsto N(y,A)\in[0,1]$ is measurable, and for all bounded (or positive) measurable test function $h:E\to\mathbb{R}$,

$$\mathbb{E}(h(X)\mid Y)=\int_E h(x)N(Y,\cdot).$$

For all $y\in F$, we also say that $N(y,\cdot)$ is the conditional law of X given $Y=y$, in other words

$$\mathbb{E}(h(X)\mid Y=y)=\int_E h(x)N(y,\cdot).$$

In particular $\mathbb{P}(X\in A\mid Y)=N(Y,A)$ for all $A\in\mathscr{E}$. We sometimes speak about disintegration of measure.

The random variables X and Y are independent if and only if $N(y,\cdot)$ does not depend on y in the sense that for almost all $y\in F$, $N(y,\cdot)=\mathbb{P}_X$ where $\mathbb{P}_X$ is the law of X.

If (X,Y) has Lebesgue density $f_{X,Y}$ then X and Y have densities $f_X=\int f(\cdot,y)\mathrm{d}y$ and $f_Y=\int f(x,\cdot)\mathrm{d}x$ and the conditional law $\mathrm{Law}(X\mid Y=y)$ has density $f_{X\mid Y=y}=f_{X,Y}(x,y)/f_Y(y)$, in such a way that

$$f_{X,Y}(x,y)=f_{X\mid Y=y}(x)f_Y(y)=f_X(x)f_{Y\mid X=x}(y).$$

1.6 Gaussian random vectors

A random vector $X=(X_1,\dots,X_n)$ of $\mathbb{R}^n$ is a Gaussian random vector when every linear combination of its components is Gaussian, namely for all $\alpha_1,\dots,\alpha_n\in\mathbb{R}$ the real random variable $\alpha_1X_1+\cdots+\alpha_nX_n$ is Gaussian.

Let X be a random vector with mean vector and the covariance matrix

$$m=\mathbb{E}(X)=(\mathbb{E}(X_1),\dots,\mathbb{E}(X_n))\quad\text{and}\quad\Sigma=\left(\mathbb{E}((X_j-m_j)(X_k-m_k))\right)_{1\le j,k\le n}$$

Then X is Gaussian iff its characteristic function is given for all $t\in\mathbb{R}^n$ by

$$\varphi_X(t)=\mathbb{E}\left(\mathrm{e}^{\mathrm{i}tX}\right)=\mathrm{e}^{\mathrm{i}tm-\frac{1}{2}\langle\Sigma t,t\rangle}.$$

We denote this law $\mathcal{N}(m,\Sigma)$. Beware that when $n=1$, we denote $\Sigma=\sigma^2$.

We say that $\mathcal{N}(0,I_d)$ is the standard Gaussian.

The law $\mathcal{N}(m,\Sigma)$ has a density iff Σ is invertible, given by

$$x\in\mathbb{R}^n\mapsto\frac{\exp\left(-\frac{1}{2}\langle\Sigma^{-1}(x-m),x-m\rangle\right)}{\sqrt{(2\pi)^n\det(\Sigma)}},$$

otherwise $\mathcal{N}(m,\Sigma)$ is supported by a strict sub-vector space of $\mathbb{R}^n$.

If $(X_1,\dots,X_n)$ is a Gaussian random vector, then $X_1,\dots,X_n$ are independent iif Σ is diagonal.

If $Z\sim\mathcal{N}(0,I_n)$ and $m\in\mathbb{R}^d$ and $A\in\mathcal{M}_{d,n}(\mathbb{R})$ then $AZ\sim\mathcal{N}(m,AA^\top)$ is a Gaussian random vector of $\mathbb{R}^d$.

Coding in action 1.6.1. Simulation.

Write a Python[a] or Julia[b] program for the simulation of a sample of $\mathcal{N}(m,\Sigma)$ knowing m and Σ. What is the best way to reduce to the one-dim. case? What is the best way to find A such that $AA^\top=\Sigma$?

[a] `https://en.wikipedia.org/wiki/Python_(programming_language)`
[b] `https://en.wikipedia.org/wiki/Julia_(programming_language)`

1.7 Bounded variation and Lebesgue – Stieltjes integral

Definition 1.7.1. p-variation of a function on a finite interval.

Let $[a,b]\subset\mathbb{R}$ be a finite interval. For all $p\geq 1$, the p-variation of a function $f:[a,b]\to\mathbb{R}$ is defined by

$$\|f\|_{p-\mathrm{var}}=\Big(\sup_{t_k}\sum_k|f(t_{k+1})-f(t_k)|^p\Big)^{1/p}\in[0,+\infty]$$

where the supremum runs over all finite partitions or sub-divisions of the interval I namely the finite sequences $(t_k)_{0\leq k\leq n}$ in $[a,b]$ such that $n\geq 0$ and $a=t_0<\cdots<t_{n+1}=b$.

- $\|f\|_{1\text{-var}}$ is called sometimes the total variation of f
- if $f:[a,b]\to\mathbb{R}$ has finite 1-variation, we say that f has finite variation or is of bounded variation
- if $f:[a,b]\to\mathbb{R}$ is of bounded variation then f is bounded (the boundedness of $[a,b]$ plays a role here).
- if $f:[a,b]\to\mathbb{R}$ if of bounded variation and is differentiable with integrable derivative then

$$\|f\|_{1\text{-var}}=\int_a^b|f'(t)|\mathrm{d}t.$$

- if f is continuously differentiable then f has bounded variation and the latter holds true.

Theorem 1.7.2. Representation of bounded variation functions on a finite interval.

Let $[a,b]\subset\mathbb{R}$ be a finite interval. For all $f:[a,b]\mapsto\mathbb{R}$, the following properties are equivalent:

1. f is of bounded variation
2. f is the difference of two positive increasing functions $[a,b]\to\mathbb{R}$.

Such a decomposition is not unique in general.

Proof. $1\Rightarrow 2$. Let f be a function of bounded variation on $[a,b]$. For all $t\in[a,b]$, let

$$F(t)=\sup_\delta\sum_{k=0}^{n-1}|f(t_{k+1})-f(t_k)|$$

where the supremum runs over the set of partitions or sub-divisions $\delta: a=t_0<\cdots<t_n=t$ of $[a,t]$, $n=n_\delta\geq 1$. Now F is increasing (and bounded) by definition. It suffices now to show that $G=F-f$ is increasing. We observe that for all $t_1<t_2$ in $[a,b]$, we have $F(t_1)+f(t_2)-f(t_1)\leq F(t_1)+|f(t_2)-f(t_1)|\leq F(t_2)$, and thus

$$G(t_2)-G(t_1)=F(t_2)-f(t_2)-F(t_1)+f(t_1)\geq 0.$$

$2\Rightarrow 1$. If f and g have bounded variation on $[a,b]$, then it is also the case for $f-g$. On the other hand, if f is monotonic on I then it is of bounded variation since for all sub-division $a=t_0<\cdots<t_n=b$, $n\geq 1$,

$$\sum_{k=0}^{n-1}|f(t_{k+1})-f(t_k)|=|f(b)-f(a)|.$$

■

The notion of bounded variation is used for the Lebesgue – Stieltjes integral in stochastic calculus.

Theorem 1.7.3. Lebesgue – Stieltjes integral of continuous finite variation integrators.

Let $[a,b]\subset\mathbb{R}$ be a finite interval. Let $f:[a,b]\to\mathbb{R}$ be right continuous and of bounded variation.

Then there exists a unique finite signed Borel measure μ_f on $([a,b],\mathscr{B}_{[a,b]})$ such that

$$\mu_f(\{a\})=0, \quad \text{and for all } t\in[a,b], \quad \mu_f((a,t])=f(t)-f(a).$$

It is customary to denote $\mathrm{d}\mu_f=\mathrm{d}f$, and for all measurable $g:[a,b]\to\mathbb{R}$, positive or in $\mathrm{L}^1(|\mu_f|)$,

$$\int_a^b g(t)\mathrm{d}f(t)=\int g\mathrm{d}\mu_f.$$

Moreover, for all bounded and continuous $g:[a,b]\to\mathbb{R}$, and for all sequence $(\delta_n)_{n\geq1}$ of partitions or sub-divisions of $[a,b]$, $\delta_n: a=t_0^{(n)}<\cdots<t_{m_n}^{(n)}=b$, $m_n\geq1$, with $\lim_{n\to\infty}\max_k(t_{k+1}^{(n)}-t_k^{(n)})=0$, we have

$$\int_a^b g(t)\mathrm{d}f(t)=\lim_{n\to\infty}\sum_k g(t_k^{(n)})(f(t_{k+1}^{(n)})-f(t_k^{(n)})).$$

Furthermore, $h:t\in[a,b]\mapsto h(t)=\int_a^t g(s)\mathrm{d}f(s)$ is continuous and of bounded variation, and $\mu_h=g\mu_f$ in other words $\mathrm{d}h(t)=g(t)\mathrm{d}f(t)$, in the sense that for all bounded and measurable $k:[a,b]\to\mathbb{R}$,

$$\int_a^b k(t)\mathrm{d}h(t)=\int_a^b k(t)\mathrm{d}\int_a^t g(s)\mathrm{d}f(s)=\int_a^b k(t)g(t)\mathrm{d}f(t).$$

In particular when $f(t)=t$ for all $t\geq0$ then on all $[a,b]\subset[0,\infty)$, the measure μ_f is the Lebesgue measure and for all measurable $g:\mathbb{R}_+\mapsto\mathbb{R}$ which is locally bounded or positive, we have, for all $t\geq0$,

$$\int_0^t g(s)\mathrm{d}f(s)=\int g\mathrm{d}\mu_f=\int_0^t g(s)\mathrm{d}s.$$

Theorem 1.7.3 is used in stochastic calculus with $f(t)=V_t(\omega)$, $t\geq0$, and for almost all fixed $\omega\in\Omega$ where $V=(V_t)_{t\geq0}$ is a finite variation process, for instance $V=\langle M\rangle$ where M is a continuous local martingale. In particular when $M=B$ is Brownian motion then $V_t=t$ is deterministic and we recover the example above.

Theorem 3.2.1 says that Brownian motion has a.s. sample paths of infinite variation on any interval. In particular the assumptions of Theorem 1.7.3 are not satisfied when $f(t)=B_t(\omega)$, $t\in[a,b]\subset[0,+\infty)$.

Proof. First part. Theorem 1.7.2 gives $f=f_+-f_-$ where $f_\pm\geq0$ are bounded and increasing. This reduces the problem to the case where f is increasing and μ_f is a positive Borel measure. In this case, the result follows from the Carathéodory extension theorem (Theorem 1.8.5). Note: μ_f is unique even if $f_\pm$ are not.

Second part. For all $n\geq1$, set $g^{(n)}(a)=g(a)$, and for all $t\in(a,b]$, $g^{(n)}(t)=g(t_k^{(n)})$ if $t\in(t_k^{(n)},t_{k+1}^{(n)}]$ for some $k\in\{0,\ldots,m_n-1\}$. Then $g^{(n)}$ is measurable, we have $\lim_{n\to\infty}g^{(n)}(t)=g(t)$ for all $t\in[a,b]$, and moreover $\sup_n\sup_{t\in[a,b]}|g^{(n)}(t)|\leq\sup_{t\in[a,b]}|g(t)|<\infty$. By dominated convergence in $\mathrm{L}^1(|\mu|)$, we obtain

$$\sum_k g(t_k^{(n)})(f(t_{k+1}^{(n)})-f(t_k^{(n)}))=\int g^{(n)}\mathrm{d}\mu_f\underset{n\to\infty}{\longrightarrow}\int g\mathrm{d}\mu_f=\int_a^b g(t)\mathrm{d}f(t).$$

Note that if g is measurable and not continuous, then $g^{(n)}\to g$ as $n\to\infty$, almost everywhere on $[a,b]$, which is suitable for the Lebesgue measure but not necessarily for the measure $|\mu|$ which is of interest here.

Third part. First of all, for all $s\in[a,b]$, we have $\mu_{f|_{[a,s]}}=\mu_f|_{[a,s]}$.

$$\int_a^s g(t)\mathrm{d}f(t)=\int g\mathrm{d}\mu_{f\mathbf{1}_{[a,s]}}=\int g\mathbf{1}_{[a,s]}\mathrm{d}\mu_f.$$

The continuity of h follows now by dominated convergence. For the 1-variation, we write

$$\sum_k|h(t_{k+1})-h(t_k)|\leq\sum_k\int|g|\mathbf{1}_{(t_{k+1},t_k]}\mathrm{d}|\mu_f|=\int|g|\mathrm{d}|\mu_f|<\infty.$$

Finally, to prove the formula, it suffices to check it for $k=\mathbf{1}_{[a,c]}$ for $c\in[a,b]$. This writes $\mu_h(c)-\mu_h(a)=\int_a^c g(t)\mathrm{d}\mu_f(t)=h(c)-h(a)$, which is the definition of μ_h. Note that by construction we have $h(a)=0$. ■

Remark 1.7.4. Riemann – Stieltjes – Young integral.

Following L.C. Young, it can be shown that if $f, g : [a,b] \to \mathbb{R}$ are continuous with f of finite p-var. and g of finitie q-var. with $1/p + 1/q > 1$, then the Riemann – Stieltjes integral is well defined:

$$\int_a^b f(t)\mathrm{d}g(t) = \lim_{n\to\infty} \sum_{k=0}^{m_n} f(t_k^{(n)})(g(t_{k+1}^{(n)}) - g(t_k^{(n)})),$$

where $(\delta_n)_{n\geq 1}$ is an arbitrary sequence of partitions of $[a,b]$, $\delta_n : a = t_0 < \cdots < t_{m_n} = b$, $m_n \geq 1$.

1.8 Monotone class theorem and Carathéodory extension theorem

Definition 1.8.1. π-systems and λ-systems.

- We say that $\mathscr{C} \subset \mathscr{P}(\Omega)$ is a π-system when $A \cap B \in \mathscr{C}$ for all $A, B \in \mathscr{C}$
- We say that $\mathscr{S} \subset \mathscr{P}(\Omega)$ is a λ-system (or monotone class or Dynkin[6] system) when
  - $\cup_n A_n \in \mathscr{S}$ for all $(A_n)_n$ such that $A_n \subset A_{n+1}$ and $A_n \in \mathscr{S}$ for all n
  - $A \setminus B \in \mathscr{S}$ for all $A, B \in \mathscr{S}$ such that $B \subset A$.

[6] Named after Eugene Dynkin (1924 – 2014), Soviet and American mathematician.

Basic examples of π-systems are given by the class of singletons $\{\{x\} : x \in \mathbb{R}\} \cup \{\varnothing\}$, the class of product subsets $\{A \times B : A, B \in \mathscr{P}(\Omega)\}$, and the class of intervals $\{(-\infty, x] : x \in \mathbb{R}\}$.

A basic yet important example of λ-system is given by $\{A \in \mathscr{A} : \mathbb{P}(A) = \mathbb{Q}(A)\}$ where $\mathbb{P}$ and $\mathbb{Q}$ are probability measures on $(\Omega, \mathscr{A})$, see Corollary 1.8.4 for an application.

Lemma 1.8.2. σ-algebras.

A λ-system that contains Ω and which is a π-system is a σ-algebra.

Note that conversely, a σ-algebra is always a π-system, but not a λ-system in general, due to the second property of λ-systems which is not necessarily valid for a σ-algebra when $A \neq \Omega$.

Proof. If a λ-system $\mathscr{S} \subset \mathbb{P}(\Omega)$ contains Ω and is a π-system then for all $A, B \in \mathscr{S}$ we have

$$A \cup B = \Omega \setminus ((\Omega \setminus A) \cap (\Omega \setminus B)),$$

which means that $\mathscr{S}$ is table by finite union. This allows to drop the non-decreasing condition in the stability of $\mathscr{S}$ by countable union, which simply means finally that $\mathscr{S}$ is a σ-algebra. ■

Theorem 1.8.3. Dynkin π-λ Theorem.

If $\mathscr{S} \subset \mathscr{P}(\Omega)$ is a λ-system containing Ω and including a π-system $\mathscr{C}$,
then $\mathscr{S}$ contains also the σ-algebra $\sigma(\mathscr{C})$ generated by $\mathscr{C}$.

Proof. The λ-system generated by a subset of $\mathscr{P}(\Omega)$ is by definition the intersection of all λ-systems which include this subset. This intersection is not empty since it contains $\mathscr{P}(\Omega)$, and we can check that it is a λ-system. It is the smallest (for the inclusion) λ-system containing the initial subset of $\mathscr{P}(\Omega)$.

Let $\mathscr{S}'$ be the λ-system generated by $\mathscr{C}$ and Ω. It suffices to show that $\mathscr{S}'$ is a σ-algebra. For that, and thanks to lemma 1.8.2, it suffices to show that $\mathscr{S}'$ is a π-system. To do so, let us define

$$\mathscr{S}_1 = \{A \in \mathscr{S}' : A \cap B \in \mathscr{S}' \text{ for all } B \in \mathscr{C}\},$$

which is a λ-system including Ω and containing $\mathscr{C}$, hence $\mathscr{S}_1 \subset \mathscr{S}'$, and thus $\mathscr{S}_1 = \mathscr{S}'$. Now,

$$\mathscr{S}_2 = \{A \in \mathscr{S}' : A \cap B \in \mathscr{S}' \text{ for all } B \in \mathscr{S}'\}$$

is a λ-system containing Ω and including $\mathscr{S}$ and thus $\mathscr{S}_2 = \mathscr{S}'$, hence $\mathscr{S}'$ is a π-system. ■

Corollary 1.8.4. Sierpiński[a] – Dynkin (functional) monotone class theorem.

[a] Named after Wacław Sierpiński (1882 – 1969), Polish mathematician.

1. For all probability measures $\mathbb{P}$ and $\mathbb{Q}$ on a measurable space $(\Omega, \mathscr{A})$, if $\mathbb{P}(A) = \mathbb{Q}(A)$ for all $A \in \mathscr{C}$ where $\mathscr{C}$ is a π-system such that $\sigma(\mathscr{C}) = \mathscr{A}$, then $\mathbb{P} = \mathbb{Q}$
2. Let H be a vector space of bounded measurable functions $(\Omega, \mathscr{A}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ such that
   (a) H is stable by monotone convergence namely if $(f_n)_n$ is a sequence in H such that $f_n \nearrow f$ pointwise with f bounded then $f \in H$
   (b) H contains constant functions namely $\mathbf{1}_\Omega \in H$, is stable by product namely if $f, g \in H$ then $fg \in H$, and contains all $\mathbf{1}_A$ for all A in a π-system $\mathscr{C}$ on Ω such that $\sigma(\mathscr{C}) = \mathscr{A}$

   then H contains all $\mathscr{A}$-measurable bounded functions $\Omega \to \mathbb{R}$.

Note that H is an algebra in the sense that it is a vector space stable by product.
The second statement can be seen as some sort of Stone – Weierstrass theorem of measure theory.

Proof.

1. Take $\mathscr{S} = \{A \in \mathscr{A} : \mathbb{P}(A) = \mathbb{Q}(A)\}$ and use Theorem 1.8.3.
2. Take $\mathscr{S} = \{A \in \mathscr{A} : \mathbf{1}_A \in H\}$ and use Theorem 1.8.3.

■

Theorem 1.8.5. Carathéodory extension theorem.

Let $\Omega \neq \varnothing$, $\mathscr{A} \subset \mathscr{P}(\Omega)$, and $\mu : \mathscr{A} \mapsto \mathbb{R}_+$. Let $\sigma(A)$ be the σ-algebra generated by $\mathscr{A}$. If

1. $\Omega \in \mathscr{A}$
2. (stability by complement) for all $A \in \mathscr{A}$, we have $A^c = \Omega \setminus A \in \mathscr{A}$
3. (stability by intersection) for all $A, B \in \mathscr{A}$, we have $A \cap B \in \mathscr{A}$
4. μ is σ-additive and σ-finite

then there exists a unique σ-additive measure μ_{ext} on $(\Omega, \sigma(A))$ such that $\mu_{\text{ext}} = \mu$ on $\mathscr{A}$.

Proof. See for instance [1]. The uniqueness can be deduced from Corollary 1.8.4. ■

Chapter 2

Processes, filtrations, stopping times, martingales

A stochastic process or process is a family of random variables $X = (X_t)_{t\geq 0}$, indexed by a parameter $t \in \mathbb{R}_+$ interpreted as a time, defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and taking values in some measurable space $(G, \mathscr{B})$. By default a process takes real values. In general G is a metric space, with distance denoted d, complete, separable, and $\mathscr{B}$ is its Borel σ-algebra.

2.1 Measurability

The natural filtration of a process $(X_t)_{t\geq 0}$ is the increasing family $(\mathscr{F}_t)_{t\geq 0}$ of sub-σ-algebras of $\mathscr{F}$ defined for all $t \geq 0$ by $\mathscr{F}_t = \sigma(X_s : 0 \leq s \leq t)$. More generally, an increasing family $(\mathscr{F}_t)_{t\geq 0}$ of sub-σ-algebras of $\mathscr{F}$ is called a filtration. For a given filtration $(\mathscr{F}_t)_{t\geq 0}$ on $(\Omega, \mathscr{F}, \mathbb{P})$, we say that the process X is...

- real when $G = \mathbb{R}$ in other words X takes real values (this is the default in this course)
- d-dimensional when $G = \mathbb{R}^d$ in other words X takes its values in $\mathbb{R}^d$, $d \geq 1$
- issued from the origin when $X_0 = 0$ (makes sense when G is a vector space)
- adapted when for all $t \geq 0$, X_t is $\mathscr{F}_t$ measurable
- measurable when for all $t \geq 0$, $(s, \omega) \in [0, t] \times \Omega \mapsto X_s(\omega)$ is $\mathscr{B}_{[0,t]} \otimes \mathscr{F}$ measurable
- progressive when for all $t \geq 0$, $(s, \omega) \in [0, t] \times \Omega \mapsto X_s(\omega)$ is $\mathscr{B}_{[0,t]} \otimes \mathscr{F}_t$ measurable
- right-continuous (respectively left-continuous, continuous) when for almost all $\omega \in \Omega$, the sample path $t \in \mathbb{R}_+ \mapsto X_t(\omega) \in G$ is right-continuous (respectively left-continuous, continuous)
- square integrable when for all $t \geq 0$, $\mathbb{E}(X_t^2) < \infty$
- bounded in L^p, $p \geq 1$, when $\sup_{t\geq 0} \mathbb{E}(|X_t|^p) < \infty$
- bounded when there exists a finite $C > 0$ such that almost surely, $\sup_{t\geq 0} |X_t| \leq C$
- locally bounded when for almost all $\omega \in \Omega$ and all $t \geq 0$, $\sup_{s\in[0,t]} |X_s(\omega)| < \infty$
- of finite variation when almost surely $t \mapsto X_t$ is of bounded variation on all finite intervals of $\mathbb{R}_+$, equivalently is the difference of two positive increasing processes, see Theorem 1.7.2.
- Feller continuous when $x \mapsto \mathbb{E}(f(X_t) \mid X_0 = x)$ is continuous for all $t \geq 0$ and bounded measurable f.

Theorem 2.1.1. Progressive σ-field and progressive processes.

1. The family $\mathscr{P}$ of all $A \in \mathscr{F} \otimes \mathscr{B}_{\mathbb{R}_+}$ such that the process $(\omega, t) \mapsto \mathbf{1}_{(\omega,t)\in A}$ is progressive is a σ-field on $\Omega \times \mathbb{R}_+$ called the progressive σ-field. Moreover the following properties hold:
   - For all $A \subset \Omega \times \mathbb{R}_+$, we have $A \in \mathscr{P}$ if and only if for all $t \geq 0$, $A \cap (\Omega \times [0, t]) \in \mathscr{F}_t \otimes \mathscr{B}_{[0,t]}$.
   - A process $X = (X_t)_{t\geq 0}$ is progressive if and only if it is measurable with respect to the progressive σ-algebra $\mathscr{P}$ on $\Omega \times \mathbb{R}_+$ as a random variable $X : (\omega, t) \in \Omega \times \mathbb{R}_+ \mapsto X_t(\omega)$

2. If $X = (X_t)_{t\geq 0}$ is adapted right-continuous or left-continuous defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ and taking its values in a metric space (E, d) equipped with its Borel σ-algebra, then X is progressive. In particular continuous adapted implies progressive.

Proof.

1. Exercise
2. We give the proof in the right-continuous case, the left-continuous case being entirely similar. For all $n \geq 1$, $t > 0$, $s \in [0, t]$, we define the random variable

$$X_s^n = \begin{cases} X_{kt/n}^n & \text{if } s \in [(k-1)t/n, kt/n),\ 1 \leq k \leq n, \\ X_t & \text{if } s = t. \end{cases}$$

Since $(X_t)_{t\geq 0}$ is right-continuous, it follows that $X_s(\omega) = \lim_{n\to\infty} X_s^n(\omega)$ for all $t > 0$ and $s \in [0, t]$ and all $\omega \in \Omega$. On the other hand, for every Borel subset A of E,

$$\{(\omega, s) \in \Omega \times [0, t] : X_s^n(\omega) \in A\} = (\{X_t \in A\} \times \{t\}) \bigcup \Big(\bigcup_{k=1}^{n} (\{X_{kt/n} \in A\} \times [(k-1)t/n, kt/n)) \Big).$$

Since $(X_t)_{t\geq 0}$ is adapted, this set belongs to $\mathscr{F}_t \otimes \mathscr{B}_{[0,t]}$. Therefore, for all $n \geq 1$, the function $(\omega, s) \in \Omega \times [0, t] \mapsto X_s^n(\omega)$ is measurable for $\mathscr{F}_t \otimes \mathscr{B}_{[0,t]}$. Now a pointwise limit of measurable functions is measurable, and therefore the function $(\omega, s) \in \Omega \times [0, t] \mapsto X_s(\omega)$ is also measurable for $\mathscr{F}_t \otimes \mathscr{B}_{[0,t]}$, which means, since $t > 0$ is arbitrary, that $(X_t)_{t\geq 0}$ is progressive.

■

A process $X = (X_t)_{t\geq 0}$ taking its values in $\mathbb{R}^d$ can be seen as a random variable taking its values in the "path space" $\mathscr{P}(\mathbb{R}_+, \mathbb{R}^d)$ of functions from $\mathbb{R}_+$ to $\mathbb{R}^d$. The measurability is for free if we equip $\mathscr{P}(\mathbb{R}_+, \mathbb{R}^d)$ with the σ-algebra $\mathscr{A}_{\mathscr{P}(\mathbb{R}_+,\mathbb{R}^d)}$ generated by the cylindrical events

$$\{f \in \mathscr{P}(\mathbb{R}_+, \mathbb{R}^d) : f(t_1) \in I_1, \ldots, f(t_n) \in I_n\}$$

where $n \geq 1$, $t_1, \ldots, t_n \in \mathbb{R}_+$, and where $I_1, \ldots, I_n$ are products of intervals in $\mathbb{R}^d$ of the form $\prod_{i=1}^d (a_i, b_i]$. Unfortunately $\mathscr{P}(\mathbb{R}_+, \mathbb{R}^d)$ is so big that $\mathscr{A}_{\mathscr{P}(\mathbb{R}_+,\mathbb{R}^d)}$ turns out to be too small, and does not contain for instance events of interest such that $\{f \in \mathscr{P}(\mathbb{R}_+, \mathbb{R}^d) : \sup_{t\in[0,1]} f(t) < 1\}$.

We focus in this course on continuous processes. This suggests to consider $\mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)$ and the σ-algebra $\mathscr{A}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)}$ generated by the cylindrical events $\{f \in \mathscr{C}(\mathbb{R}_+, \mathbb{R}^d) : f(t_1) \in I_1, \ldots, f(t_n) \in I_n\}$ where $n \geq 1$, $t_1, \ldots, t_n \in \mathbb{R}_+$, $n \geq 1$, $I_1, \ldots, I_n$ are products of intervals in $\mathbb{R}^d$ of the form $\prod_{i=1}^d (a_i, b_i]$. We have then the following:

Theorem 2.1.2. What a wonderful world.

On $\mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)$, the following σ-algebras coincide:

- σ-algebra $\mathscr{A}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)}$ generated by the cylindrical events
- Borel σ-algebra $\mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)}$ generated by the open sets of the topology of uniform convergence on compact intervals of $\mathbb{R}_+$.

Proof. Take $d = 1$ for simplicity. It can be shown that $\mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)$ equipped with the distance

$$d(f, g) = \sum_{n=1}^{\infty} 2^{-n} (1 \wedge \max_{t\in[0,n]} |f(t) - g(t)|)$$

is a Polish space in other words a complete and separate metric space, and the associated topology is the one of uniform convergence on compact subsets of $\mathbb{R}_+$. First we have the inclusion $\mathscr{A}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)} \subset \mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R})}$ since the σ-algebra $\mathscr{A}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R})}$ is generated by the cylinders

$$\{f \in \mathscr{C}(\mathbb{R}_+, \mathbb{R}) : f(t_1) < a_1, \ldots, f(t_n) < a_n\}, \quad n \geq 1, t_1, \ldots, t_n \in \mathbb{R}_+, a_1, \ldots, a_n \in \mathbb{R},$$

which are open subsets. Conversely, for all $g \in \mathscr{C}(\mathbb{R}_+,\mathbb{R})$, all $n \geq 1$, and all $r > 0$,

$$\{f \in \mathscr{C}(\mathbb{R}_+,\mathbb{R}) : \max_{t\in[0,n]} |f(t) - g(t)| \leq r\} = \cap_{t\in\mathbb{Q}\cap[0,n]}\{f \in \mathscr{C}(\mathbb{R}_+,\mathbb{R}) : |f(t) - g(t)| \leq r\}$$

belongs to $\mathscr{A}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R})}$, and since these sets generate $\mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)}$, we get $\mathscr{A}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)} = \mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R})}$. ∎

Theorem 2.1.3. Continuous processes as random variables on path space.

Let $X = (X_t)_{t\geq 0}$ be a continuous d-dimensional process defined on $(\Omega,\mathscr{F},\mathbb{P})$. Let $\Omega' \in \mathscr{F}$ such that $\mathbb{P}(\Omega') = 1$ and $\Omega' \subset \{X_\bullet \in \mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)\}$. Then the map $X|_{\Omega'} : \omega \in \Omega' \to X_\bullet(\omega) \in \mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)$ is measurable with respect to the σ-algebras $\mathscr{F}' = \{F \cap \Omega' : A \in \mathscr{A}\}$ and $\mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)}$.

Proof. Let us consider an arbitrary cylindrical event

$$F = \{f \in \mathscr{C}(\mathbb{R}_+,\mathbb{R}^d) : f(t_1) \in I_1,\ldots, f(t_n) \in I_n\},$$

where $n \geq 1$, $t_1,\ldots,t_n \in \mathbb{R}_+$, and $I_1,\ldots,I_n$ are product of intervals as $\prod_{i=1}^d (a_i,b_i]$. Then

$$\Omega' \cap \{X_\bullet \in F\} = \Omega' \cap \{X_{t_1} \in I_1,\ldots, X_{t_n} \in I_n\} \in \mathscr{F}'.$$

Now $\mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)}$ is generated by cylindrical events (Theorem 2.1.2). ∎

Remark 2.1.4. Equality of processes, modification and indistinguishability.

Two processes $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ defined on the same probability space $(\Omega,\mathscr{F},\mathbb{P})$ are indistinguishable when for almost all $\omega \in \Omega$ the sample paths $t \mapsto X_t(\omega)$ and $t \mapsto Y_t(\omega)$ coincide, namely

$$\mathbb{P}(\forall t \geq 0 : X_t = Y_t) = 1.$$

There is a weaker notion in which the almost sure event depends on time, namely we say that Y is a modification of X if for all $t \geq 0$ the event $\Omega_t = \{\omega \in \Omega : X_t(\omega) \neq Y_t(\omega)\}$ is negligible, in other words

$$\forall t \geq 0 : \mathbb{P}(X_t = Y_t) = 1.$$

If X and Y are continuous then the two notions of indistinguishable and modification coincide.

If $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ are two processes taking values in $\mathbb{R}^d$ with same finite dimensional marginal distributions, in the sense that for all $n \geq 1$ and all $t_1,\ldots,t_n \in \mathbb{R}_+$, the random vectors $(X_{t_1},\ldots,X_{t_n})$ and $(Y_{t_1},\ldots,Y_{t_n})$ have same law in $(\mathbb{R}^d)^n$, then X and Y have same law as random variables on the path space $(\mathscr{P}(\mathbb{R}_+,\mathbb{R}),\mathscr{A}_{\mathscr{P}(\mathbb{R}_+,\mathbb{R}^d)})$. The following theorem provides a sort of converse, stated when $d = 1$ for simplicity.

Theorem 2.1.5. Kolmogorov extension theorem.

For all $n \geq 1$ and all $t \in \mathbb{R}^n$ with $0 \leq t_1 \leq \cdots \leq t_n$, let $\mu_{t_1,\ldots,t_n}$ be a probability measure on $\mathbb{R}^n$. Let us assume the following consistency condition:

- for all $n \geq 1$, $t \in \mathbb{R}^n$ with $0 \leq t_1 \leq \cdots \leq t_n$, and all $A_1,\ldots,A_{n-1} \in \mathscr{B}_\mathbb{R}$, we have

$$\mu_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_{n-1} \times \mathbb{R}) = \mu_{t_1,\ldots,t_{n-1}}(A_1 \times \cdots \times A_{n-1}).$$

Then there exists a unique probability measure μ on the path space $(\mathscr{P}(\mathbb{R}_+,\mathbb{R}),\mathscr{A}_{\mathscr{P}(\mathbb{R}_+,\mathbb{R})})$ such that for all $n \geq 1$, all $t \in \mathbb{R}^n$ with $0 \leq t_1 \leq \cdots \leq t_n$, and all $A_1,\ldots,A_n \in \mathscr{B}_\mathbb{R}$, we have

$$\mu(\pi_{t_1} \in A_1,\ldots,\pi_{t_n} \in A_n) = \mu_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_n),$$

where $\pi_t(\omega) = \omega_t$, namely $\pi_t : \omega \in \mathscr{P}(\mathbb{R}_+,\mathbb{R}) \mapsto \omega_t \in \mathbb{R}$ for all $t \geq 0$.

Proof. For a cylindrical event $A_{t_1,\dots,t_n}(B) = \{f \in \mathscr{P}(\mathbb{R}_+,\mathbb{R}) : (f(t_1),\dots,f(t_n)) \in B\}$ where $n \geq 1$, $t \in \mathbb{R}^n$ with $0 \leq t_1 \leq \cdots \leq t_n$, and where $B \in \mathscr{B}_{\mathbb{R}^n}$, we define $\mu(B) = \mu_{t_1,\dots,t_n}(B)$. This makes sense thanks to the consistency condition. Note that we could drop the ordering on the coordinates of t by defining $\mu_{t_1,\dots,t_n} = \mu_{t_{(1)},\dots,t_{(n)}}$ where $t_{(1)} \leq \cdots \leq t_{(n)}$ is the reordering. Moreover $\mu(\mathscr{P}(\mathbb{R}_+,\mathbb{R})) = 1$. Since the set of cylinders satisfies the assumptions of the Carathéodory extension theorem (Theorem 1.8.5), and generates the σ-algebra $\mathscr{A}_{\mathscr{P}(\mathbb{R}_+,\mathbb{R})}$, it remains to show that μ is a σ-finite measure, which is the difficult part of the proof. See instance [1]. ■

2.2 Completeness

Contrary to discrete processes, continuous processes lead naturally to measurability issues.

In a probability space $(\Omega,\mathscr{F},\mathbb{P})$, we say that $A \subset \Omega$ is negligible when there exists $A' \in \mathscr{F}$ with $A \subset A'$ and $\mathbb{P}(A') = 0$. We say that the $(\Omega,\mathscr{F},\mathbb{P})$ is complete when $\mathscr{F}$ contains the negligible subsets of Ω.

A filtration $(\mathscr{F}_t)_{t\geq 0}$ on $(\Omega,\mathscr{F},\mathbb{P})$ is complete when $\mathscr{F}_0$ contains the negligible subsets of $\mathscr{F}$.

Completeness emerges naturally via almost sure events which are complement of negligible subsets.

Theorem 2.2.1. Measurability of running supremum from completeness.

Let $(X_t)_{t\geq 0}$ be a continuous process defined on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ and taking values in a topological space E equipped with its Borel σ-field $\mathscr{E}$. Let $f : E \to \mathbb{R}$ be a measurable function.

- If $(\Omega,\mathscr{F},\mathbb{P})$ is complete then $\sup_{s\in[0,t]} f(X_s)$ is measurable for all $t \geq 0$.
- If X is adapted with respect to a complete filtration $(\mathscr{F}_t)_{t\geq 0}$ then $(\sup_{s\in[0,t]} f(X_s))_{t\geq 0}$ is adapted.

Proof. Let $\Omega' \in \mathscr{F}$ be an almost sure event on which X is continuous. Set $S_t = \sup_{s\in[0,t]} f(X_s)$.

- For all $t \geq 0$ and $A \in \mathscr{E}$, we have
$$\Omega' \cap \{S_t \in A\} = \Omega' \cap \Big\{\sup_{s\in[0,t]\cap\mathbb{Q}} f(X_s) \in A\Big\} \in \mathscr{F},$$
while $(\Omega \setminus \Omega') \cap \{S_t \in A\} \subset \Omega \setminus \Omega'$ is negligible and thus belongs to $\mathscr{F}$ by completeness of $(\Omega,\mathscr{F},\mathbb{P})$.
- Same argument as before with $\mathscr{F}_t$ instead of $\mathscr{F}$.

■

The notion of completeness is relative to the probability measure $\mathbb{P}$. There is also a notion of universal completeness, see [5], that do not depend on the probability measure, but we do not use it in these notes.

2.3 Stopping times

Definition 2.3.1. Stopping time.

A map $T : \Omega \to [0,+\infty]$ is a stopping time or optional time for a filtration $(\mathscr{F}_t)_{t\geq 0}$ on $(\Omega,\mathscr{F},\mathbb{P})$ when $\{T \leq t\} \in \mathscr{F}_t$ for all $t \geq 0$. All constant non-negative random variables are stopping times.

Contrary to discrete time filtrations, the notion of stopping times for continuous time filtrations is deeply subtle and leads naturally to the notions of complete filtration and right continuous filtration.

Theorem 2.3.2. Hitting times as archetypal examples of stopping times.

Let $X = (X_t)_{t\geq 0}$ be a continuous and adapted process on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ with respect to a complete filtration $(\mathscr{F}_t)_{t\geq 0}$, and taking its values in a metric space G equipped with its Borel σ-field. Then, for all closed subset $A \subset G$, the hitting time $T_A : \Omega \to [0,+\infty]$ of A, defined by
$$T_A = \inf\{t \geq 0 : X_t \in A\},$$

with convention $\inf\varnothing = +\infty$, is a stopping time.

For instance $T_n = T_{[n,\infty)} = \inf\{t \geq 0 : |X_t| \geq n\}$ when $G = \mathbb{R}^d$.

Proof. Let Ω' be the almost sure event on which X is continuous. On Ω', since X is continuous and A is closed, we have $\{t \geq 0 : X_t \in A\} = \{t \geq 0 : \mathrm{dist}(X_t, A) = 0\}$, the map $t \geq 0 \mapsto \mathrm{dist}(X_t, A)$ is continuous, and the inf in the definition of T_A is a min. Now, since X is adapted, we have, for all $t \geq 0$,

$$\Omega' \cap \{T_A \leq t\} = \Omega' \cap \bigcap_{s \in [0,t] \cap \mathbb{Q}} \{X_s \in A\} \in \mathscr{F}_t,$$

where we have also used the fact that $\Omega' \in \mathscr{F}_t$ for all $t \geq 0$ since $(\mathscr{F}_t)_{t\geq 0}$ is complete. On the other hand, $(\Omega \setminus \Omega') \cap \{T_A \leq t\} \subset \Omega \setminus \Omega'$ is negligible, and belongs then to $\mathscr{F}_t$ for all $t \geq 0$ since $(\mathscr{F}_t)_{t\geq 0}$ is complete. ∎

We say that a filtration $(\mathscr{F}_t)_{t\geq 0}$ is right-continuous when $\mathscr{F}_t = \mathscr{F}_{t^+}$ for all $t \geq 0$ where

$$\mathscr{F}_{t+} = \bigcap_{\varepsilon > 0} \mathscr{F}_{t+\varepsilon} = \bigcap_{s > t} \mathscr{F}_s.$$

Theorem 2.3.3. Stopping times: alternative definition.

If $T : \Omega \to [0, +\infty]$ is a stopping time with respect to a filtration $(\mathscr{F}_t)_{t\geq 0}$ then $\{T < t\} \in \mathscr{F}_t$ for all $t \geq 0$. Conversely this property implies that T is a stopping time when the filtration is right-continuous.

Proof. If T is a stopping time then for all $t \geq 0$ we have

$$\{T < t\} = \bigcup_{n=1}^{\infty} \{T \leq t - \tfrac{1}{n}\} \in \mathscr{F}_t,$$

(note also that $\{T = t\} = \{T \leq t\} \cap \{T < t\}^c \in \mathscr{F}_t$). Conversely $\{T \leq t\} \in \cap_{s>t} \mathscr{F}_s = \mathscr{F}_{t+}$ since for all $s > t$,

$$\{T \leq t\} = \bigcap_{n=1}^{\infty} \{T < (t + \tfrac{1}{n}) \wedge s\} \in \mathscr{F}_s.$$

∎

This can be skipped at first reading.

The following generalizes Theorem 2.3.2 to hitting times of arbitrary measurable subsets by progessive processes, at the price of assuming right continuity of the filtration in addition to completeness.

Theorem 2.3.4: Hitting times are stopping times reloaded.

Let $X = (X_t)_{t\geq 0}$ be a progressive process defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ equipped with a right continuous and complete filtration $(\mathscr{F}_t)_{t\geq 0}$, and taking its values in a measurable space G. Then for all measurable subset $A \subset G$, the hitting time $T_A : \Omega \to [0, +\infty]$ defined by

$$T_A = \inf\{t \geq 0 : X_t \in A\},$$

with convention $\inf\varnothing = +\infty$, is a stopping time.

Example of progressive processes include adapted right-continuous processes.

Proof. The debut D_B of any $B \in \mathscr{F} \otimes \mathscr{B}(\mathbb{R}_+)$ is defined for all $\omega \in \Omega$ by

$$D_B(\omega) = \inf\{t \geq 0 : (\omega, t) \in B\} \in [0, +\infty].$$

If B is progressive, then D_B is a stopping time (this is known as the debut theorem). Indeed, for all $t \geq 0$ the set $\{D_B < t\}$ is then the projection on Ω of $\{s \in [0, t) : (\omega, s) \in B\}$, which belongs to $\mathscr{B}(\mathbb{R}_+) \otimes \mathscr{F}_t$

since B is progressive. Since the filtration is right-continuous and complete, this projection[a] belongs to $\mathscr{F}_t$. Now $\{D_B < t\} \in \mathscr{F}_t$ for all $t \geq 0$ implies that D_B is a stopping time since the filtration is right continuous (Theorem 2.3.3). Finally it remains to note that $T_A = D_B$ with $B = \{(\omega, t) : X_t \in A\}$, which is progressive as the pre-image of $\mathbb{R}_+ \times A$ by the map $(\omega, t) \mapsto X_t(\omega)$ (recall that X is progressive). ■

[a] See [5, Th. IV.50 page 116]. This is related to a famous mistake made by the French Henri Lebesgue (1875 – 1941) on the measurability of projections of measurable sets in product spaces, that motivated the Russian Nikolai Luzin (1883 – 1950) and his student Mikhail Yakovlevich Suslin (1894 – 1919) to forge the concept of analytic set and descriptive set theory.

Remark 2.3.5. Canonical filtration.

It is customary to assume that the underlying filtration is right-continuous and complete. For a given filtration $(\mathscr{F}_t)_{t\geq 0}$, it is always possible to consider its completion $(\sigma_t)_{t\geq 0} = (\sigma(\mathscr{N} \cup \mathscr{F}_t))_{t\geq 0}$ where $\mathscr{N}$ is the collection of negligible subsets of $\mathscr{F}$. It is also customary to consider the right-continuous version $(\sigma_{t+})_{t\geq 0}$, called the canonical filtration. A process is always adapted with respect to the canonical filtration constructed from its completed natural filtration.

From now on and unless otherwise stated we make the "canonical assumption": we assume that the underlying filtration is complete and right-continuous.

Remark 2.3.6. Subtleties about righ-continuity of filtrations.

The natural filtration of a right-continuous process is not right-continuous in general, indeed a counter example is given by $X_t = tZ$ for all $t \geq 0$ where Z is a non-constant random variable, since $\sigma(X_0) = \{\varnothing, \Omega\}$ while $\sigma(X_{0+\varepsilon} : \varepsilon > 0) = \sigma(Z) \neq \sigma(X_0)$. However it can be shown that the completion of the natural filtration of a "Feller Markov process" – including all Lévy processes and in particular Brownian motion – is always right-continuous.

Theorem 2.3.7. Stopping times properties.

Let S, T, and T_n, $n \geq 0$ be stopping times for some underlying filtration $(\mathscr{F}_t)_{t\geq 0}$ on an underlyning probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Then:

1. the following family is a σ-algebra called the stopping σ-algebra:
$$\mathscr{F}_T = \{A \in \mathscr{F} : \forall t \geq 0, A \cap \{T \leq t\} \in \mathscr{F}_t\}.$$
Moreover the stopping time T is $\mathscr{F}_T$-measurable
2. $X = (X_t)_{t\geq 0}$ is adapted then the stopped process $X^T = (X_{t\wedge T})_{t\geq 0}$ is also adapted. Moreover
$$(X^T)^S = X^{S\wedge T} = (X^S)^T$$
3. if $(X_t)_{t\geq 0}$ is adapted and progressive and if T is a.s. finite then $X^T = (X_{t\wedge T})_{t\geq 0}$ is progressive
4. if $X = (X_t)_{t\geq 0}$ is adapted and right-continuous then $Z = X_T \mathbf{1}_{T<\infty}$ is $\mathscr{F}_T$-mesurable
5. if $S \leq T$ then $\mathscr{F}_S \subset \mathscr{F}_T$
6. $S \wedge T$ and $S \vee T$ are stopping times and in particular $\mathscr{F}_{S\wedge T} \subset \mathscr{F}_{S\vee T}$
7. if $(\mathscr{F}_t)_{t\geq 0}$ is right-continuous then $\underline{\lim}_n T_n$ and $\overline{\lim}_n T_n$ are stopping times and
$$\cap_n \mathscr{F}_{T_n} = \mathscr{F}_{\inf_n T_n}.$$

Proof. The proof of the first three items are left as exercises.

4. Let $B \in \mathscr{B}_{\mathbb{R}}$ and $t \geq 0$. Then we have:
$$\{Z \in B\} \cap \{T \leq t\} = \{X_{T\wedge t} \in B\} \cap \{T \leq t\}.$$
Now we consider the composition of measurable maps:
$$\omega \in (\Omega, \mathscr{F}_t) \mapsto (\sigma(\omega) \wedge t, \omega) \in ([0,t] \times \Omega, \mathscr{B}_{[0,1]} \otimes \mathscr{F}_t) \mapsto X_{\sigma(\omega)\wedge t}(\omega) \in (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$$
and we use the fact that X is progressive.
5. If $A \in \mathscr{F}_S$ then, for all $t \geq 0$, $A \cap \{T \leq t\} = A \cap \{S \leq t\} \cap \{T \leq t\} \in \mathscr{F}_t$, hence $A \in \mathscr{F}_T$.
6. For all $t \geq 0$ we have
$$\{S \wedge T > t\} = \{S > t\} \cap \{T > t\} \in \mathscr{F}_t \quad \text{and} \quad \{S \vee T \leq t\} = \{S \leq t\} \cap \{T \leq t\} \in \mathscr{F}_t.$$
7. It suffice to show that $\sup_n T_n$ and $\inf_n T_n$ are stopping times. But
$$\{\sup_n T_n \leq t\} = \cap_n \{T_n \leq t\} \in \mathscr{F}_t \quad \text{and} \quad \{\inf_n T_n < t\} = \cup_n \{T_n < t\} \in \mathscr{F}_t$$
and therefore
$$\{\inf_n T_n \leq t\} = \cap_{\varepsilon>0} \{\inf_n T_n < t + \varepsilon\} \in \mathscr{F}_{t+} = \mathscr{F}_t.$$
Let $A \in \cap_n \mathscr{F}_{T_n}$. Then
$$A \cap \{\inf_n T_n < t\} = \cup_n A \cap \{T_n < t\} \in \mathscr{F}_t.$$
Therefore
$$A \cap \{\inf_n T_n \leq t\} \in \mathscr{F}_{t+} = \mathscr{F}_t.$$

■

Remark 2.3.8. Truncation via cutoff stopping times for continuous processes.

Truncation is an important tool in probability theory, and allows for instance to prove the strong law of large numbers for i.i.d. integrable random variables by reduction to the case of more integrable random variables. This tool is also available for stochastic processes, and its version with cutoff stopping times has the advantage of keeping the martingale structure (Doob stopping, Theorem 2.5.1). Let $X = (X_t)_{t\geq 0}$ be adapted. For all n we introduce the "truncation" or "cutoff" stopping time

$$T_n = \inf\{t \geq 0 : |X_t| \geq n\},$$

which takes its values in $[0, +\infty]$. We have $T_n \leq T_{n+1}$ for all n. If X is continuous then almost surely[a].

$$T_n \underset{n\to\infty}{\nearrow} +\infty.$$

Still if X is additionally continuous then almost surely and for all $n \geq 1$ and all $t \geq 0$,

$$|X_{t\wedge T_n}| \leq n\mathbf{1}_{|X_0|\leq n} + |X_0|\mathbf{1}_{|X_0|>n}.$$

If $X_0 = 0$ then the process $|X^{T_n}|$ is bounded by n for all $n \geq 1$. This is useful in this course[b].

[a] Indeed, almost surely, either the trajectory of X is bounded then $T_n = +\infty$ for large enough n beyond a (random) threshold, or the trajectory of X is unbounded and then by definition of being continuous and unbounded we have $T_n \nearrow +\infty$ as $n \to \infty$. Without continuity X_t could take arbitrary large values near a finite time forcing $(T_n)_n$ to be bounded.

[b] Localization is efficient for continuous processes issued from the origin. If X is discontinuous and in particular if it is a discrete time process, then, due to a possible jump at time T_n, we could have $|X_{T_n}| > n$ even if $X_0 = 0$ and n is large.

2.4 Martingales, sub-martingales, super-martingales

We restrict for simplicity to continuous martingales/sub-martingales/super-martingales. But many of the results remain actually valid for right-continuous martingales/sub-martingales/super-martingales.

The notion of martingale implements the idea of updating with a conditionally independent ingredient.

Definition 2.4.1. Martingales, sub-martingales, super-martingales.

Let $X = (X_t)_{t\geq 0}$ be a real adapted and integrable process in the sense that for all $t \geq 0$, X_t is measurable for $\mathscr{F}_t$ and $X_t \in \mathrm{L}^1$. Then, when

- $\mathbb{E}(X_t \mid \mathscr{F}_s) \geq X_s$ for all $t \geq 0$ and all $s \in [0,t]$, we say that X is a sub-martingale,
- $\mathbb{E}(X_t \mid \mathscr{F}_s) = X_s$ for all $t \geq 0$ and all $s \in [0,t]$, we say that X is a martingale
- $\mathbb{E}(X_t \mid \mathscr{F}_s) \leq X_s$ for all $t \geq 0$ and all $s \in [0,t]$, we say that X is a super-martingale.

These three notions can be seen in a sense as a probabilistic counterpart of the notions of increasing sequence, constant sequence, and decreasing sequence in basic classical analysis.

- For a sub-martingale, $t \mapsto \mathbb{E}(X_t)$ grows and in particular $\mathbb{E}(X_t) \geq \mathbb{E}(X_0)$ for all $t \geq 0$
- For a martingale, $t \mapsto \mathbb{E}(X_t)$ is constant, namely $\mathbb{E}(X_t) = \mathbb{E}(X_0)$ for all $t \geq 0$. It is a conservation law
- For a super-martingale, $t \mapsto \mathbb{E}(X_t)$ decreases and in particular $\mathbb{E}(X_t) \leq \mathbb{E}(X_0)$ for all $t \geq 0$.

The set of martingales is the intersection of the set of sub-martingales and the set of super-martingales.
A super-martingale or sub-martingale is a martingale if and only if its expectation is constant along time.
Being a martingale for a given filtration is a property stable by linear combinations.

If M is a martingale and if $(t_n)_{n\geq 0}$ is a strictly increasing sequence of times then the sequence of random variables $(M_{t_n})_{n\geq 0}$ is a discrete time martingale. We will try to avoid using discrete time martingales, but we will sometimes discretize time, notably to handle stopping times, which is roughly the same. The theory of discrete time martingales is similar to the theory of continuous time martingales that we develop here and comes with very similar theorems. In this course, most stochastic processes are in continuous time, and when we say "continuous process/martingale/etc", we mean that the process has continuous sample paths.

Example 2.4.2. Martingales.

1. If $Y \in \mathrm{L}^1$ then the process $(X_t)_{t\geq 0}$ defined by $X_t = \mathbb{E}(Y \mid \mathscr{F}_t)$ for all $t \geq 0$ is a martingale with respect to $(\mathscr{F}_t)_{t\geq 0}$ known as the Doob martingale or a closed martingale. It is uniformly integrable. Corollary 4.4.5 provides a sort of converse (u.i. martingales are closed)
2. If $(X_t)_{t\geq 0}$ is a martingale and if $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and such that $\varphi(X_t) \in \mathrm{L}^1$ for all $t \geq 0$, then by the Jensen inequality for conditional expectation, $(Y_t)_{t\geq 0} = (\varphi(X_t))_{t\geq 0}$ is a sub-martingale for the same filtration. In particular $(|X_t|)_{t\geq 0}$, $(X_t^2)_{t\geq 0}$, and $(\mathrm{e}^{X_t})_{t\geq 0}$ are sub-martingales
3. If $(X_t)_{t\geq 0}$ is a sub-martingale and if $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and non-decreasing such that $\varphi(X_t) \in \mathrm{L}^1$ for all $t \geq 0$, then by the Jensen inequality for condition expectation, $(Y_t)_{t\geq 0} = (\varphi(X_t))_{t\geq 0}$ is a sub-martingale for the same filtration. In particular $(\mathrm{e}^{X_t})_{t\geq 0}$ is a sub-martingale
4. A martingale $X = (X_t)_{t\geq 0}$ is also a martingale for its natural filtration $(\sigma(X_s : s \in [0,t]))_{t\geq 0}$
5. If $(E_n)_{n\geq 1}$ are independent and identically distributed exponential random variables of mean $1/\lambda$, then, for all $t \geq 0$, the number of these random variables falling in the interval $[0,t]$ is $N_t = \mathrm{card}\{n \geq 1 : E_n \in [0,t]\}$. It is known that the counting process $(N_t)_{t\geq 0}$ has independent and stationary increments of Poisson law, namely for all $n \geq 1$ and $0 = t_0 \leq \cdots \leq t_n$, the random variables $N_{t_1} - N_{t_0}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent of law $\mathrm{Poi}(\lambda(t_1 - t_0)), \ldots, \mathrm{Poi}(\lambda(t_n - t_{n-1}))$. We say that $(N_t)_{t\geq 0}$ is the simple Poisson process of intensity λ. Now for the (natural) filtration $(\mathscr{F}_t)_{t\geq 0}$, $\mathscr{F}_t = \sigma(N_s : 0 \leq s \leq t)$, and for all $c \in \mathbb{R}$, the process $(N_t - ct)_{t\geq 0}$ is a sub-martingale if $c < \lambda$, a martingale if $c = \lambda$, and a super-martingale if $c > \lambda$. Namely, for all $0 \leq s \leq t$,

$$\begin{aligned}\mathbb{E}(N_t - ct \mid \mathscr{F}_s) &= \mathbb{E}(N_t - N_s - c(t-s) + N_s - cs \mid \mathscr{F}_s)\\ &= \mathbb{E}(N_t - N_s) - c(t-s) + N_s - cs\end{aligned}$$

$$= (\lambda - c)(t-s) + N_s - cs.$$

This process is not continuous, but has right-continuous and left limited trajectories (càdlàg[a]).

6. If $(N_t)_{t\geq 0}$ is the simple Poisson process of intensity λ as above, then, for all $0 \leq s \leq t$,

$$\mathbb{E}(\mathrm{e}^{N_t - ct} \mid \mathscr{F}_s) = \mathrm{e}^{N_s - cs}\mathbb{E}(\mathrm{e}^{N_t - N_s})\mathrm{e}^{-c(t-s)} = \mathrm{e}^{N_s - cs}\mathrm{e}^{\lambda(t-s)(\mathrm{e}-1)-c(t-s)}.$$

It follows that for the natural filtration of $(N_t)_{t\geq 0}$, the process $(\mathrm{e}^{N_t-ct})_{t\,t\geq 0}$ is a sub-martingale if $c < \lambda(\mathrm{e}-1)$, a martingale if $c = \lambda(\mathrm{e}-1)$, and a super-martingale if $c > \lambda(\mathrm{e}-1)$. We often say that $(\mathrm{e}^{N_t-ct})_{t\geq 0}$ is an exponential (sub/super-)martingale.

7. The Brownian motion $(B_t)_{t\geq 0}$ of Chapter 3 has independent and stationary Gaussian increments: for all $n \geq 1$ and $0 = t_0 \leq \cdots \leq t_n$ the random variables $B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent of law $\mathcal{N}(0, t_1 - t_0), \ldots, \mathcal{N}(0, t_n - t_{n-1})$. Thus the process $(B_t)_{t\geq 0}$ is a martingale for its natural filtration, indeed, for all $0 \leq s \leq t$,

$$\mathbb{E}(B_t \mid \mathscr{F}_s) = \mathbb{E}(B_t - B_s + B_s \mid \mathscr{F}_s) = \mathbb{E}(B_t - B_s) + B_s = B_s.$$

This process has continuous trajectories. Moreover and similarly, for all $c \in \mathbb{R}$, the process $(B_t^2 - ct)_{t\geq 0}$ is a sub-martingale if $c < 1$, a martingale if $c = 1$, and a super-martingale if $c > 1$. The key is to use the decomposition $B_t = (B_t - B_s)^2 + 2B_sB_t - B_s^2$. We can also study the process $\mathrm{e}^{B_t - ct}$ and seek for a condition on c to get a martingale, and we speak about an exponential martingale. For simplicity, most of the martingales encountered in this course are continuous.

[a] *Continu à droite avec limites à gauche.*

2.5 Doob stopping theorem and maximal inequalities

Stopped martingales are martingales, and the conservation law extends to stopping times:

Theorem 2.5.1. Doob[a] stopping theorem.

[a] Named after Joseph L. Doob (1910–2004), American mathematician.

If M is a continuous martingale and $T : \Omega \to [0, +\infty]$ is a stopping time then $M^T = (M_{t\wedge T})_{t\geq 0}$ is a (continuous) martingale, namely for all $t \geq 0$ and $s \in [0, t]$, we have

$$M_{t\wedge T} \in \mathrm{L}^1 \quad \text{and} \quad \mathbb{E}(M_{t\wedge T} \mid \mathscr{F}_s) = M_{s\wedge T}.$$

Moreover, if T is bounded, or if T is almost surely finite and $(M_{t\wedge T})_{t\geq 0}$ is u.i.[a], then

$$M_T \in \mathrm{L}^1 \quad \text{and} \quad \mathbb{E}(M_T) = \mathbb{E}(M_0).$$

[a] For instance dominated by an integrable random variable, or even bounded by a constant.

In practice, the best is to retain that $(M_{t\wedge T})_{t\geq 0}$ is a martingale. We have $\lim_{t\to\infty} M_{T\wedge t}\mathbf{1}_{T<\infty} = M_T\mathbf{1}_{T<\infty}$ a.s. When $T < \infty$ a.s. we could use what we know on M and T to deduce by monotone or dominated convergence that this holds in L^1, giving $\mathbb{E}(M_T) = \mathbb{E}(\lim_{t\to\infty} M_{t\wedge T}) = \lim_{t\to\infty}\mathbb{E}(M_{t\wedge T}) = \mathbb{E}(M_0)$. Theorem 2.5.1 states that this is automatically the case when T is bounded or when M^T is u.i. Furthermore, if M^T is u.i. then it can be shown that M_∞ exists, giving a sense to M_T even on $\{T = \infty\}$, and then $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.

Proof. Let assume first that T takes a finite number of values $t_1 < \cdots < t_n$. Let us show that $M_T \in \mathrm{L}^1$ and $\mathbb{E}(M_T) = \mathbb{E}(M_0)$. We have $M_T = \sum_{k=1}^n M_{t_k}\mathbf{1}_{T=t_k} \in \mathrm{L}^1$, and moreover, using $\{T \geq t_k\} = (\cup_{i=1}^{k-1}\{T = t_i\})^c \in \mathscr{F}_{t_{k-1}}$, and the martingale property $\mathbb{E}(M_{t_k} - M_{t_{k-1}} \mid \mathscr{F}_{t_{k-1}}) = 0$, for all k, we get

$$\mathbb{E}(M_T) = \mathbb{E}(M_0) + \mathbb{E}\Big(\sum_{k=1}^{n} \mathbb{E}(M_{t_k} - M_{t_{k-1}} \mid \mathscr{F}_{t_{k-1}})\mathbf{1}_{T \geq t_k}\Big) = \mathbb{E}(M_0).$$

Suppose now that T takes an infinite number of values but is bounded by some constant C. For all $n \geq 0$, we approximate T by the piecewise constant random variable (discretization of $[0, C]$)[1]

$$T_n = C\mathbf{1}_{T=C} + \sum_{k=1}^{n} t_k \mathbf{1}_{t_{k-1} \leq T < t_k} \quad \text{where} \quad t_k = t_{n,k} = C\frac{k}{n}.$$

This is a stopping time since it takes discrete values and for all $m \geq 0$,

$$\{T_n = m\} = \begin{cases} \varnothing \in \mathscr{F}_0 & \text{if } m \not\in \{t_k : 1 \leq k \leq n\} \\ \{T = C\} \in \mathscr{F}_C & \text{if } m = C \\ \{T < t_{k-1}\}^c \cap \{T < t_k\} \in \mathscr{F}_{t_k} & \text{if } m = t_k,\ 1 \leq k \leq n \end{cases}$$

where we used the fact that $\{T = t\} = \{T \leq t\} \cap \{T < t\}^c = \{T \leq t\} \cap \cap_{r=1}^{\infty} \{T > t - 1/r\} \in \mathscr{F}_t$ for all $t \geq 0$.

Since T_n takes a finite number of values, the previous step gives $\mathbb{E}(M_{T_n}) = \mathbb{E}(M_0)$. On the other hand, almost surely, $T_n \to T$ as $n \to \infty$. Since M is continuous, it follows that almost surely $M_{T_n} \to M_T$ as $n \to \infty$. Let us show now that $(M_{T_n})_{n\geq 1}$ is uniformly integrable. Since for all $n \geq 0$, T_n takes its values in a finite set $t_1 < \cdots < t_{m_n} \leq C$, the martingale property[2] and the Jensen inequality give, for all $R > 0$,

$$\begin{aligned} \mathbb{E}(|M_{T_n}|\mathbf{1}_{|M_{T_n}| \geq R}) &= \sum_k \mathbb{E}(|M_{t_k}|\mathbf{1}_{|M_{t_k}| \geq R, T_n = t_k}) \\ &= \sum_k \mathbb{E}(|\mathbb{E}(M_C \mid \mathscr{F}_{t_k})|\mathbf{1}_{|M_{t_k}| \geq R, T_n = t_k}) \\ &\leq \sum_k \mathbb{E}(\mathbb{E}(|M_C| \mid \mathscr{F}_{t_k})\mathbf{1}_{|M_{t_k}| \geq R, T_n = t_k}) \\ &= \sum_k \mathbb{E}(|M_C|\mathbf{1}_{|M_{t_k}| \geq R, T_n = t_k}) \\ &= \mathbb{E}(|M_C|\mathbf{1}_{|M_{T_n}| \geq R}). \end{aligned}$$

Now M is continuous and thus locally bounded, and $M_C \in \mathrm{L}^1$, thus, by dominated convergence,

$$\sup_n \mathbb{E}(|M_{T_n}|\mathbf{1}_{|M_{T_n}| > R}) \leq \mathbb{E}(|M_C|\mathbf{1}_{\sup_{s\in[0,C]}|M_s| \geq R}) \underset{R\to\infty}{\longrightarrow} 0.$$

Therefore $(M_{T_n})_{n\geq 0}$ is uniformly integrable. As a consequence

$$\overset{\text{a.s.}}{\lim_{n\to\infty}} M_{T_n} = M_T \in \mathrm{L}^1 \quad \text{and} \quad \mathbb{E}(M_T) = \lim_{n\to\infty} \mathbb{E}(M_{T_n}) = \mathbb{E}(M_0).$$

Let us suppose now that T is an arbitrary stopping time. For all $0 \leq s \leq t$ and $A \in \mathscr{F}_s$, the random variable $S = s\mathbf{1}_A + t\mathbf{1}_{A^c}$ is a (finite) stopping time, and what precedes for the finite stopping time $t \wedge T \wedge S$ gives $M_{t\wedge T\wedge S} \in \mathrm{L}^1$ and $\mathbb{E}(M_{t\wedge T\wedge S}) = \mathbb{E}(M_0)$. Now, using the definition of S, we have

$$\mathbb{E}(M_0) = \mathbb{E}(M_{t\wedge T\wedge S}) = \mathbb{E}(\mathbf{1}_A M_{s\wedge T}) + \mathbb{E}(\mathbf{1}_{A^c} M_{t\wedge T}) = \mathbb{E}(\mathbf{1}_A(M_{s\wedge T} - M_{t\wedge T})) + \mathbb{E}(M_{t\wedge T}).$$

Since $\mathbb{E}(M_{t\wedge T}) = \mathbb{E}(M_0)$, we get $\mathbb{E}((M_{t\wedge T} - M_{s\wedge T})\mathbf{1}_A) = 0$, namely the martingale property for $(M_{t\wedge T})_{t\geq 0}$.

Finally, suppose that $T < \infty$ a.s. and $(M_{t\wedge T})_{t\geq 0}$ is u.i. The random variable M_T is well defined and $\lim_{t\to\infty} M_{t\wedge T} = M_T$ a.s. because M is continuous. Furthermore, since $(M_{t\wedge T})_{t\geq 0}$ is u.i., it follows that $M_T \in \mathrm{L}^1$ and $\lim_{t\to\infty} M_{t\wedge T} = M_T$ in L^1. In particular $\mathbb{E}(M_0) \underset{\forall t}{=} \mathbb{E}(M_{t\wedge T}) = \lim_{t\to\infty} \mathbb{E}(M_{t\wedge T}) = \mathbb{E}(M_T)$. ■

[1] By using dyadics, we could define T_n in such a way that $T_n \searrow T$, giving $M_{T_n} \to M_T$ pointwise when M is right-continuous.
[2] It also works for non-negative sub-martingales.

Example 2.5.2. Example of application of Doob stopping theorem.

Let $(M_t)_{t\geq 0}$ be a continuous martingale, $a<b$, and $T=\inf\{t\geq 0 : M_t\in\{a,b\}\}$ the hitting time of the boundary of $[a,b]$. Suppose that M_0 takes its values in $[a,b]$ and that T is almost surely finite[a]. Then on the one hand, we have the equation $\mathbb{P}(M_T=a)+\mathbb{P}(M_T=b)=1$. On the other hand, by definition of T the process $(M_{t\wedge T})_{t\geq 0}$ is bounded and thus u.i. and the Doob stopping theorem (Theorem 2.5.1) gives then $x=\mathbb{E}(M_0)=\mathbb{E}(M_T)=a\mathbb{P}(M_T=a)+b\mathbb{P}(M_T=b)$. It follows by combining the equations that

$$\mathbb{P}(M_T=a)=\frac{b-x}{b-a}\quad\text{and}\quad\mathbb{P}(M_T=b)=\frac{x-a}{b-a}$$

(note that $x\in[a,b]$). This holds in particular for Brownian motion started from $x\in[a,b]$, and by using an exponential martingale, it is then even possible to compute the Laplace transform of T.

[a] Holds for BM B with $B_0=0\in(a,b)$ since $\mathbb{P}(T=\infty)\leq\mathbb{P}(T>t)\leq\mathbb{P}(B_t\in(a,b))=\mathbb{P}(\sqrt{t}Z\in(a,b))\to 0$ as $t\to\infty$.

Coding in action 2.5.3. Gambler's ruin.

Physically Brownian motion and the simple symmetric random walk are the same, it is just a matter of scale. Fix $a\leq b$ in $\mathbb{Z}$. Write a code to plot on the same graphic multiple trajectories of such a random walk started from various values of $x\in[a,b]$ and stopped when it reaches a or b. Could you verify numerically the formulas of Example 2.5.2? And mathematically?

Remark 2.5.4. Counter example with an unbounded stopping time.

If M is a continuous martingale with $M_0=0$, then, for all $a>0$, $T_a=\inf\{t>0 : M_t=a\}$ is a stopping time, but it cannot be bounded since this would give $0=\mathbb{E}(M_0)=\mathbb{E}(M_{T_a})=a>0$, a contradiction!

The following variant of the Doob stopping is useful in many applications.

Theorem 2.5.5. Doob stopping theorem for sub-martingales.

If M is a continuous sub-martingale and S and T are bounded stopping times such that $S\leq T$, $M_S\in\mathrm{L}^1$, and $M_T\in\mathrm{L}^1$, then $\mathbb{E}(M_S)\leq\mathbb{E}(M_T)$.

Proof. We proceed as in the proof of Theorem 2.5.1, by assuming first that S and T take their values in the finite set $\{t_1,\ldots,t_n\}$ where $t_1<\cdots<t_n$. In this case M_T and M_S are in L^1 automatically. The inequality $S\leq T$ gives $\mathbf{1}_{S\geq t}\leq\mathbf{1}_{T\geq t}$ for all t. Using this fact and the sub-martingale property of M, we get

$$\mathbb{E}(M_S)=\mathbb{E}(M_0)+\mathbb{E}\Big(\sum_{k=1}^n\underbrace{\mathbb{E}(M_{t_k}-M_{t_{k-1}}\mid\mathcal{F}_{t_{k-1}})}_{\geq 0}\mathbf{1}_{S\geq t_k}\Big)\leq\mathbb{E}(M_0)+\mathbb{E}\Big(\sum_{k=1}^n\mathbb{E}(M_{t_k}-M_{t_{k-1}}\mid\mathcal{F}_{t_{k-1}})\mathbf{1}_{T\geq t_k}\Big)=\mathbb{E}(M_T).$$

More generally, when S and T are arbitrary bounded stopping times satisfying $S\leq T$, and at least when M is a non-negative sub-martingale, we can proceed by approximation as in the proof of Theorem 2.5.1. ■

This can be skipped at first reading.

Theorem 2.5.6: Doob stopping theorem for non-negative super-martingales

If M is a continuous non-negative super-martingale and S and T are stopping times such that $S\leq T$, then $M_S\in\mathrm{L}^1$ and $M_T\in\mathrm{L}^1$ and $\mathbb{E}(M_S)\geq\mathbb{E}(M_T\mid\mathcal{F}_S)$, in particular $\mathbb{E}(M_S)\geq\mathbb{E}(M_T)$.

When S and T are bounded we recover Theorem 2.5.6 in the special case where $M\leq 0$.

Proof. See for instance [14, Theorem 3.25 pages 64–65]. Note that since M is a non-negative super-martingale, it is automatically bounded in L^1 since $0\leq\mathbb{E}(M_t)\leq\mathbb{E}(M_0)$ for all $t\geq 0$. ■

The following theorem allows to control the tail of the supremum of a martingale over a time interval by the moment at the terminal time. It is a continuous time martingale version of the simpler Kolmogorov maximal inequality for sums of independent and identically distributed random variables.

Theorem 2.5.7. Doob maximal inequalities.

1. If M is a continuous martingale or non-negative sub-martingale then for all $p \geq 1$, $t \geq 0$, $\lambda > 0$,

$$\mathbb{P}\Big(\max_{s\in[0,t]} |M_s| \geq \lambda\Big) \leq \frac{\mathbb{E}(|M_t|^p)}{\lambda^p}.$$

2. If M is a continuous martingale then for all $p > 1$ and $t \geq 0$,

$$\mathbb{E}\Big(\max_{s\in[0,t]} |M_s|^p\Big) \leq \Big(\frac{p}{p-1}\Big)^p \mathbb{E}(|M_t|^p) \quad \text{in other words} \quad \Big\|\max_{s\in[0,t]} |M_s|\Big\|_p \leq \frac{p}{p-1}\|M_t\|_p.$$

   In particular if $M_t \in \mathrm{L}^p$ then $M_t^* = \max_{s\in[0,t]} |M_s| \in \mathrm{L}^p$.

Note that $q = 1/(1-1/p) = p/(p-1)$ is the Hölder conjugate of p in the sense that $1/p + 1/q = 1$. The Doob inequality is often used with $p = 2$, for which $(p/(p-1))^p = 4$.

Proof. We can assume that the right hand side is finite ($M_t \in \mathrm{L}^p$), otherwise the inequalities are trivial.

1. If M is a martingale, then by the Jensen inequality for the convex function $u \in \mathbb{R} \mapsto |u|^p$, the process $|M|^p$ is a sub-martingale. Similarly, If M is a non-negative sub-martingale then, since $u \in [0,+\infty) \mapsto u^p$ is convex and non-decreasing it follows that $M^p = |M|^p$ is a sub-martingale. Therefore in all cases $(|M_s|^p)_{s\in[0,t]}$ is a sub-martingale. Let us define the bounded stopping time

$$T = t \wedge \inf\{s \geq 0 : |M_s| \geq \lambda\}.$$

   Since M is continuous we have $|M_T| \leq \max(|M_0|, \lambda)$ and thus $M_T \in \mathrm{L}^1$. The Doob stopping Theorem 2.5.5 for the sub-martingale $|M|^p$ and the bounded stopping times T and t that satisfy $T \leq t$ gives

$$\mathbb{E}(|M_T|^p) \leq \mathbb{E}(|M_t|^p).$$

   On the other hand the definition of T gives

$$|M_T|^p \geq \lambda^p \mathbf{1}_{\max_{s\in[0,t]}|M_s|\geq\lambda} + |M_t|^p \mathbf{1}_{\max_{s\in[0,t]}|M_s|<\lambda} \geq \lambda^p \mathbf{1}_{\max_{s\in[0,t]}|M_s|\geq\lambda}.$$

   It remains to combine these inequalities to get the desired result

2. We first reduce to the case where M satisfies $\max_{s\in[0,t]}|M_s| \in \mathrm{L}^p$. To do so, we introduce for all $n \geq 1$ the truncation or localization stopping time[3] $T_n = \inf\{s \geq 0 : |M_s| \geq n\}$. By the Doob stopping theorem (Theorem 2.5.1), the process $(M_{s\wedge T_n})_{s\in[0,t]}$ is a martingale. Moreover, since M is continuous, we have the domination $|M_{s\wedge T_n}| \leq |M_0| \wedge n$, and since $M_t \in \mathrm{L}^p$ gives $M_0 \in \mathrm{L}^p$, we obtain $\max_{s\in[0,t]}|M_{s\wedge T_n}| \in \mathrm{L}^p$. The desired inequality for the dominated martingale $(M_{s\wedge T_n})_{s\in[0,t]}$ would give

$$\mathbb{E}(\max_{s\in[0,t]} |M_{s\wedge T_n}|^p) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}(|M_{t\wedge T_n}|^p),$$

   and the desired result for $(M_s)_{s\in[0,t]}$ would then follow by monotone convergence theorem as $n \to \infty$ since then $|M_{s\wedge T_n}| \nearrow |M_s|$ for all $s \in [0,t]$. Thus this shows that we can assume without loss of generality that $\sup_{s\in[0,t]}|M_s| \in \mathrm{L}^p$. This is our first martingale localization argument!

   By using the proof of the first item with $p = 1$ and decomposing M_t as we did for M_T, we get

$$\mathbb{P}(\max_{s\in[0,t]} |M_s| \geq \lambda) \leq \frac{\mathbb{E}(|M_t|\mathbf{1}_{\max_{s\in[0,t]}|M_s|\geq\lambda})}{\lambda}$$

[3] Since we are only interested by the time interval $[0,t]$, we could take $\wedge t$ which makes the stopping time bounded.

for all $\lambda > 0$, and thus

$$\int_0^\infty \lambda^{p-1}\mathbb{P}(\max_{s\in[0,t]}|M_s| \geq \lambda)\mathrm{d}\lambda \leq \int_0^\infty \lambda^{p-2}\mathbb{E}(|M_t|\mathbf{1}_{\max_{s\in[0,t]}|M_s|\geq\lambda})\mathrm{d}\lambda.$$

Now the Fubini–Tonelli theorem gives

$$\int_0^\infty \lambda^{p-1}\mathbb{P}(\max_{s\in[0,t]}|M_s| \geq \lambda)\mathrm{d}\lambda = \mathbb{E}\int_0^{\max_{s\in[0,t]}|M_s|}\lambda^{p-1}\mathrm{d}\lambda = \frac{1}{p}\mathbb{E}(\max_{s\in[0,t]}|M_s|^p).$$

and similarly (here we need $p > 1$)

$$\int_0^\infty \lambda^{p-2}\mathbb{E}(|M_t|\mathbf{1}_{\max_{s\in[0,t]}|M_s|\geq\lambda})\mathrm{d}\lambda = \frac{1}{p-1}\mathbb{E}(|M_t|\max_{s\in[0,t]}|M_s|^{p-1}).$$

Combining all this gives

$$\mathbb{E}(\max_{s\in[0,t]}|M_s|^p) \leq \frac{p}{p-1}\mathbb{E}(M_t\max_{s\in[0,t]}|M_s|^{p-1}).$$

But since the Hölder inequality gives

$$\mathbb{E}(|M_t|\max_{s\in[0,t]}|M_s|^{p-1}) \leq \mathbb{E}(|M_t|^p)^{1/p}\mathbb{E}(\max_{s\in[0,t]}|M_s|^p)^{\frac{p-1}{p}},$$

we obtain

$$\mathbb{E}(\max_{s\in[0,t]}|M_s|^p) \leq \frac{p}{p-1}\mathbb{E}(|M_t|^p)^{1/p}\mathbb{E}(\max_{s\in[0,t]}|M_s|^p)^{\frac{p-1}{p}}.$$

Consequently, since $\mathbb{E}(\max_{s\in[0,t]}|M_s|^p) < \infty$, we obtain the desired inequality.

■

Example 2.5.8. A consequence of Doob maximal inequality.

Let $(M_t)_{t\geq 0}$ be a continuous martingale bounded in L^p, $p > 1$, namely

$$C_p = \sup_{t\geq 0}\mathbb{E}(|M_t|^p) < \infty.$$

It follows that M is u.i. But the Doob maximal inequality says more. Namely, by Theorem 2.5.7, for all $t \geq 0$, $\mathbb{E}(\max_{s\in[0,t]}|M_s|^p) \leq C_p$. The monotone convergence theorem gives then

$$\mathbb{E}(\sup_{t\geq 0}|M_t|^p) \leq C_p < \infty.$$

Therefore, almost surely $\sup_{t\geq 0}|M_t| < \infty$. In other words M has almost surely bounded trajectories. Beware however that the bound is random and may depend on the trajectory.

The following version of Doob maximal inequality is useful for some applications.

Theorem 2.5.9. Doob maximal inequality for super-martingales.

If M is a continuous super-martingale, then for all $t \geq 0$ and $\lambda > 0$, denoting $M^- = \max(0, -M)$,

$$\mathbb{P}\Big(\max_{s\in[0,t]}|M_s| \geq \lambda\Big) \leq \frac{\mathbb{E}(M_0) + 2\mathbb{E}(M_t^-)}{\lambda}.$$

In particular when M is non-negative then $\mathbb{E}(M^-) = 0$ and the upper bound becomes $\mathbb{E}(M_0)/\lambda$.

This can be skipped at first reading.

Proof. Let us define the bounded stopping time

$$T = t \wedge \inf\{s \in [0,t] : M_s \geq \lambda\}.$$

We have $M_T \in \mathrm{L}^1$ since $|M_T| \le \max(|M_0|, |M_t|, \lambda)$. By the Doob stopping Theorem 2.5.5 with the sub-martingale $-M$ and the bounded stopping times 0 and T that satisfy $M_0 \in \mathrm{L}^1$ and $M_T \in \mathrm{L}^1$, we get

$$\mathbb{E}(M_0) \ge \mathbb{E}(M_T) \ge \lambda \mathbb{P}(\max_{s \in [0,t]} M_s \ge \lambda) + \mathbb{E}(M_t \mathbf{1}_{\max_{s \in [0,t]} M_s < \lambda})$$

hence, recalling that $M^- = \max(-M, 0)$,

$$\lambda \mathbb{P}(\max_{s \in [0,t]} M_s \ge \lambda) \le \mathbb{E}(M_0) + \mathbb{E}(M_t^-).$$

This produces the desired inequality when M is non-negative. For the general case, we observe that the Jensen inequality for the non-decreasing convex function $u \in \mathbb{R} \mapsto \max(u, 0)$ and the sub-martingale $-M$ shows that M^- is a non-negative sub-martingale. Thus, by Theorem 2.5.1,

$$\lambda \mathbb{P}(\max_{s \in [0,t]} M_s^- \ge \lambda) \le \mathbb{E}(M_t^-).$$

Finally, putting both inequalities together gives

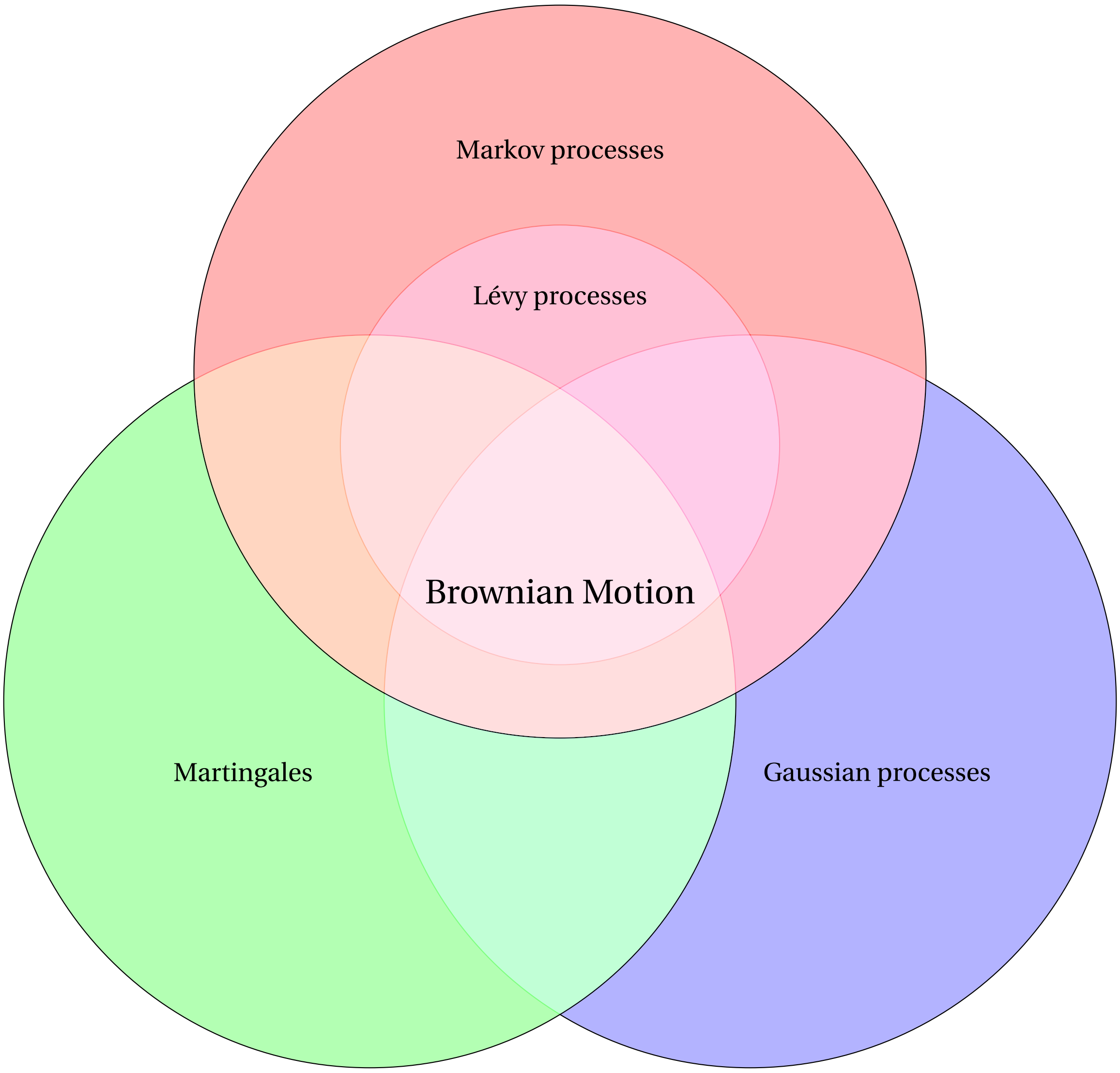
$$\lambda \mathbb{P}(\max_{s \in [0,t]} |M_s| \ge \lambda) \le \lambda \mathbb{P}(\max_{s \in [0,t]} M_s \ge \lambda) + \lambda \mathbb{P}(\max_{s \in [0,t]} M_s^- \ge \lambda) \le \mathbb{E}(M_0) + 2\mathbb{E}(M_t^-).$$

■

Chapter 3

Brownian motion

Just like the central limit theorem, Brownian motion is a physical as well as a mathematical phenomenon, see figures 3.1, 3.2, and 3.3. In this chapter, we study some properties of the mathematical Brownian motion.



For all $t > 0$, $d \geq 1$, the density of the Gaussian distribution $\mathcal{N}(0, tI_d)$ on $\mathbb{R}^d$ is

$$x \in \mathbb{R}^d \mapsto p_t(x) = \frac{e^{-\frac{|x|^2}{2t}}}{(\sqrt{2\pi t})^d} \quad \text{where} \quad |x|^2 = x_1^2 + \cdots + x_d^2.$$

We have, for all $s, t > 0$,

$$p_{t+s}(x) = (p_t * p_s)(x) = \int_{\mathbb{R}^d} p_t(x-z) p_s(z) \mathrm{d}z.$$

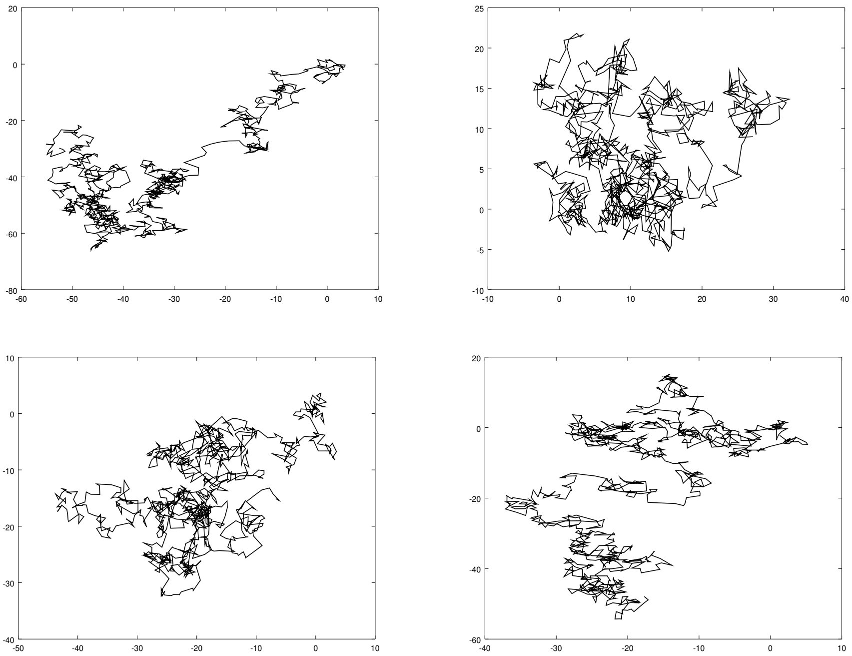


Figure 3.1: First steps of four approximated sample paths of 2-dimensional Brownian motion issued from the origin, numerically simulated with a Gaussian random walk via code `plot(cumsum(randn(2,1000)))`.

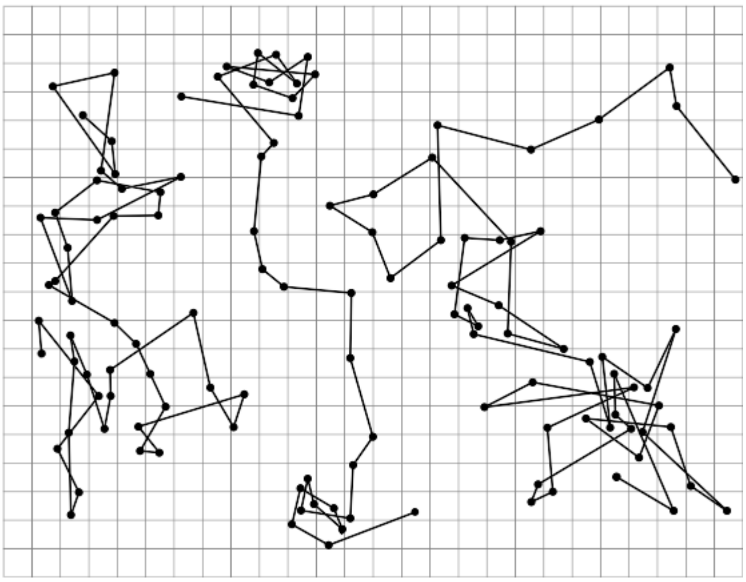


Figure 3.2: From the famous book [16] of Jean Perrin (1870 – 1942), three tracings of the motion of colloidal particles of radius 0.53 μm, as seen under the microscope are displayed. Successive positions every 30 seconds are joined by straight line segments (mesh size is 3.2 μm). These precise and systematic experiments, inspired by the historical ones by Robert Brown (1773 – 1858), allowed to test the atomistic theory of Ludwig Boltzmann (1944 – 1906), Albert Einstein (1879 – 1955), Marian Schmoluchovski (1872 – 1917), and others. "*Ainsi, la théorie moléculaire du mouvement brownien peut-être regardée comme expérimentalement établie, et, du même coup, il devient assez difficile de nier la réalité objective des molécules.*". Louis Bachelier (1870 – 1946) identified independently a similar physical phenomenon in the behavior of stock markets.

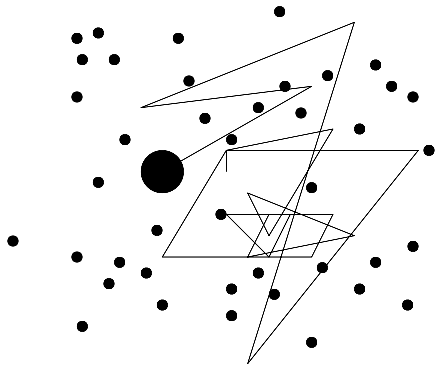


Figure 3.3: Atomistic interpretation of physical Brownian motion: a big particle of dust in a liquid is subject to a high number of collisions with the molecules of the liquid, which are much smaller and disordered by heat. This leads to the kinetic interpretation behind the Langevin equation. In reality, the diameter ratio is high, for instance the colloidal particle observed by Perrin has diameter of 0.57 µm while a molecule of water has a diameter of 0.27 nm, which gives a diameter ratio of about 2000. Moreover in reality the molecules density is high, the distance between molecules being of 0.31 nm for water. With this atomistic interpretation, physical Brownian motion is essentially a random walk, seen at a space-time scale which makes it close to mathematical Brownian motion, its idealistic scaling limit.

Definition 3.0.1. Brownian motion[a] or Wiener[b] process.

[a] Named after Robert Brown (1773 – 1858), Scottish botanist.
[b] Named after Norbert Wiener (1894 – 1964), American mathematician.

A d-dimensional Brownian motion (BM) is a d-dimensional process $B = (B_t)_{t\geq 0}$ which has:

1. Almost surely continuous trajectories, in the sense that B is a continuous process.
2. Stationary, Gaussian, independent increments:
   - for all $0 \leq s \leq t$, $B_t - B_s \sim \mathcal{N}(0,(t-s)I_d)$
   - for all $t_0 = 0 < t_1 < \cdots < t_n$, $n \geq 0$, $B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent.

Beware that there are no conditions on B_0, and in particular $B_t = B_0 + B_t - B_0$ may not be Gaussian.

```python
# Python program generating the graphic used for the lecture notes cover
import numpy as np ; import matplotlib.pyplot as pp
for i in range(1,11):
    pp.plot(np.cumsum(np.random.randn(1,1000)[0]),'k-',linewidth=1)
pp.axis('off') ; pp.show()
```

```julia
# Julia program generating the graphic used for the lecture notes cover
using Pkg ; Pkg.add("Plots") ; using Plots
for i=1:10
plot!(cumsum(randn(1000,1),dims = 1), lw = 1, legend = false, grid = false,
 ↪  axes=([],false))
end
gui()
```

Coding in action 3.0.2. Stochastic simulation.

By using the structure of the increments, write your own program simulating and plotting approximated trajectories of BM. Can we check numerically that the mathematical object of Brownian motion exists? Let D be a closed domain of $\mathbb{R}^d$ such as a disc or a square, containing the origin 0. Let ∂D be its boundary, let B be a BM with $B_0 = 0$, and let $T = \inf\{t \geq 0 : B_t \in \partial D\}$. Write a program simulating the law of T and the law of B_T, and producing nice plots when $d = 1$, $d = 2$, $d = 3$.

Remark 3.0.3. Gaussian[a] and Lévy[b] processes.

[a] Named after Carl Friedrich Gauss (1777 – 1855), German mathematician.
[b] Named after Paul Lévy (1886 – 1971), French mathematician.

For all $n \geq 1$ and $0 \leq t_1 < \cdots < t_n$ the random vector $(B_{t_1}, \ldots, B_{t_n})$ is Gaussian, and we say that Brownian motion is a Gaussian process. On the other hand, for all $n \geq 1$ and $0 = t_0 < \cdots < t_n$ the increments $B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent and stationary in the sense that their law depends only on the differences $t_1 - t_0, \ldots, t_n - t_{n-1}$ between successive times. Also Brownian motion has independent and stationary increments and such processes are called Lévy processes. They form a special subclass of Markov processes. The Poisson process considered in Example 2.4.2 is also a Lévy process, for which the increments are Poisson and the trajectories right continuous with left limits.

Remark 3.0.4. Finite dimensional laws.

A d-dimensional continuous process $X = (X_t)_{t \geq 0}$ issued from $x \in \mathbb{R}^d$ is a Brownian Motion iff for all $n \geq 0$, all $0 < t_1 < \cdots < t_n$, all $A_i \in \mathcal{B}_{\mathbb{R}^d}$, $1 \leq i \leq n$, we have

$$\mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n) = \int_{A_1 \times \cdots \times A_n} p_{t_1}(x_1 - x) p_{t_2 - t_1}(x_2 - x_1) \cdots p_{t_n - t_{n-1}}(x_n - x_{n-1}) \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

Remark 3.0.5. Reduction to centered case.

From the definition, we get that if $B = (B_t)_{t \geq 0}$ is a Brownian motion issued from the origin namely $B_0 = 0$ and if H is a random variable then $(H + B_t)_{t \geq 0}$ is also a Brownian motion, issued from H.

Remark 3.0.6. Reduction to one-dimensional case.

From the definition, if $X = (X_t)_{t \geq 0}$ is d-dimensional with coordinates $X_t = (X_t^1, \ldots, X_t^d)$ in $\mathbb{R}^d$, then X is a Brownian motion issued from the origin iff the following two properties hold true:

1. for all $1 \leq i \leq d$, $(X_t^i)_{t \geq 0}$ is a Brownian motion issued from the origin
2. the processes $(X_t^1)_{t \geq 0}, \ldots, (X_t^d)_{t \geq 0}$ are independent.

3.1 Characterizations and martingales

Theorem 3.1.1. Characterization of BM by Gaussianity and covariance.

If $X = (X_t)_{t \geq 0}$ is real, continuous, issued from the origin, then X is a Brownian motion if and only if X is a Gaussian process, centered, with covariance given by $\mathbb{E}(X_t X_s) = s \wedge t$ for all $s, t \geq 0$.

Proof.

1. Suppose that $X = (X_t)_{t \geq 0}$ is a Brownian motion issued from the origin, then for all $0 < t_1 < \cdots < t_n$ the random variables $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are Gaussian, centered, and independent, and $X_0 = 0$,

and $(X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}})$ and $(X_{t_1}, \ldots, X_{t_n})$ are (centered) Gaussian random vectors in the sense that all linear combinations of their coordinates are Gaussian. Moreover, for all $0 \le s \le t$, we have

$$\mathbb{E}(X_s X_t) = \mathbb{E}(X_s(X_t - X_s)) + \mathbb{E}(X_s^2) = 0 + s = s = s \wedge t.$$

2. Conversely, if $X = (X_t)_{t\ge 0}$ is a Gaussian process, centered, with $\mathbb{E}(X_s X_t) = s \wedge t$ for all $s, t \ge 0$, then for all $0 < t_1 < \cdots < t_n$, the random vector $(X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}})$ is Gaussian, centered, with diagonal covariance $\operatorname{diag}(t_1, t_2 - t_1, \ldots, t_n - t_{n-1})$, which implies that $(X_t)_{t\ge 0}$ is a Brownian motion.

■

Corollary 3.1.2. Scale invariance by space-time scaling.

If $B = (B_t)_{t\ge 0}$ is a BM on $\mathbb{R}$, issued form the origin, then for all $c \in (0, +\infty)$, $\left(\frac{1}{\sqrt{c}} B_{ct}\right)_{t\ge 0}$ is a BM.

Proof. The process $\left(\frac{1}{\sqrt{c}} B_{ct}\right)_{t\ge 0}$ is continuous, Gaussian, centered, with same covariance as BM. ■

Theorem 3.1.3. Fourier and Laplace martingale characterizations of Brownian motion.

Let $X = (X_t)_{t\ge 0}$ be a d-dimensional continuous process issued from the origin.
The following properties are equivalent:

1. X is a Brownian motion
2. For all $\lambda \in \mathbb{R}^d$, $(M_t^\lambda)_{t\ge 0} = (\mathrm{e}^{\mathrm{i}\lambda\cdot X_t + \frac{|\lambda|^2 t}{2}})_{t\ge 0}$ is a martingale[a] for the natural filtration of X
3. For all $\lambda \in \mathbb{R}^d$, $(N_t^\lambda)_{t\ge 0} = (\mathrm{e}^{\lambda\cdot X_t - \frac{|\lambda|^2 t}{2}})_{t\ge 0}$ is a martingale for the natural filtration of X.

[a] The notion of martingale remains valid for complex valued processes.

Proof. Let us define $\mathscr{G}_t = \sigma(X_s : s \in [0, t])$ for all $t \ge 0$. The process X is a BM iff for all $0 \le s < t$, $X_t - X_s$ is independent of $\mathscr{G}_s$ and $X_t - X_s \sim \mathcal{N}(0, (t - s)I_d)$, in other words if and only if for all $0 \le s < t$ and $\lambda \in \mathbb{R}^d$,

$$\mathbb{E}(\mathrm{e}^{\mathrm{i}\lambda\cdot(X_t - X_s)} \mid \mathscr{G}_s) = \mathrm{e}^{-\frac{|\lambda|^2(t-s)}{2}}.$$

By multiplying both sides by $\mathrm{e}^{\mathrm{i}uZ}$ for an arbitrary bounded $\mathscr{G}_s$ measurable random variable Z and taking the expectation we get that $X_t - X_s$ is independent of $\mathscr{G}_s$ and $X_t - X_s \sim \mathcal{N}(0, (t-s)I_d)$. This shows the equivalence of the first two properties. The third property is the Laplace (instead of Fourier) transform version. ■

Definition 3.1.4. Brownian motion with respect to a filtration.

Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\ge 0}, \mathbb{P})$ be a filtered probability space. We say that a continuous d-dimensional process $X = (X_t)_{t\ge 0}$ is an $(\mathscr{F}_t)_{t\ge 0}$ Brownian motion when it is $(\mathscr{F}_t)_{t\ge 0}$ adapted and for all $t \ge 0$ and $s \in [0, t]$, the increment $X_t - X_s$ is independent of $\mathscr{F}_s$ and follows the Gaussian law $\mathcal{N}(0, (t - s)I_d)$, which is equivalent to say that for all $\lambda \in \mathbb{R}^d$, the process $(\exp(\mathrm{i}\lambda \cdot X_t + \frac{1}{2}|\lambda|^2 t))_{t\ge 0}$ is an $(\mathscr{F}_t)_{t\ge 0}$-martingale.

Remark 3.1.5. Definitions of Brownian motion (BM).

If $X = (X_t)_{t\ge 0}$ is an $(\mathscr{F}_t)_{t\ge 0}$ BM, then X is a BM in the sense of Definition 3.0.1. Conversely, a BM $(X_t)_{t\ge 0}$ in the sense of Definition 3.0.1 is an $(\mathscr{G}_t)_{t\ge 0}$ BM where $\mathscr{G}_t = \sigma(X_s : s \le t)$ for all $t \ge 0$ is the natural filtration associated to X (see Theorem 3.1.3).

Theorem 3.1.6. Martingale properties.

Let $B=(B_t)_{t\geq 0}$ be an $(\mathscr{F}_t)_{t\geq 0}$ d-dimensional Brownian motion and let $B_t=(B_t^1,\dots,B_t^d)$ be the coordinates of the random vector B_t. Then for all $0\leq s<t$ an $1\leq j,k\leq d$,

$$\mathbb{E}(B_t^j-B_s^j\mid\mathscr{F}_s)=0\quad\text{and}\quad\mathbb{E}((B_t^j-B_s^j)(B_t^k-B_s^k)\mid\mathscr{F}_s)=(t-s)\mathbf{1}_{j=k}.$$

As a consequence, for all $1\leq j,k\leq d$,

- $(B_t^j)_{t\geq 0}$ is a continuous $(\mathscr{F}_t)_{t\geq 0}$-martingale, provided that $B_0\in\mathrm{L}^1$
- $(B_t^jB_t^k-\mathbf{1}_{j=k}t)_{t\geq 0}$ is a continuous $(\mathscr{F}_t)_{t\geq 0}$-martingale, provided that $B_0\in\mathrm{L}^2$.

Actually it turns out that these properties characterize Brownian motion.

Proof. The first property follows from the fact that $(B_t^j)_{t\geq 0}$ is a BM. For the second property, we write

$$\begin{aligned}\mathbb{E}((B_t^j-B_s^j)(B_t^k-B_s^k)\mid\mathscr{F}_s)&=\mathbb{E}((B_t^j-B_s^j)(B_t^k-B_s^k))\\&=\mathbb{E}((B_t^j-B_s^j))\mathbb{E}((B_t^k-B_s^k))\mathbf{1}_{j\neq k}+\mathbb{E}((B_t^j-B_s^k)^2)\mathbf{1}_{j=k}\\&=0+(t-s)\mathbf{1}_{j=k}.\end{aligned}$$

As a consequence, for all $0\leq s\leq t$ and $1\leq j,k\leq d$,

$$\mathbb{E}(B_t^j\mid\mathscr{F}_s)=B_s^j=\mathbb{E}(B_s^j\mid\mathscr{F}_s)$$

and

$$\mathbb{E}(B_t^jB_t^k-t\mathbf{1}_{j=k}\mid\mathscr{F}_s)=B_s^jB_s^k-s\mathbf{1}_{j=k}=\mathbb{E}(B_s^jB_s^k-s\mathbf{1}_{j=k}\mid\mathscr{F}_s).$$

■

Up to now, we study BM but it is unclear if BM exists or not! Actually an explicit construction of BM is given in Section 3.6. Other constructions are available, see for instance [13].

3.2 Variation of trajectories and quadratic variation

See Definition 1.7.1 (finite variation functions) and Definition 4.1.1 (quadratic variation of processes).

Theorem 3.2.1. Variation and quadratic variation of Brownian motion.

Let $B=(B_t)_{t\geq 0}$ be a BM issued from the origin, let $[u,v]$ be a finite interval, $0\leq u<v$, and let δ be a partition or sub-division of $[u,v]$, $\delta: u=t_0<\cdots<t_n=v$, $n\geq 1$. Let us consider the quantities

$$r_1(\delta)=\sum_{i=1}^{n-1}|B_{t_{i+1}}-B_{t_i}|\quad\text{and}\quad r_2(\delta)=\sum_{i=0}^{n-1}|B_{t_{i+1}}-B_{t_i}|^2.$$

Then the following properties hold true:

1. $\lim_{|\delta|\to 0}r_2(\delta)=v-u$ in L^2 and thus in $\mathbb{P}$, where $|\delta|=\sup_{0\leq i\leq n}(t_{i+1}-t_i)$. In other words, the quadratic variation of B on a finite interval is equal to the length of the interval.
2. $\sup_{\delta\in\mathscr{P}}r_1(\delta)=+\infty$ almost surely, where $\mathscr{P}$ is the set of subdivision of $[u,v]$. In other words the sample paths of B are almost surely of infinite variation on all intervals.

The second proprery implies that we cannot hope to define an integral $\int_a^b\varphi_t\mathrm{d}B_t(\omega)$ with φ continuous as in Theorem 1.7.2 because $t\mapsto B_t(\omega)$ is of infinite variation on all intervals for almost all ω. However, and following Itô, the first property will be the key to give a sort of L^2 or in $\mathbb{P}$ meaning to such stochastic integrals.

The quadratic variation of square integrable continuous martingales is considered in Theorem 4.1.4.

Proof. We could use Lemma 4.1.2 to get that the sample path of B have infinite variation on the time interval $[0,t]$. Let us be more precise by using the special explicit nature of Brownian motion.

1. If $Z \sim \mathcal{N}(0,1)$ then $\mathbb{E}(Z^4) = 3$, hence

$$\begin{aligned}\mathbb{E}((r_2(\delta))^2) &= \mathbb{E}\Big(\Big(\sum_i |B_{t_{i+1}} - B_{t_i}|^2\Big)^2\Big)\\ &= \sum_i \mathbb{E}(|B_{t_{i+1}} - B_{t_i}|^4) + 2\sum_{i<j} \mathbb{E}(|B_{t_{i+1}} - B_{t_i}|^2 |B_{t_{j+1}} - B_{t_j}|^2)\\ &= 3\sum_i (t_{i+1}-t_i)^2 + 2\sum_{i<j}(t_{i+1}-t_i)(t_{j+1}-t_j)\\ &= 2\sum_i (t_{i+1}-t_i)^2 + \Big(\sum_i (t_{i+1}-t_i)\Big)^2\\ &= 2\sum_i (t_{i+1}-t_i)^2 + (v-u)^2.\end{aligned}$$

Moreover $\mathbb{E}(r_2(\delta)) = \sum_i (t_{i+1}-t_i) = v-u$. Thus

$$\mathbb{E}((r_2(\delta)-(v-u))^2) = 2\sum_i (t_{i+1}-t_i)^2 \le 2\max_i (t_{i+1}-t_i)(v-u) \underset{|\delta|\to 0}{\longrightarrow} 0.$$

2. From the first part, there exists a sequence of subdivisions $(\delta^k)_k$ of $[u,v]$ such that

$$\lim_{k\to\infty} r_2(\delta^k) = \lim_{k\to\infty} \sum_i |B_{t^k_{i+1}} - B_{t^k_i}|^2 = v-u \quad \text{almost surely}$$

and thus, almost surely,

$$\sup_\delta r_1(\delta) \ge r_1(\delta^k) = \sum_i |B_{t^k_{i+1}} - B_{t^k_i}| \ge \frac{\sum_i |B_{t^k_{i+1}} - B_{t^k_i}|^2}{\max_i |B_{t^k_{i+1}} - B_{t^k_i}|} \underset{k\to\infty}{\longrightarrow} +\infty,$$

where used the fact that almost surely, $\max_i |B_{t^k_{i+1}} - B_{t^k_i}| \to 0$ as $k \to \infty$ since $B_\bullet$ is continuous and hence uniformly continuous on every compact interval such as $[u,v]$ (Heine theorem).

■

3.3 Blumenthal zero-one law and its consequences on the trajectories

This can be skipped at first reading.

Theorem 3.3.1: Properties of Brownian trajectories

If $B = (B_t)_{t\ge 0}$ is a one-dimensional BM on $\mathbb{R}$ issued form the origin, and $\mathcal{F}_t = \sigma(B_t)$, then:

1. Blumenthal[a] 0-1 law. The σ-algebra $\mathcal{F}_{0^+} = \cap_{t>0}\mathcal{F}_t$ is trivial: for all $A \in \mathcal{F}_{0^+}$, $\mathbb{P}(A) \in \{0,1\}$
2. Almost surely, for all $\varepsilon > 0$, $\inf_{s\in[0,\varepsilon]} B_s < 0$ and $\sup_{s\in[0,\varepsilon]} B_s > 0$
3. For all $a \in \mathbb{R}$, almost surely[b], $T_a = \inf\{t \ge 0 : B_t = a\} < \infty$
4. Almost surely[c], $\underline{\lim}_{t\to\infty} B_t = -\infty$ and $\overline{\lim}_{t\to\infty} B_t = +\infty$
5. Almost surely, the function $t \in \mathbb{R}_+ \to B_t$ is not monotone on any non singleton interval.

[a] Named after Robert McCallum Blumenthal (1931 – 2012), American mathematician.
[b] Howerver T_a is not bounded, see Remark 2.5.4.
[c] This does not imply that a.s. $\lim_{t\to\infty} |B_t| = +\infty$.

Proof.

1. The idea is to show that $\mathscr{F}_{0^+}$ is independent of itself. For all $A \in \mathscr{F}_{0^+}$, all $k \geq 1$, all bounded continuous $f : \mathbb{R}^k \to \mathbb{R}$, and all $0 < t_1 < \cdots < t_k$, we have

$$\mathbb{E}(\mathbf{1}_A f(B_{t_1}, \ldots, B_{t_k})) = \lim_{\varepsilon \to 0^+} \mathbb{E}(\mathbf{1}_A f(B_{t_1} - B_\varepsilon, \ldots, B_{t_k} - B_\varepsilon)).$$

Now when $0 < \varepsilon < t_1$, the random variables $B_{t_1} - B_\varepsilon, \ldots, B_{t_k} - B_\varepsilon$ are independent of $\mathscr{F}_\varepsilon$ (structure of the increments of simple Markov property), and thus independent of $\mathscr{F}_{0^+}$. It follows that

$$\mathbb{E}(\mathbf{1}_A f(B_{t_1}, \ldots, B_{t_k})) = \lim_{\varepsilon \to 0^+} \mathbb{P}(A)\mathbb{E}(f(B_{t_1} - B_\varepsilon, \ldots, B_{t_k} - B_\varepsilon)) = \mathbb{P}(A)\mathbb{E}(f(B_{t_1}, \ldots, B_{t_k})).$$

Hence $\mathscr{F}_{0^+}$ is independent of $\sigma(B_{t_1}, \ldots, B_{t_k})$ for all t_i's, and thus is independent of $\sigma(B_t, t > 0)$. But $\sigma(B_t, t > 0) = \sigma(B_t, t \geq 0)$ since $B_0 = 0$. It remains to note that $\mathscr{F}_{0^+} \subset \sigma(B_t, t \geq 0)$.

2. For the statement with the sup, it suffices to show that $\mathbb{P}(A) = 1$ where

$$A = \bigcap_n \Big\{ \sup_{s \in [0,1/n]} B_s > 0 \Big\}.$$

We can restrict the intersection to $n \geq N$ for an arbitrary large threshold N, therefore $A \in \mathscr{F}_{0^+}$. Next, thanks to the Blumenthal zero-one law, it suffices to show that $\mathbb{P}(A) > 0$. Now

$$\mathbb{P}\Big(\sup_{s \in [0,1/n]} B_s > 0 \Big) \underset{n \to \infty}{\searrow} \mathbb{P}(A)$$

while

$$\mathbb{P}\Big(\sup_{s \in [0,1/n]} B_s > 0 \Big) \geq \mathbb{P}(B_{1/n} > 0) = \frac{1}{2},$$

giving $\mathbb{P}(A) \geq 1/2$ and thus $\mathbb{P}(A) = 1$. The statement with inf follows by using $-B$ instead of B.

3. Thanks to the previous property,

$$\mathbb{P}\Big(\sup_{s \in [0,1]} B_s > \varepsilon \Big) \underset{\varepsilon \to 0}{\nearrow} \mathbb{P}\Big(\sup_{s \in [0,1]} B_s > 0 \Big) = 1.$$

But by the scale invariance (Corollary 3.1.2),

$$\mathbb{P}\Big(\sup_{s \in [0,1]} B_s > \varepsilon \Big) = \mathbb{P}\Big(\sup_{s \in [0,\varepsilon^{-2}]} \varepsilon^{-1} B_{\varepsilon^2 s} > 1 \Big) = \mathbb{P}\Big(\sup_{s \in [0,\varepsilon^{-2}]} B_s > 1 \Big)$$

Now, since

$$\mathbb{P}\Big(\sup_{s \in [0,\varepsilon^{-2}]} B_s > 1 \Big) \underset{\varepsilon \to 0}{\nearrow} \mathbb{P}\Big(\sup_{s \geq 0} B_s > 1 \Big),$$

we get

$$\mathbb{P}\Big(\sup_{s \geq 0} B_s > 1 \Big) = 1.$$

Again by scaling, we obtain, for all $R > 0$, and also by replacing B by $-B$,

$$\mathbb{P}\Big(\sup_{s \geq 0} B_s > R \Big) = 1 \quad \text{and} \quad \mathbb{P}\Big(\inf_{s \geq 0} B_s < -R \Big) = 1.$$

This implies that for all $a \in \mathbb{R}$, almost surely $T_a < \infty$.

4. This is implied directly by the end of the proof of the previous item.

5. From the item about the inf and sup, and the structure of increments, we have, almost surely, for all $t \in \mathbb{Q} \cap \mathbb{R}_+$ and all $\varepsilon > 0$, $\inf_{s \in [t,t+\varepsilon]} B_s < B_t$ and $\sup_{s \in [t,t+\varepsilon]} B_s > B_t$, hence the result.

■

3.4 Strong law of large numbers, invariance by time inversion, law of iterated logarithm

The nature of the increments of Brownian motion leads to formulate the following theorem.

Theorem 3.4.1. Strong law of large numbers.

If $(B_t)_{t\geq 0}$ is BM on $\mathbb{R}$ with $B_0=0$ then $\displaystyle\lim_{t\to\infty}\frac{B_t}{t}=0$ almost surely and in L^p for all $p\in[1,\infty)$.

The central limit theorem would be the trivial statement $\sqrt{t}\frac{B_t}{t}\xrightarrow[t\to\infty]{\text{law}}\mathcal{N}(0,1)$.
The a.s. remains valid for an arbitrary B_0, and the L^p convergence if $B_0\in\mathrm{L}^p$.

Proof. Since for all $t>0$ and all $p>0$, $\mathbb{E}\left(\left|\frac{B_t}{t}\right|^p\right)=\frac{\mathbb{E}(|B_1|^p)}{t^{p/2}}$ and $B_1\sim\mathcal{N}(0,1)$, we have immediately

$$\frac{B_t}{t}\xrightarrow[t\to\infty]{\mathrm{L}^p}0\quad\text{and in particular}\quad\frac{B_t}{t}\xrightarrow[t\to\infty]{\mathbb{P}}0.$$

To get the almost sure convergence, we need some tightness, a control of tails that can be done via moments. Let us prove the a.s. convergence. Let a and b be real numbers such that $0<a<b$. We have

$$\mathbb{E}\left(\sup_{a\leq t\leq b}\left(\frac{B_t}{t}\right)^2\right)\leq\frac{1}{a^2}\mathbb{E}\left(\sup_{a\leq t\leq b}B_t^2\right).$$

The Doob maximal inequality of Theorem 2.5.7 applied to the martingale $(B_{a+t})_{t>0}$ on $[0,b-a]$ yields

$$\mathbb{E}\left(\sup_{a\leq t\leq b}\left(\frac{B_t}{t}\right)^2\right)\leq\frac{1}{a^2}4\mathbb{E}(B_b^2)=\frac{4b}{a^2}.$$

Applying this to $a=2^n$ and $b=2^{n+1}$ we obtain

$$\mathbb{E}\left(\sup_{2^n\leq t\leq 2^{n+1}}\left(\frac{B_t}{t}\right)^2\right)\leq\frac{8}{2^n}$$

Thus, by the Markov inequality, for any $\varepsilon>0$,

$$\mathbb{P}\left(\sup_{2^n\leq t\leq 2^{n+1}}\left|\frac{B_t}{t}\right|>\varepsilon\right)\leq\frac{1}{\varepsilon^2}\mathbb{E}\left(\sup_{2^n\leq t\leq 2^{n+1}}\left(\frac{B_t}{t}\right)^2\right)\leq\frac{8}{2^n\varepsilon^2},$$

which gives

$$\sum_{n=0}^{\infty}\mathbb{P}\left(\sup_{2^n\leq t\leq 2^{n+1}}\left|\frac{B_t}{t}\right|>\varepsilon\right)<\infty.$$

Now, according to the Borel–Cantelli lemma, there exists an almost sure event A_ε such that for all $\omega\in A_\varepsilon$, there exists a threshold n_ω such that for all $n\geq n_\omega$, $\sup_{2^n\leq t\leq 2^{n+1}}\left|\frac{B_t(\omega)}{t}\right|\leq\varepsilon$. Thus, for all $\varepsilon>0$, there exists an a.s. event A_ε such that for all $\omega\in A_\varepsilon$, there exists t_ω, such that for all $t\geq t_\omega$,

$$\left|\frac{B_t(\omega)}{t}\right|\leq\varepsilon.$$

It remains to consider the almost sure event $A=\bigcap_{r=1}^{\infty}A_{1/r}$, on which $\lim_{t\to\infty}\frac{B_t}{t}=0$. ∎

Corollary 3.4.2. Invariance by time inversion.

If $B=(B_t)_{t\geq 0}$ is a BM on $\mathbb{R}$ with $B_0=0$ then $X=(tB_{1/t})_{t\geq 0}$ with the convention $X_0=0$ is also BM.

Proof. The process X is Gaussian, centered, with $\mathbb{E}(X_sX_t)=s\wedge t$ for all $s,t\geq 0$. It remains to prove that X is continuous. By definition X is continuous on $(0,\infty)$. It remains to prove the almost sure continuity at $t=0$. This follows from Theorem 3.4.1, namely, almost surely, $\lim_{t\to 0^+}X_t=\lim_{t\to 0^+}tB_{1/t}=\lim_{t\to+\infty}\frac{B_t}{t}=0$. ∎

Theorem 3.4.3. Law of Iterated Logarithm.

If $(B_t)_{t\geq 0}$ is a Brownian motion on $\mathbb{R}$ then

$$\mathbb{P}\Big(\underline{\lim}_{t\searrow 0}\frac{B_t}{\sqrt{2t\log(\log(1/t))}}=-1,\quad \overline{\lim}_{t\searrow 0}\frac{B_t}{\sqrt{2t\log(\log(1/t))}}=1\Big)=1$$

and

$$\mathbb{P}\Big(\underline{\lim}_{t\to\infty}\frac{B_t}{\sqrt{2t\log(\log(t))}}=-1,\quad \overline{\lim}_{t\to\infty}\frac{B_t}{\sqrt{2t\log(\log(t))}}=1\Big)=1.$$

This can be skipped at first reading.

Proof. The second property follows from the first one by using invariance by time inversion (Corollary 3.4.2). Let us prove the first property. We can assume without loss of generality that $B_0=0$. Since the intersection of two almost sure events is almost sure, and since the law of B is symmetric in the sense that $-B$ and B have same law, it follows that it suffices to show that

$$\mathbb{P}\Big(\overline{\lim}_{t\searrow 0}\frac{B_t}{\sqrt{2t\log(\log(1/t))}}=1\Big)=1.$$

Let us first prove that

$$\mathbb{P}\Big(\overline{\lim}_{t\searrow 0}\frac{B_t}{\sqrt{2t\log(\log(1/t))}}\leq 1\Big)=1. \tag{$\star$}$$

Let us define $h(t)=\sqrt{2t\log(\log(1/t))}$. For all $\alpha>0$ and $\beta>0$, the Doob maximal inequality of Theorem 2.5.7 used for the "exponential" martingale $(\mathrm{e}^{\alpha B_t-\frac{\alpha^2}{2}t})_{t\geq 0}$ gives, for all $t\geq 0$,

$$\mathbb{P}\Big(\max_{s\in[0,t]}\Big(B_s-\frac{\alpha}{2}s\Big)>\beta\Big)=\mathbb{P}\Big(\max_{s\in[0,t]}\mathrm{e}^{\alpha B_s-\frac{\alpha^2}{2}s}\geq \mathrm{e}^{\alpha\beta}\Big)\leq \mathrm{e}^{-\alpha\beta}.$$

For all $\theta,\delta\in(0,1)$ and $n\geq 1$, this inequality with $t=\theta^n$, $\alpha=(1+\delta)h(\theta^n)/\theta^n$ and $\beta=h(\theta^n)/2$ gives

$$\mathbb{P}\Big(\max_{s\in[0,\theta^n]}\Big(B_s-\frac{(1+\delta)h(\theta^n)}{2\theta^n}s\Big)>\frac{h(\theta^n)}{2}\Big)=O_{n\to\infty}(n^{-(1+\delta)}).$$

By the Borel–Cantelli lemma, we get that for almost all $\omega\in\Omega$, there exists n_ω such that for all $n\geq n_\omega$,

$$\max_{s\in[0,\theta^n]}\Big(B_s-\frac{(1+\delta)h(\theta^n)}{2\theta^n}s\Big)\leq\frac{1}{2}h(\theta^n).$$

This inequality implies that for all $t\in[\theta^{n+1},\theta^n]$,

$$B_t(\omega)\leq\max_{s\in[0,\theta^n]}B_s(\omega)\leq\frac{1}{2}(2+\delta)h(\theta^n)\leq\frac{(2+\delta)h(t)}{2\sqrt{\theta}}.$$

Therefore

$$\mathbb{P}\Big(\overline{\lim}_{t\searrow 0}\frac{B_t}{\sqrt{2t\log(\log(1/t))}}\leq\frac{2+\delta}{2\sqrt{\theta}}\Big)=1.$$

Now we let $\theta\to 1$ and $\delta\to 0$ to get ($\star$). It remains to prove that

$$\mathbb{P}\Big(\overline{\lim}_{t\searrow 0}\frac{B_t}{\sqrt{2t\log(\log(1/t))}}\geq 1\Big)=1.$$

For that, for all $n\geq 1$ and $\theta\in(0,1)$, we define the event

$$A_n=\{\omega\in\Omega:B_{\theta^n}(\omega)-B_{\theta^{n+1}}(\omega)\geq(1-\sqrt{\theta})h(\theta^n)\}.$$

We have, denoting $a_n = (1-\sqrt{\theta})h(\theta^n)/(\theta^{n/2}\sqrt{1-\theta})$,

$$\mathbb{P}(A_n) = \frac{1}{\sqrt{2\pi}}\int_{a_n}^{\infty} e^{-\frac{u^2}{2}}\,du \geq \frac{a_n}{1+a_n^2}e^{-\frac{a_n^2}{2}} = O_{n\to\infty}\Big(n^{-\frac{1+\theta-2\sqrt{\theta}}{1-\theta}}\Big).$$

Thus $\sum_{n\geq 1}\mathbb{P}(A_n) = +\infty$. Now the independence of the increments of B and the Borel–Cantelli lemma give that almost surely, for an infinite number of values of n, we have

$$B_{\theta^n} - B_{\theta^{n+1}} \geq (1-\sqrt{\theta})h(\theta^n).$$

But the first part of the proof gives, for almost all $\omega\in\Omega$, that there exists n_ω such that for all $n\geq n_\omega$,

$$B_{\theta^{n+1}} > -2h(\theta^{n+1}) \geq -2\sqrt{\theta}h(\theta^n).$$

Therefore, almost surely, for an infinite number of values of n, we have

$$B_{\theta^n} > h(\theta^n)(1-3\sqrt{\theta}).$$

This gives

$$\mathbb{P}\Big(\varlimsup_{t\searrow 0}\frac{B_t}{\sqrt{2t\log(\log(1/t))}} \geq 1-3\sqrt{\theta}\Big) = 1.$$

It remains to send θ to 0. Note that this proof uses both sides of the Borel–Cantelli lemma. ∎

Corollary 3.4.4. Regularity of Brownian motion sample paths.

If $(B_t)_{t\geq 0}$ is a Brownian motion on $\mathbb{R}$ then for all $s\geq 0$, we have

$$\mathbb{P}\Big(\varliminf_{t\searrow 0}\frac{B_{t+s}-B_s}{\sqrt{2t\log(\log(1/t))}} = -1,\quad \varlimsup_{t\searrow 0}\frac{B_{t+s}-B_s}{\sqrt{2t\log(\log(1/t))}} = 1\Big) = 1.$$

In particular almost surely the sample paths $t\in\mathbb{R}_+\mapsto B_t$ of B are not $\frac{1}{2}$-Hölder[a] continuous on finite intervals and in particular are nowhere differentiable on $\mathbb{R}_+$.

[a] $f : I\to\mathbb{R}$ is γ-Hölder continuous when $(\forall\varepsilon>0)(\exists\eta>0)(\forall s,t\in I)(|s-t|^\gamma\leq\eta\Rightarrow|f(s)-f(t)|\leq\varepsilon)$.

Proof. Follows from Theorem 3.4.3 and the fact that $(B_{t+s}-B_s)_{t\geq 0}$ and $(B_t)_{t\geq 0}$ have same law. ∎

3.5 Strong Markov property, reflection principle, hitting time

If $(B_t)_{t\geq 0}$ is BM then we easily check that for all fixed $T>0$, the process $(B_{t+T}-B_T)_{t\geq 0}$ is a BM, issued form the origin, independent of $\mathcal{F}_T$. This is the simple Markov property. It extends to stopping times T:

Theorem 3.5.1. Strong Markov[a] property.

[a] Named after Andrey Markov (1856 – 1922), Russian mathematician.

If $B=(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion issued from the origin, then for all stopping time T such that $\mathbb{P}(T<\infty)>0$, under the probability measure $\mathbb{P}(\cdot\mid T<\infty)$, the following properties hold:

1. $((B_{t+T}-B_T)\mathbf{1}_{\{T<\infty\}})_{t\geq 0}$ is a Brownian motion issued from the origin, independent of $\mathcal{F}_T$
2. For all measurable and bounded $f:\mathbb{R}^d\to\mathbb{R}$, we have, for all $t>0$,

$$\mathbb{E}(f(B_{t+T})\mathbf{1}_{\{T<\infty\}}\mid\mathcal{F}_T) = P_t(f)(B_T)\mathbf{1}_{\{T<\infty\}}$$

where

$$P_t(f)(x) = \mathbb{E}(f(x+B_t)) = \frac{1}{(\sqrt{2\pi t})^d}\int_{\mathbb{R}^d} e^{-\frac{|x-y|^2}{2t}}f(y)\mathrm{d}y = (p_t * f)(x).$$

We say then that Brownian motion is a strong Markov process.

Proof. Suppose first that $\mathbb{P}(T<\infty)=1$. Let us define $B^* = (B_{T+t}-B_T)_{t\geq 0}$. For all $n\geq 1$, let us define

$$T_n = \sum_{k\geq 0} \frac{k+1}{2^n} \mathbf{1}_{T\in[\frac{k}{2^n},\frac{k+1}{2^n})}.$$

We have that $T\leq T_n$, and T_n takes its values in the set of dyadics $D_n = \{k/2^n : k\geq 0\}$. We check easily that T_n is a stopping time, and that $T_n \searrow T$ as $n\to\infty$. Let $A\in\mathscr{F}_T$, $m\geq 0$, and $0=t_0<\cdots<t_m<\infty$. By the dominated convergence theorem, we have, for all continuous and bounded $\varphi : (\mathbb{R}^d)^m \to \mathbb{R}$,

$$\begin{aligned}\mathbb{E}(\mathbf{1}_A\varphi(B^*_{t_1},\ldots,B^*_{t_m})) &= \mathbb{E}(\mathbf{1}_A\varphi(B_{t_1+T}-B_T,\ldots,B_{t_m+T}-B_T))\\ &= \lim_{n\to\infty}\mathbb{E}(\mathbf{1}_A\varphi(B_{t_1+T_n}-B_{T_n},\ldots,B_{t_m+T_n}-B_{T_n})).\end{aligned}$$

Moreover, for all $n\geq 1$, we have $A\in\mathscr{F}_T\subset\mathscr{F}_{T_n}$ since $T\leq T_n$ and, using the fact that $A\in\mathscr{F}_{T_n}$,

$$\begin{aligned}\mathbb{E}(\mathbf{1}_A\varphi(B_{t_1+T_n}-B_{T_n},\ldots,B_{t_m+T_n})) &= \sum_{r\in D_n}\mathbb{E}(\mathbf{1}_{A\cap\{T_n=r\}}\varphi(B_{t_1+r}-B_r,\ldots,B_{t_m+r}-B_r))\\ &= \sum_{r\in D_n}\mathbb{E}(\mathbf{1}_{A\cap\{T_n=r\}}\mathbb{E}(\varphi(B_{t_1+r}-B_r,\ldots,B_{t_m+r}-B_r)\mid\mathscr{F}_r))\\ &= \sum_{r\in D_n}\mathbb{P}(A\cap\{T_n=r\})\mathbb{E}(\varphi(B_{t_1+r}-B_r,\ldots,B_{t_m+r}-B_r))\\ &= \mathbb{P}(A)\mathbb{E}(\varphi(B_{t_1}-B_0,\ldots,B_{t_m}-B_0)).\end{aligned}$$

This implies the first property since $(B_t-B_0)_{t\geq 0}$ is a Brownian motion issued from the origin. Note that this proves in the same time the fact that B^* has the law of B and is independent of $\mathscr{F}_T$. To prove only the identity in law, we can remove $\mathbf{1}_A$ in other words take $A=\Omega$.

The second property follows immediately from the first one, namely since for all $t\geq 0$, B^*_t is independent of $\mathscr{F}_T$ while B_T is measurable with respect to $\mathscr{F}_T$ we get, using Remark 1.5.2,

$$\mathbb{E}(f(B_{t+T})\mid\mathscr{F}_T) = \mathbb{E}(f(B^*_t+B_T)\mid\mathscr{F}_T) = g_t(B_T)$$

where

$$g_t(x) = \mathbb{E}(f(x+B^*_t)) = \mathbb{E}(f(x+B_t)) = (p_t * f)(x).$$

Finally, for a T taking values in $[0,+\infty]$, the same argument works with A replaced by $A\cap\{T<\infty\}$. ■

This can be skipped at first reading.

Corollary 3.5.2: Reflection principle

Let B be a one-dimensional Brownian motion issued from the origin. For all $t\geq 0$, let us define $S_t = \sup_{s\in[0,t]} B_s$. Then, for all $t\geq 0$, the following properties hold:

- For all $a\geq 0$ and all $b\in(-\infty,a]$, $\mathbb{P}(S_t\geq a, B_t\leq b) = \mathbb{P}(B_t\geq 2a-b)$.
- The random variables S_t and $|B_t|$ have same law.

The reflection principle simply says that on the event $\{T_a\leq t\}$, the probability of being, at time t, below level $b=a-(a-b)$, is equal to the one of being above level $a+(a-b)$, hence the name. This is related to the fact that the process after time T_a is again BM, which has a symmetric law.

Proof.

- We know from Theorem 3.3.1 that $T_a = \inf\{t\geq 0 : B_t = a\}<\infty$ almost surely. We have

$$\mathbb{P}(S_t\geq a, B_t\leq b) = \mathbb{P}(T_a\leq t, B_t\leq b) = \mathbb{P}(T_a\leq t, B'_{t-T_a}\leq b-a)$$

with $B'_t = B_{T_a+t}-B_{T_a}$, where we have used in the last step $B'_{t-T_a} = B_{T_a+t-T_a} - B_{T_a} = B_t - a$

which makes sense on $\{T_a \leq t\}$. Now by the strong Markov property (Theorem 3.5.1), B' is independent of T_a and has the same law as B. Since B' and $-B'$ have same law, it follows that (T_a, B') has the same law as $(T_a, -B')$. Also

$$\begin{aligned}\mathbb{P}(T_a \leq t, B'_{t-T_a} \leq b-a) &= \mathbb{P}(T_a \leq t, -B'_{t-T_a} \leq b-a)\\ &= \mathbb{P}(T_a \leq t, -(B_t - a) \leq b-a)\\ &= \mathbb{P}(T_a \leq t, B_t \geq 2a-b)\\ &= \mathbb{P}(B_t \geq 2a-b)\end{aligned}$$

where we have use in the last step the fact that $\{T_a \leq t\}$ contains a.s. $\{B_t \geq 2a-b\}$.

- Follows from the first identity with $b = a$, the inequality $B_t \leq S_t$, and the fact that B_t and $-B_t$ have same law, which give, for all $a \geq 0$,

$$\begin{aligned}\mathbb{P}(S_t \geq a) &= \mathbb{P}(S_t \geq a, B_t \leq a) + \mathbb{P}(S_t \geq a, B_t \geq a)\\ &= \mathbb{P}(B_t \geq a) + \mathbb{P}(B_t \geq a)\\ &= \mathbb{P}(B_t \geq a) + \mathbb{P}(B_t \leq -a)\\ &= \mathbb{P}(|B_t| \geq a).\end{aligned}$$

■

Corollary 3.5.3: Densities

Let $(B_t)_{t\geq 0}$ be a one-dimensional Brownian motion issued from the origin.

- For all $t > 0$, the law of the couple $(\sup_{s\in[0,t]} B_s, B_t)$ has density

$$(a,b) \in \mathbb{R}^2 \mapsto \frac{2(2a-b)}{\sqrt{2\pi t^3}} \mathrm{e}^{-\frac{(2a-b)^2}{2t}} \mathbf{1}_{a\geq 0, b\leq a}.$$

- For all $a \in \mathbb{R}$, the law of $T_a = \inf\{t \geq 0 : B_t = a\}$ is equal to the law of $\frac{a^2}{B_1^2}$ with density

$$t \in \mathbb{R} \mapsto \frac{|a|}{\sqrt{2\pi t^3}} \mathrm{e}^{-\frac{a^2}{2t}} \mathbf{1}_{t>0}.$$

The law of T_a is known as the Lévy or Bachelier distribution.
It is, up to scaling by a^2, an inverse χ^2 distribution.

Proof.

- Direct consequence of Corollary 3.5.2.
- Thanks to Corollary 3.5.2, we have, for all $t \geq 0$, denoting $S_t = \sup_{s\in[0,t]} B_s$,

$$\mathbb{P}(T_a \leq t) = \mathbb{P}(S_t \geq a) = \mathbb{P}(|B_t| \geq a) = \mathbb{P}(B_t^2 \geq a^2) = \mathbb{P}(tB_1^2 \geq a^2) = \mathbb{P}(a^2/B_1^2 \leq t).$$

■

3.6 A construction of Brownian motion

A natural and intuitive idea to construct Brownian motion is to try to realize it as a scaling limit of a random walk with Gaussian increments. More precisely, if $(X_n)_{n\geq 1}$ are independent and identically distributed real random variables with law $\mathcal{N}(0,1)$, then this would consist for all $n \geq 1$ to define the Gaussian process

$(X^n_t)_{t\geq 0}$ obtained by linear interpolation as

$$X^n_t = \frac{X_1+\cdots+X_{\lfloor nt\rfloor}}{\sqrt{n}} \sim \mathcal{N}\Big(0,\frac{\lfloor nt\rfloor}{n}\Big),$$

and to consider the limit in law of $(X^n_{t_1},\ldots,X^n_{t_k})$ as $n\to\infty$, for all $k\geq 1$ and $0\leq t_1\leq\cdots\leq t_k$. Actually X^n_t is a good approximation for numerical simulation. The central limit phenomenon suggests that the Brownian motion scaling limit is the same if we start from non Gaussian ingredients: we only need zero mean and unit variance. Such a functional central limit phenomenon is known as the Donsker invariance principle. From this point of view, Brownian motion is just a universal Gaussian limiting object.

Beyond intuition, the mathematical existence of Brownian motion is not obvious. Historically, Norbert Wiener seems to be the first scientist to give a rigorous construction, around 1923, and for this reason, Brownian motion is sometimes called the Wiener process. For more information on history, see [18, 8].

The construction of Brownian motion provided below is based on another very natural idea: by seeing Brownian motion as an infinite family of orthogonal Gaussian random variables, we could start from our favorite infinite dimensional Hilbert space, such as $\mathrm{L}^2(\mathbb{R},\mathrm{d}x)$, and construct a Gaussian random variable by using a linear combination of the elements of a Hilbert basis with Gaussian i.i.d. weights. This will produce a Gaussian process with the desired covariance. It will then remain to obtain the continuity, which can be done by using a general tightness criterion on the increments due to Kolmogorov.

Theorem 3.6.1. Pre-Brownian motion or Gaussian measures.

Let us consider the Hilbert space $\mathrm{G}=\mathrm{L}^2(\mathbb{R},\mathrm{d}x)$ and

$$\langle f,g\rangle_{\mathrm{G}} = \int f(x)g(x)\mathrm{d}x,\quad f,g\in G.$$

Then there exists a centered Gaussian family $\widetilde{B}=(\widetilde{B}_g)_{g\in\mathrm{G}}$ defined on a probability space $(\Omega,\mathcal{A},\mathbb{P})$ such that $g\in\mathrm{G}\mapsto\widetilde{B}_g\in\mathrm{L}^2(\Omega,\mathcal{A},\mathbb{P})$ is a linear isometry, in other words for all $f,g\in\mathrm{G}$ and $\alpha,\beta\in\mathbb{R}$,

$$\mathbb{E}(\widetilde{B}_f\widetilde{B}_g)=\langle f,g\rangle_{\mathrm{G}}\quad\text{and}\quad \widetilde{B}_{\alpha f+\beta g}=\alpha\widetilde{B}_f+\beta\widetilde{B}_g.$$

Proof. Let $(X_n)_{n\geq 0}$ be i.i.d. real random variables with law $\mathcal{N}(0,1)$, defined on a probability space $(\Omega,\mathcal{A},\mathbb{P})$, and let $(e_n)_{n\geq 0}$ be an orthonormal sequence of the Hilbert space $\mathrm{G}=\mathrm{L}^2(\mathbb{R},\mathrm{d}x)$. For all $g\in\mathrm{G}$, the series

$$\widetilde{B}_g=\sum_{n=0}^\infty X_n\langle g,e_n\rangle_{\mathrm{G}}$$

is well defined in $\mathrm{L}^2(\Omega,\mathcal{A},\mathbb{P})$. Indeed the Cauchy criterion is satisfied:

$$\mathbb{E}\Big(\Big(\sum_{n=p}^{p+q}X_n\langle g,e_n\rangle_{\mathrm{G}}\Big)^2\Big)=\sum_{n=p}^{p+q}\langle g,e_n\rangle^2_{\mathrm{G}}\underset{p,q\to\infty}{\longrightarrow}0.$$

We see from Lemma 3.6.2 that $\widetilde{B}$ is a centered Gaussian random variable and that

$$\|\widetilde{B}_g\|^2=\mathbb{E}((\widetilde{B}_g)^2)=\langle g,g\rangle_{\mathrm{G}}=\|g\|^2_{\mathrm{G}}$$

hence $g\mapsto\widetilde{B}_g$ is an isometry. Its linearity is immediate. By polarization we get, for all $f,g\in\mathrm{G}$,

$$4\mathbb{E}(\widetilde{B}_f\widetilde{B}_g)=\mathbb{E}((\widetilde{B}_f+\widetilde{B}_g)^2)-\mathbb{E}((\widetilde{B}_f-\widetilde{B}_g)^2)=\mathbb{E}(\widetilde{B}^2_{f+g})-\mathbb{E}(\widetilde{B}^2_{f-g})=\|f+g\|^2_{\mathrm{G}}-\|f-g\|^2_{\mathrm{G}}=\langle f,g\rangle_{\mathrm{G}}.$$

The sub-space $\mathrm{H}=\overline{\mathrm{span}\{\widetilde{B}_g:g\in\mathrm{G}\}}$ of $\mathrm{L}^2(\Omega,\mathcal{A},\mathbb{P})$ is isomorphic via $g\mapsto\widetilde{B}_g$ to $\mathrm{G}=\mathrm{L}^2(\mathbb{R},\mathrm{d}x)$.

Lemma 3.6.2. Convergence of Gaussians.

If $(X_n)_n$ is a sequence of Gaussian real random variables such that $X_n \xrightarrow[n\to\infty]{\mathrm{L}^2} X$ for some real random variable X, then $m = \lim_{n\to\infty} \mathbb{E}(X_n)$ and $\sigma^2 = \mathbb{E}(X_n^2) - \mathbb{E}(X_n)^2$ exist and $X \sim \mathcal{N}(m, \sigma^2)$.

Proof of Lemma 3.6.2. Since $X_n \to X$ in L^2, the convergence holds also in L^1, hence the convergence of the first two moments. Let us show now that X is Gaussian. Since $X_n \to X$ in law, we get, for all $t \in \mathbb{R}$, $\varphi_{X_n}(t) = \mathrm{e}^{\mathrm{i}tm_n - \frac{t^2}{2}\sigma_n^2} \to \varphi_X(t)$. Thus, for all $\varepsilon > 0$, $|\varphi_{X_n}(\varepsilon)| = \exp(-\frac{\varepsilon^2}{2}\sigma_n^2) \to |\varphi_X(\varepsilon)|$. Since φ_X is a characteristic function, it is non vanishing in a neighborhood of the origin, and thus $\sigma_n \to \sigma_*$ for some $\sigma_* \geq 0$. It follows in turn that for all $t \in \mathbb{R}$, $\mathrm{e}^{\mathrm{i}tm_n} \to \mathrm{e}^{\mathrm{i}tm_*}$ for some m_*. Now by dominated convergence,

$$\sqrt{2\pi}\mathrm{e}^{-\frac{m_n^2}{2}} = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tm_n}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tm_*}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t = \sqrt{2\pi}\mathrm{e}^{-\frac{m_*^2}{2}}$$

thus $m_n \to m_*$. Therefore $X \sim \mathcal{N}(m_*, \sigma_*^2)$, and necessarily $(m_*, \sigma_*^2) = (m, \sigma^2)$. ■

■

With $\widetilde{B}$ being as in Theorem 3.6.1, let us define, for all $t \geq 0$, the random variable

$$B_t = \widetilde{B}_{\mathbf{1}_{[0,t]}}.$$

Now $B = (B_t)_{t\geq 0}$ is a centered Gaussian process, with covariance given for all $s, t \geq 0$ by

$$\mathbb{E}(B_s B_t) = \langle \mathbf{1}_{[0,s]}, \mathbf{1}_{[0,t]} \rangle_{\mathrm{L}^2(\mathbb{R},\mathrm{d}x)} = s \wedge t.$$

However the "pre-BM" B has no reason to be continuous. Let us remark however that for all $0 \leq s < t$,

$$\frac{B_t - B_s}{\sqrt{t-s}} \sim \mathcal{N}(0,1), \quad \text{thus}^1 \quad \mathbb{E}((B_t - B_s)^{2n}) = c_n (t-s)^n.$$

The fourth moment case $n = 2$ allows, thanks to Theorem 3.6.3 below ($p = 4$, $\varepsilon = 1$, $\gamma < \varepsilon/p = 1/4$), to construct a continuous modification B^* of B, which is a Brownian motion on $\mathbb{R}$ issued from the origin. Moreover using the higher moments for all values of n gives the optimal Hölder regularity: $\gamma < \frac{n-1}{2n} \to \frac{1}{2}$.

Theorem 3.6.3. Kolmogorov continuity criterion.

Let $X = (X_t)_{t\geq 0}$ be a process defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ taking its values in a Banach space $\mathbb{B}$ with norm $\|\cdot\|$, and such that the following tightness of increments property holds: there exist $p \geq 1$, $\varepsilon > 0$, and $c > 0$ such that for all $s, t \geq 0$,

$$\mathbb{E}(\|X_t - X_s\|^p) \leq c|t-s|^{1+\varepsilon}.$$

Then there exists a modification[a] of X that is a continuous process whose trajectories are, on each finite interval, γ-Hölder continuous for all $\gamma \in [0, \varepsilon/p)$, in the sense that a.s. for all $t > 0$, there exists a constant $C = C(\omega, t) > 0$ such that for all $u, v \in [0, t]$ and all η, if $|u - v| \leq \eta$ then $|X_u - X_v| \leq C\eta^\gamma$.

[a] There exists $X^* = (X_t^*)_{t\geq 0}$ such that for all $t \geq 0$, $X_t = X_t^*$ as random variables in other words almost surely.

Proof. It suffices to prove the result on a finite time interval $[0, t]$. Let us first show that X is Hölder continuous on the dyadics $\mathcal{D} = \cup_{n\geq 0}\mathcal{D}_n$ where $\mathcal{D}_n = \{tk/2^n : k \in \{0, \ldots, 2^n\}\} \subset \mathcal{D}_{n+1}$. For notation simplicity we take $t = 1$. For all $n \geq 1$, all $\varepsilon > 0$, and all $\gamma > 0$, the Markov inequality gives

$$\mathbb{P}\left(\max_{1\leq k\leq 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-\gamma n}\right) \leq \sum_{k=1}^{2^n} \mathbb{P}\left(\|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-\gamma n}\right)$$

[1] We have $c_n = \mathbb{E}(Z^{2n}) = \frac{(2n)!}{2^n n!}$ where $Z \sim \mathcal{N}(0,1)$ but this explicit formula for c_n is useless for our purposes.

$$\le \sum_{k=1}^{2^n} 2^{\gamma pn} \mathbb{E}\left(\|X_{\frac{k}{2}^n} - X_{\frac{k-1}{2^n}}\|^p\right)$$
$$\le c2^n 2^{-n(1+\varepsilon)+\gamma pn} = c2^{-n(\varepsilon-\gamma p)}.$$

Now, by taking $\varepsilon > \gamma p$ we get

$$\sum_{n=1}^{\infty} \mathbb{P}(\max_{1\le k\le 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \ge 2^{-\gamma n}) < \infty.$$

Thus, the Borel–Cantelli lemma provides $A \in \mathscr{A}$ such that $\mathbb{P}(A) = 1$ and for all $\omega \in A$, there exists n_ω such that for all $n \ge n_\omega$, we have $\max_{1\le k\le 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \le 2^{-\gamma n}$. Hence there exists a random variable C such that

$$C < \infty \text{ a.s.} \quad \text{and} \quad \max_{1\le k\le 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \le C2^{-\gamma n}.$$

Let us prove that on A, the paths of X are γ-Hölder continuous on $\mathscr{D}$, Let $s, t \in \mathscr{D}$ with $s \ne t$ and $n \ge 0$ such that $|s-t| \le 2^{-n}$. Let $(s_k)_{k\ge1}$ be increasing, with $s_k = s$ for k large enough (stationarity), and $s_k \in \mathscr{D}_k$ for all k. Let $(t_k)_{k\ge1}$ be a similar sequence for t, such that s_k and t_k are neighbors in $\mathscr{D}_n$ for all k. Then

$$X_t - X_s = \sum_{k=n}^{\infty} (X_{t_{k+1}} - X_{t_k}) + X_{s_n} - X_{t_n} - \sum_{k=n}^{\infty} (X_{s_{k+1}} - X_{s_k}),$$

where the sums are actually finite since the sequences are stationary. Now

$$\|X_t - X_s\| \le \sum_{k=n}^{\infty} \|X_{t_k} - X_{t_{k+1}}\| + \|X_{s_n} - X_{t_n}\| + \sum_{k=n}^{\infty} \|X_{s_{k+1}} - X_{s_k}\|,$$

and thus

$$\|X_t - X_s\| \le C2^{-\gamma n} + 2\sum_{k=n}^{\infty} C2^{-\gamma(k+1)} \le 2C\sum_{k=n}^{\infty} 2^{-\gamma k} \le \frac{2C}{1-2^{-\gamma}} 2^{-\gamma n},$$

meaning that $|s-t| \le 2^{-n}$ implies $\|X_t - X_s\| \le C'2^{-\gamma n}$ for some random variable C'. Thus, on A, the sample paths of X are γ-Hölder continuous on $\mathscr{D}$. The set $\mathscr{D}$ is dense in $\mathbb{R}_+$. Now for all $\omega \in A$, let $t \mapsto X^*_t(\omega)$ be the unique continuous function[2] agreeing with $t \mapsto X_t(\omega)$ on $\mathscr{D}$.

It remains to show that X^* is a modification of X. By construction, $X_t = X^*_t$ for all $t \in \mathscr{D}$. Let $t \in \mathbb{R}_+$. Since $\mathscr{D}$ is dense in $\mathbb{R}_+$, there exists $(t_n)_n$ in $\mathscr{D}$ with $\lim_{n\to\infty} t_n = t$, thus $\lim_{n\to\infty} X_{t_n} = X_t$ in $\mathrm{L}^p((\Omega,\mathscr{A},\mathbb{P}),(\mathbb{B},\|\cdot\|))$ thanks to the hypothesis. Hence there exists a subsequence $(t_{n_k})_k$ such that $\lim_{k\to\infty} X_{t_{n_k}} = X_t$ almost surely (here we use $(\mathbb{B},\|\cdot\|)$). Finally, the continuity of X^* gives $X_{t_{n_k}} = X^*_{t_{n_k}} \to X^*_t = X_t$ almost surely as $k \to \infty$. ■

Corollary 3.6.4. Existence.

One-dimensional Brownian motion exists, and thus d-dimensional Brownian motion for all $d \ge 1$. Moreover, almost surely, the trajectories of real Brownian motion are, on each finite time interval, Hölder continuous of order γ for all $\gamma \in (0,1/2)$, not more.

Proof. Theorem 3.6.3 with $p = 2n$ and $n \to \infty$ gives $\gamma \in (0,1/2)$, while Theorem 3.4.4 gives the optimality. ■

3.7 Wiener measure, canonical Brownian motion

Let $(B_t)_{t\ge0}$ be an arbitrary d-dimensional Brownian motion issued from 0, and defined on a probability space $(\Omega,\mathscr{A},\mathbb{P})$. Since $(B_t)_{t\ge0}$ is a continuous process, we know, from Theorem 2.1.3, that we can consider $(B_t)_{t\ge0}$ as a random variable from $(\Omega',\mathscr{A}',\mathbb{P})$ to $(\mathrm{W},\mathscr{B}_\mathrm{W})$ where $\mathrm{W} = \mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)$ is equipped with the topology of uniform convergence on every compact subset of $\mathbb{R}_+$ and where $\mathscr{B}_\mathrm{W}$ is the associated Borel σ-algebra.

As a random variable on trajectories, Brownian motion is not unique. We can construct an infinite number of versions of it. What is unique is its law μ. This law is known as the Wiener measure. There exists however a special realization of Brownian motion as a random variable, which is called the canonical Brownian

[2] We can use here the following general property of metric spaces: if S and T are metric spaces with S complete, if D is a dense subset of S, and if $f : D \to T$ is uniformly continuous, then there exists a unique continuous $\widetilde{f} : S \to T$ that agrees with f on D.

motion, defined on a canonical space $(\mathrm{W},\mathscr{B}_W,\mu)$. Namely, on the probability space $(\mathrm{W}=\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d),\mathscr{B}_w,\mu)$, where μ is the Wiener measure, let us consider the coordinates process $\pi=(\pi_t)_{t\geq 0}$ defined by

$$\pi_t(w)=w_t$$

for all $t\geq 0$ and $w\in\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)$. Under μ, the process π is a d-dimensional Brownian motion issued from the origin. It is called the canonical Brownian motion.

Theorem 3.7.1. Wiener measure.

There exists a unique probability measure μ on the canonical space $(\mathrm{W},\mathscr{B}_\mathrm{W})$, called the Wiener measure, such that for all $n\geq 1$, $0<t_1<\cdots<t_n$, $A_1,\ldots,A_n\in\mathscr{B}_{\mathbb{R}^d}$,

$$\mu(\{w\in\mathrm{W}: w_{t_1}\in A_1,\ldots,w_{t_n}\in A_n\})=\int_{A_1\times\cdots\times A_n} p_{t_1-t_0}(x_1-x_0)\cdots p_{t_n-t_{n-1}}(x_n-x_{n-1})\mathrm{d}x_1\cdots\mathrm{d}x_n$$

where $t_0=0$, $x_0=0$, p is the heat or Gaussian kernel defined for all $t>0$ and $x\in\mathbb{R}^d$ by

$$p_t(x)=\frac{\mathrm{e}^{-\frac{|x|^2}{2t}}}{(\sqrt{2\pi t})^d}.$$

Moreover for all d-dimensional Brownian motion $B=(B_t)_{t\geq 0}$ issued from the origin, we have, for all measurable and bounded or positive $\Phi:\mathrm{W}\to\mathbb{R}$,

$$\mathbb{E}(\Phi(B))=\int_\mathrm{W}\Phi(w)\mu(\mathrm{d}w).$$

Proof. We know how to construct a d-dimensional Brownian motion $B=(B_t)_{t\geq 0}$ issued form the origin. If μ is the law of B seen as a random variable taking values on the canonical space $(\mathrm{W},\mathscr{B}_\mathrm{W})$, then it is immediate to get the first desired property since

$$\mu(B_{t_1}\in A_1,\ldots,B_{t_n}\in A_n)=\mu(\{w\in\mathrm{W}: w_{t_1}\in A_1,\ldots,w_{t_n}\in A_n\}).$$

Finally μ is unique because it is entirely determined on the family $\mathscr{C}$ of cylindrical subsets of W, which is stable by finite intersections and generates $\mathscr{B}_\mathrm{W}$ (monotone class argument, Corollary 1.8.4). ■

Recall that a notion of density of Wiener measure would require a notion of Lebesgue measure on Wiener space, which is missing[3].

[3] It can be shown that on an infinite-dimensional separable Banach space equipped with its Borel σ-algebra, the only locally finite and translation-invariant Borel measure is the trivial measure identically equal to zero. Equivalently, every translation-invariant measure that is not identically zero assigns infinite measure to all open subsets. See for instance `https://en.wikipedia.org/wiki/Infinite-dimensional_Lebesgue_measure` and references therein.

Chapter 4

More on martingales

For simplicity, this chapter is about continuous processes only.

4.1 Quadratic variation, square integrable martingales, increasing process

Definition 4.1.1. Quadratric variation if square integrable processes.

Let $X = (X_t)_{t\geq 0}$ be a square integrable real process such that $X_0 = 0$. The quadratic variation process $[X] = ([X]_t)_{t\geq 0}$ of X is defined for all $t \geq 0$ by the limit (when it exists)

$$[X]_t = \overset{\mathbb{P}}{\lim_{|\delta|\to 0}} \sum_k (X_{t_{k+1}} - X_{t_k})^2$$

where the convergence takes place in probability, and where $\delta : 0 = t_0 < \cdots < t_n = t,\ n = n_\delta \geq 1$, runs over all the partitions or sub-divisions of $[0,t]$, and where $|\delta| = \max_{1\leq k\leq n} |t_{k+1} - t_k|$ is the mesh of δ. More generally, the quadratic covariation process of a couple of square integrable real processes $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ is defined for all $t \geq 0$ by the following limit when it exists:

$$[X,Y]_t = \overset{\mathbb{P}}{\lim_{|\delta|\to 0}} \sum_k (X_{t_{k+1}} - X_{t_k})(Y_{t_{k+1}} - Y_{t_k}).$$

We have $[X] = [X,X]$. The set of processes with quadratic variation is a vector space. The operator $[\cdot]$ is bilinear on this space and we have by polarization $[X,Y] = \frac{1}{4}([X+Y] - [X-Y])$.

We use convergence in probability because we do not know if the process has high enough moments. Recall that for Brownian motion we have used the fourth moment for L^2 convergence of quadratic variation.

Theorem 3.2.1 states that for a BM B, we have, for all $t \geq 0$, $[B]_t = t$. Theorem 4.1.4 states that for all any square integrable continuous martingale M issued form the origin, for all $t \geq 0$, $\mathbb{E}([M]_t) = \mathbb{E}(M_t^2)$.

Lemma 4.1.2. Continuity and finite variation implies zero quadratic variation.

If a process $X = (X_t)_{t\geq 0}$ is continuous and has finite variation then it has zero quadratic variation. In other words, for a continuous process, non-zero quadratic variation implies infinite variation.

On the same topic, Lemma 4.1.6 states that a finite variation continuous martingale is constant.

Proof. Indeed, for all $t > 0$ and all partition $\delta : 0 = t_0 < \cdots < t_n = t$ of $[0,t]$, $n = n_\delta \geq 1$,

$$\sum_k (X_{t_{k+1}} - X_{t_k})^2 \leq \max_k |X_{t_{k+1}} - X_{t_k}| \sum_k |X_{t_{k+1}} - X_{t_k}| \xrightarrow[|\delta|\to 0]{} 0.$$

The max part of the right hand side tends to 0 since X is continuous and thus uniformly continuous (Heine), while the $\sum$ part is bounded by the 1-variation of X on $[0,t]$ which is finite since X has finite variation. ∎

Coding in action 4.1.3. Quadratic variation of BM.

Could you write a code simulating approximate trajectories of one-dimensional Brownian motion and their approximate quadratic variation, and plotting both on the same graphic?

We denote by $\mathscr{M}^2$ the set of square integrable continuous martingales.
We denote by $\mathscr{M}_0^2$ the set of square integrable continuous martingales issued from the origin.
We often use the following properties for any $M \in \mathscr{M}^2$:

- Squared L^2 norm of increments: for all $0 \le s \le t$,

$$\mathbb{E}((M_t - M_s)^2) = \mathbb{E}(\mathbb{E}(M_t^2 - 2M_sM_t + M_s^2 \mid \mathscr{F}_s)) = \mathbb{E}(M_t^2) - \mathbb{E}(M_s^2)$$

  and thus for any subdivision $s = t_0 < \cdots < t_n = t$, by telescoping summation,

$$\sum_{i=1}^{n} \mathbb{E}((M_{t_i} - M_{t_{i-1}})^2) = \mathbb{E}(M_t^2) - \mathbb{E}(M_s^2).$$

- (Conditional) orthogonal increments in L^2: for all $0 \le s \le t \le u \le v$ we have

$$\mathbb{E}((M_t - M_s)(M_v - M_u) \mid \mathscr{F}_t) = (M_t - M_s)\underbrace{\mathbb{E}(M_v - M_u \mid \mathscr{F}_t)}_{=M_t - M_t = 0.} = 0.$$

The following theorem is a crucial result of martingale theory.

Theorem 4.1.4. Increasing process or angle bracket.

Let $M \in \mathscr{M}_0^2$.

- There exists a unique continuous and non-decreasing process denoted $\langle M \rangle = (\langle M_t \rangle)_{t \ge 0}$ such that $\langle M \rangle_0 = 0$ and $(M_t^2 - \langle M_t \rangle)_{t \ge 0}$ is a martingale. In particular $\langle M \rangle$ is adapted.
- For all $t \ge 0$, the quadratic variation $[M]_t$ exists and $[M]_t = \langle M \rangle_t$.

Uniqueness is up to indistinguishability.
The process $\langle M \rangle$ is called the increasing process or angle bracket of M, or even the *compensator* of M^2.
If $M \in \mathscr{M}^2$ with $M_0 \ne 0$ then we define $[M] = [M - M_0]$ and $\langle M \rangle = \langle M - M_0 \rangle$.

If B is a Brownian motion, Theorem 3.1.6 gives that $\langle B \rangle_t = t$ for all $t \ge 0$ by showing that $(B_t^2 - t)_{t \ge 0}$ is a martingale, while Theorem 3.2.1 gives that $[B]_t = t$ for all $t \ge 0$ by computing the quadratic variation. More generally Lemma 4.2.6 states that for all continuous local martingale M issued from the origin, $[M] = \langle M \rangle$.

In Theorem 4.1.4, M^2 is a sub-martingale, and actually Theorem 4.1.4 states a special case of the more general Doob–Meyer[1] decomposition of sub-martingales which is beyond the scope of this course.

Corollary 4.1.5. Boundedness in L^2.

If $M \in \mathscr{M}_0^2$ then there exists a random variable $\langle M \rangle_\infty$ taking values in $[0, +\infty]$ such that almost surely

$$\langle M \rangle_t \underset{t \to \infty}{\nearrow} \langle M \rangle_\infty,$$

and moreover M is bounded in L^2 if and only if $\langle M \rangle_\infty \in \mathrm{L}^1$, more precisely, in $[0, +\infty]$,

$$\mathbb{E}(\langle M \rangle_\infty) = \sup_{t \ge 0} \mathbb{E}(M_t^2).$$

[1] Named after Paul-André Meyer (1934–2003), French mathematician.

Proof of Corollary 4.1.5. The first property follows from the monotony and positivity of $\langle M \rangle$. For the second property, since $M^2 - \langle M \rangle$ is a martingale we get $\mathbb{E}(M_t^2) = \mathbb{E}(\langle M \rangle_t)$ for all $t \geq 0$, and by monotone convergence,

$$\mathbb{E}(M_t^2) = \mathbb{E}(\langle M \rangle_t) \underset{t \to \infty}{\nearrow} \mathbb{E}(\langle M \rangle_\infty) \in [0, +\infty].$$

■

This can be skipped at first reading.

Proof of Theorem 4.1.4.

- Existence of $\langle M \rangle$ and $[M]$ and their equality when M is bounded. Let us fix $t > 0$ and let $(\delta_n)_n$ be a sequence of partitions of $[0, t]$, $\delta_n : 0 = t_0^n < \cdots < t_{r_n}^n = t$ with $|\delta_n| = \max_{1 \leq k \leq r_n}(t_k^n - t_{k-1}^n) \to 0$ as $n \to \infty$. It can be checked that the process $X = (X_s)_{s \in [0,t]}$ defined by

  $$X_s^n = \sum_{k=1}^{r_n} M_{t_{i-1}^n}(M_{t_i^n \wedge s} - M_{t_{i-1}^n \wedge s})$$

  is a (bounded) martingale (it is crucially zero when $s \leq t_i^n$), and that

  $$M_{t_k^n}^2 - 2X_{t_k^n}^n = \sum_{i=1}^{k}(M_{t_i^n} - M_{t_{i-1}^n})^2.$$

  Now it turns out that

  $$\lim_{m,n \to \infty} \mathbb{E}((X_s^n - X_s^m)^2) = 0.$$

  It follows by the Doob maximal inequality (Theorem 2.5.7) that

  $$\lim_{m,n \to \infty} \mathbb{E}\Big(\sup_{s \in [0,t]}(X_s^n - X_s^m)^2\Big) = 0.$$

  Next, for some subsequence n_k and continuous process Y, we have that almost surely $X^{n_k} \to Y$ as $k \to \infty$. Moreover Y inherits the martingale property from X. Now the process

  $$M_{t_k^n}^2 - 2X_{t_k^n}^n = \sum_{i=1}^{k}(M_{t_i^n} - M_{t_{i-1}^n})^2$$

  is non-decreasing along t_k^n, $1 \leq k \leq r_n$. Letting $n \to \infty$ gives that $M^2 - 2Y$ is almost surely non-decreasing. This shows that $[M]$ exists, is equal to $M^2 - 2Y$, and that we can take $\langle M \rangle = [M]$.

- Existence of $\langle M \rangle$ and $[M]$ and their equality when M is not bounded. For all N, we introduce the stopping time $T_N = \inf\{t \geq 0 : |M_t| \geq N\}$. From the bounded case applied to the bounded martingale $(M_{t \wedge T_N})_{t \geq 0}$, there exists a unique increasing process $(A_t^N)_{t \geq 0}$ such that $(M_{t \wedge T_n}^2 - A_t^N)_{t \geq 0}$ is a martingale. The uniqueness gives $A_{t \wedge T_N}^N = A_t^N$, and the we can define a process $(A_t)_{t \geq 0}$ by setting $A_t = A_t^N$ on the event $\{T_N \geq t\}$. Finally by monotone and dominated converge, $(M_t^2 - A_t)_{t \geq 0}$ is a martingale.

  For the quadratic variation, it suffices to write

  $$\mathbb{P}\Big(|A_t - \sum_{k=1}^{n}(M_{t_k^n} - M_{t_{k-1}^n})^2| \geq \varepsilon\Big) \leq \mathbb{P}(T_N \leq t) + \mathbb{P}\Big(|A_t^N - \sum_{k=1}^{n}(M_{t_k^n \wedge T_N} - M_{t_{k-1}^n \wedge T_N})^2| \geq \varepsilon\Big).$$

  In contrast with the bounded case, here $A_t = \langle M \rangle_t$ belongs to L^1 but not necessarily to L^2, and in particular, the convergence of $S(\delta^n) = \sum_k (M_{t_k^n} - M_{t_{k-1}^n})^2$ holds in $\mathbb{P}$ but not necessarily in L^2.

- Uniqueness of $\langle M \rangle$. If $(A_t)_{t \geq 0}$ and $(A'_t)_{t \geq 0}$ are continuous, increasing, issued from 0, such that $(M_t^2 - A_t)_{t \geq 0}$ and $(M_t^2 - A'_t)_{t \geq 0}$ are continuous martingales, then $(A_t - A'_t)_{t \geq 0}$ is a continuous finite variation martingale, and by Lemma 4.1.6, it is constant. Since $A_0 = A'_0 = 0$, we get $A = A'$.

  Lemma 4.1.6

  If $(M_s)_{s \in [0,t]}$ is a finite variation continuous martingale then it is constant.

Proof of Lemma 4.1.6. Let $(M_s)_{s\in[0,t]}$ be a finite variation continuous martingale. We may assume without loss of generality that $M_0 = 0$. For all $N \geq 1$, we introduce the stopping time

$$T_N = t \wedge \inf\{s \in [0,t] : |M_s| \geq N, \sup \sum_k |M_{t_{k+1}} - M_{t_k}| \geq N\}$$

where the supremum runs over all sub-divisions of $[0,t]$. By Theorem 2.5.1, the stopped process $(M_{s\wedge T_N})_{s\in[0,t]}$ is a bounded martingale and thus, for all $s \leq t$,

$$\mathbb{E}((M_{t\wedge T_N} - M_{s\wedge T_N})^2) = \mathbb{E}(\mathbb{E}((M_{t\wedge T_N} - M_{s\wedge T_N})^2 \mid \mathscr{F}_s)) = \mathbb{E}(M_{t\wedge T_N}^2) - \mathbb{E}(M_{s\wedge T_N}^2).$$

This gives, using a telescoping sum, for an arbitrary sub-division $\delta : 0 = t_0 < \cdots < t_n = t$,

$$\begin{aligned}
\mathbb{E}(M_{T_N}^2) &= \mathbb{E}(M_{t\wedge T_N}^2) - \mathbb{E}(M_{0\wedge T_N}^2) \\
&= \mathbb{E}\sum_k (M_{t_{k+1}\wedge T_N} - M_{t_k\wedge T_N})^2 \\
&\leq \mathbb{E}\sup_k |M_{t_{k+1}\wedge T_N} - M_{t_k\wedge T_N}| \sum_k |M_{t_{k+1}\wedge T_N} - M_{t_k\wedge T_N}| \\
&\leq N\mathbb{E}\sup_k |M_{t_{k+1}\wedge T_N} - M_{t_k\wedge T_N}|.
\end{aligned}$$

Since M is continuous, the sup in the right hand side tends a.s. to 0 as $|\delta| = \max_i(t_{i+1} - t_i) \to 0$. Since it is bounded, by dominated convergence, $\mathbb{E}(M_{T_N}^2) = 0$. Thus $M_{T_N} = 0$, which gives in turn $M_t = 0$ by sending N to ∞ and using the fact that M is continuous with finite variation. ■

■

Corollary 4.1.7. Angle bracket, square bracket, quadratic covariation.

Let $M, N \in \mathscr{M}_0^2$.

- There exists a unique continuous finite variation process $\langle M,N\rangle = (\langle M,N\rangle_t)_{t\geq 0}$ such that $\langle M,N\rangle_0 = 0$ and $(M_tN_t - \langle M_t,N_t\rangle)_{t\geq 0}$ is a martingale. In particular $\langle M,N\rangle$ is adapted.
- The quadratic covariation of (M,N) exists and $[M,N]_t = \langle M,N\rangle_t$ for all $t \geq 0$.

It is important that M and N are martingales with respect to the same filtration, the underlying $(\mathscr{F}_t)_{t\geq 0}$. By Theorem 3.1.6, if B is a d-dimensional Brownian motion then for all $1 \leq j,k \leq d$ and all $t \geq 0$,

$$\langle B^j, B^k\rangle_t = [B^j, B^k]_t = t\mathbf{1}_{j=k}.$$

Proof. We proceed by quadratic polarization. First the processes $(M_t + N_t)_{t\geq 0}$ and $(M_t - N_t)_{t\geq 0}$ are square integrable continuous martingales with respect to $(\mathscr{F}_t)_{t\geq 0}$. Next, for all $t \geq 0$, if we define $\langle M,N\rangle_t$ as

$$\langle M,N\rangle_t = \frac{1}{4}(\langle M+N\rangle_t - \langle M-N\rangle_t),$$

then $M_tN_t - \langle M,N\rangle_t = \frac{1}{4}((M_t+N_t)^2 - \langle M+N\rangle_t - ((M_t-N_t)^2 - \langle M-N\rangle_t))$, and thus $(M_tN_t - \langle M,N\rangle_t)_{t\geq 0}$ is a martingale by Theorem 4.1.4. Moreover $\langle M,N\rangle$ is continuous with finite variation as the difference of continuous and increasing processes. The uniqueness follows as in the proof of Theorem 4.1.4. The link with the quadratic covariation follows by polarization and Theorem 4.1.4. ■

Corollary 4.1.8. Stopped angle brackets.

If $M, N \in \mathscr{M}_0^2$ and S, T are stopping times then $\langle M^T, N^S\rangle = \langle M,N\rangle^{S\wedge T}$.

Proof. Theorem 2.5.1 gives that $(M^2 - \langle M\rangle)^T = (M^T)^2 - \langle M\rangle^T$ is a martingale. Now $(\langle M\rangle^T)_0 = \langle M\rangle_0 = 0$ and $\langle M\rangle^T$ is a continuous increasing process, and thus, by the uniqueness property of the increasing process provided by Theorem 4.1.4, we have $\langle M^T\rangle = \langle M\rangle^T$. By polarization we get $\langle M^T, N^T\rangle = \langle M, N\rangle^T$. Finally, $\langle M^T, N\rangle = \langle M^T, N^T\rangle$ from the equality with quadratic covariation (sum of products of increments). ■

4.2 Local martingales and localization by stopping times

If $(M_t)_{t\geq 0}$ is a martingale, then the Doob stopping theorem states that for every stopping time T, the stopped process $(M_{t\wedge T})_{t\geq 0}$ is again a martingale. Stopping can be used in general to truncate the trajectories of a process with a cutoff, in order to gain more integrability or tightness. Typically if $(X_t)_{t\geq 0}$ is an adapted process, we could consider the sequence of stopping times $(T_n)_{n\geq 0}$ defined by $T_n = \inf\{t \geq 0 : |X_t| \geq n\}$, which satisfies almost surely $T_n \nearrow +\infty$ as $n \to \infty$ and for which for all n the stopped process $(X_{t\wedge T_n})_{t\geq 0}$ is bounded. We say that $(T_n)_{n\geq 0}$ is a localizing sequence. Now a local martingale is simply an adapted processes $(X_t)_{t\geq 0}$ such that for all $n \geq 0$ the stopped process $(X_{t\wedge T_n})_{t\geq 0}$ is a (bounded) martingale. Every martingale is a local martingale. However the converse is false, and strict local martingales do exist.
Local martingales popup naturally when constructing the stochastic integral.

Definition 4.2.1. Local martingale.

- A continuous process $(M_t)_{t\geq 0}$ issued from the origin is a local martingale if it is adapted and for all $n \geq 0$, introducing the stopping time $T_n = \inf\{t \geq 0 : |M_t| \geq n\}$, the stopped process $M^{T_n} = (M_{t\wedge T_n})_{t\geq 0}$ is a martingale. It is bounded since $\sup_{t\geq 0}|M_{t\wedge T_n}| \leq |M_0| \vee n = n < \infty$.
- Since the process M is continuous, almost surely $T_n \nearrow +\infty$ as $n \to \infty$, and thus, for all $t \geq 0$, $\lim_{n\to\infty} M_{t\wedge T_n} = M_t$ almost surely. We say that the sequence $(T_n)_{n\geq 0}$ localizes or reduces M.
- If we do not have $M_0 \neq 0$, then we say that M is a local martingale when $M - M_0$ is a local martingale however we still impose that M is adapted and in particular that M_0 is $\mathscr{F}_0$ measurable.
- We denote by $\mathscr{M}^{\mathrm{loc}}$ the set of continuous local martingales w.r.t. the default filtration $(\mathscr{F}_t)_{t\geq 0}$. We denote by $\mathscr{M}_0^{\mathrm{loc}}$ the subset issued from the origin.

Remark 4.2.2. Alternative or relaxed definitions.

Equivalently we could say that a continuous adapted process $(M_t)_{t\geq 0}$ issued from the origin is a local martingale when there exists a sequence $(S_n)_{n\geq 0}$ of stopping times such that

1. almost surely $S_n \nearrow +\infty$ as $n \to \infty$
2. for all $n \geq 1$, the continuous process $M^{S_n} = (M_{t\wedge S_n})_{t\geq 0}$ is a martingale.

Moreover in this definition we could replace martingale by the stronger conditions square integrable martingale, or u.i. martingale, or bounded in L^2 martingale, or bounded martingale. Indeed, it suffices to show that M is then localized by $T_n = \inf\{t \geq 0 : |M_t| \geq n\}$. Indeed, since M is continuous, almost surely $T_n \nearrow +\infty$ as $n \to \infty$. Next, if $(S_n)_{n\geq 0}$ localizes M, then for all $n, k \geq 0$, by the Doob stopping theorem (Theorem 2.5.1) for the martingale M^{S_k} and the stopping time T_n, the process $(M^{S_k})^{T_n} = (M_{t\wedge S_k\wedge T_n})_{t\geq 0}$ is a martingale, thus for all $0 \leq s \leq t$, $\mathbb{E}(M_{t\wedge S_k\wedge T_n} \mid \mathscr{F}_s) = M_{s\wedge S_k\wedge T_n}$. Moreover since $M_0 = 0$ and M is continuous, by definition of T_n, we have $\sup_{t\geq 0}|M_{t\wedge S_k\wedge T_n}| \leq n$, and by dominated convergence, as $k \to \infty$, we have $\mathbb{E}(M_{t\wedge T_n} \mid \mathscr{F}_s) = M_{s\wedge T_n}$, hence $(M_{t\wedge T_n})_{t\geq 0}$ is a martingale.

- Localization is a truncation for processes by cutoff that has the advantage of preserving the continuity of the process and the martingale structure thanks to Doob stopping theorems.
- A martingale is always a local martingale: take $T_n = \inf\{t \geq 0 : |M_t| \geq n\}$ and use Doob stopping (Theorem 2.5.1). Note that thanks to the convention $\inf\varnothing = \infty$ we have $T_n = \infty$ on $\{\sup_{t\geq 0}|M_t| < n\}$.
- If M is a local martingale, then no integrability is guaranteed for M_t for a fixed deterministic $t \geq 0$, and we may have $M_t \not\in \mathrm{L}^1$. Moreover for every stopping time T, the stopped process $M^T = (M_{t\wedge T})_{t\geq 0}$ is a local martingale but the Doob stopping theorem does not hold in general even if T is bounded.

Remark 4.2.3. Domination as a martingale criterion.

If M is a continuous local martingale dominated by an integrable random variable, in the sense that $\mathbb{E}\sup_{t\geq 0}|M_t| < \infty$, then, for all $t \geq 0$ and $s \in [0,t]$, by continuity and dominated convergence,

$$M_s = \lim_{n\to\infty} M_{s\wedge T_n} = \lim_{n\to\infty} \mathbb{E}(M_{t\wedge T_n} \mid \mathscr{F}_s) = \mathbb{E}(\lim_{n\to\infty} M_{t\wedge T_n} \mid \mathscr{F}_s) = \mathbb{E}(M_t \mid \mathscr{F}_s)$$

for any localization sequence $(T_n)_n$ for M, hence M is a u.i. martingale. However, there exists continuous local martingales which are bounded in L^2 and thus u.i. and which are not a martingale!

Remark 4.2.4. Strict local martingales.

Are there local martingales which are not martingales? Yes[a].

- If M is a martingale, for instance Brownian motion, and if U is measurable with respect to $\mathscr{F}_0$, then $(U+M_t)_{t\geq 0}$ is a local martingale, and a martingale if and only if $U \in \mathrm{L}^1$. Note that if M_0 is constant and $\mathscr{F} = \sigma(M_0) = \{\Omega, \varnothing\}$ then necessarily U is constant and we cannot have $U \notin \mathrm{L}^1$.
- Let M be a martingale such that $M_0 = 1$, such as the Doléans-Dade exponential. Let U be a random variable independent of M. Then $(UM_t)_{t\geq 0}$ is a local martingale with respect to the enlarged filtration $(\sigma(\sigma(U)\cup\mathscr{F}_t))_{t\geq 0}$, localized by $T_n = \inf\{t \geq 0 : |UM_t| \geq n\}$. This is in fact an Itô stochastic integral, see Exercise 4 of the 2020-2021 exam.
- Let $(B_t)_{t\geq 0}$ be a 3-dimensional BM with $B_0 = x \neq 0$. The process $(|B_t|)_{t\geq 0}$ is a Bessel process. It can be shown that the inverse Bessel process $(|B_t|^{-1})_{t\geq 0}$ is a local martingale, localized by $T_n = \inf\{t \geq 0 : |B_t| \leq 1/(|x|+n)\}$, but is not a martingale. Moreover it is bounded in L^2 and thus u.i.! For a proof, see Exercise 3 of the 2020-2021 exam[b].

[a] Some other famous examples are listed on `https://en.wikipedia.org/wiki/Local_martingale`
[b] Or `https://djalil.chafai.net/blog/2020/10/31/back-to-basics-local-martingales/`

Remark 4.2.5. Vector spaces.

The set $\mathscr{M}^{\mathrm{loc}}$ and $\mathscr{M}_0^{\mathrm{loc}}$ are real vector spaces. Indeed if $M, M' \in \mathscr{M}_0^{\mathrm{loc}}$ are localized respectively by $(T_n)_{n\geq 1}$ and $(T'_n)_{n\geq 1}$, then by the Doob stopping theorem (Theorem 2.5.1), $(S_n)_{n\geq 0} = (T_n \wedge T'_n)_{n\geq 0}$ localizes both M and M'. For all $n \geq 0$, the process $(M+M')^{S_n} = M^{S_n} + M'^{S_n}$ is a square integrable martingale. Note that we have also the following (strict) inclusions:

$$\begin{array}{ccccc} \mathbb{M}_0^2 & \subset & \mathscr{M}_0^2 & \subset & \mathscr{M}_0^{\mathrm{loc}} \\ \cap & & \cap & & \cap \\ \mathbb{M}^2 & \subset & \mathscr{M}^2 & \subset & \mathscr{M}^{\mathrm{loc}} \end{array}$$

Lemma 4.2.6. Increasing process, angle bracket, quadratic variation, square bracket.

Let $M, N \in \mathscr{M}_0^{\mathrm{loc}}$.

1. there exists a unique continuous finite variation process denoted $(\langle M,N\rangle_t)_{t\geq 0}$ with
$$\langle M,N\rangle_0 = 0 \quad \text{and} \quad (M_tN_t - \langle M,N\rangle_t)_{t\geq 0} \in \mathscr{M}_0^{\mathrm{loc}}.$$
Moreover $\langle M,N\rangle = \frac{1}{4}(\langle M+N\rangle - \langle M-N\rangle)$ where $\langle M\rangle = \langle M,M\rangle$.
2. $\langle M\rangle$ is the unique non-decreasing process such that $M^2 - \langle M\rangle$ is a continuous local martingale
3. M is localized by $T_n = \inf\{t \geq 0 : |M_t| \geq n \text{ or } \langle M\rangle_t \geq n\}$ and for all $n \geq 0$,
$$\sup_{t\geq 0} |M_t^{T_n}| \leq n \quad \text{and} \quad \sup_{t\geq 0} \langle M^{T_n}\rangle_t \leq n.$$

4. For all $t \geq 0$ if $(\delta_n)_{n\geq 1}$ is a sequence of sub-divisions of $[0,t]$, $\delta_n : 0 = t_0^n < \cdots < t_{m_n}^n = t$, then

$$S(\delta_n) = \sum_{k=1}^{m_n} (M_{t_k^n} - M_{t_{k-1}^n})(N_{t_k^n} - N_{t_{k-1}^n}) \xrightarrow[n\to\infty]{\mathbb{P}} [M,N]_t = \langle M,N\rangle_t$$

provided that $|\delta_n| = \max_{1\leq k\leq m_n}(t_k^n - t_{k-1}^n) \to 0$ as $n\to\infty$. Furthermore

$$[M,N] = \frac{1}{4}([M+N]-[M-N]) \quad \text{where} \quad [M] = [M,M].$$

We say that $\langle M\rangle$ is the increasing process of M.
We say that $\langle M,N\rangle$ is the finite variation process or angle bracket of the couple (M,N).
We say that $[M]$ is the quadratic variation of M.
We say that $[M,N]$ is the quadratic covariation or square bracket of the couple (M,N).
As for martingales, if $M \in \mathscr{M}^{\text{loc}}$ then we set $\langle M\rangle = \langle M - M_0\rangle$ and $[M] = [M-M_0]$, in particular $\langle M\rangle = [M]$.
As for martingales, $\langle M\rangle_t$ is not necessarily in L^2, and in particular $S(\delta) \to \langle M\rangle$ may not hold in L^2.

Proof.

1. If $(S_n)_{n\geq 0}$ localizes M and $(T_n)_{n\geq 0}$ localizes N then $(U_n)_{n\geq 0} = (T_n \wedge S_n)_{n\geq 0}$ localizes both M and N. By uniqueness of the increasing process of square integrable continuous martingales (Theorem 4.1.4) used for the square integrable martingales M^{U_n} and N^{U_n}, we get that for all $0 \leq n \leq m$ and $t \geq 0$,

$$\langle M^{U_m}, N^{U_m}\rangle_{t\wedge U_n} = \langle M^{U_n}, N^{U_n}\rangle_t,$$

hence $(\langle M^{U_m}, N^{U_m}\rangle)_{t\geq 0}$ and $(\langle M^{U_n}, N^{U_n}\rangle)_{t\geq 0}$ are equal up to time U_n. We then define, for all $t \geq 0$,

$$\langle M,N\rangle_t = \lim_{n\to\infty} \langle M^{U_n}, N^{U_n}\rangle_t.$$

This is the unique continuous process with finite variations and issued from the origin, denoted $\langle M,N\rangle$ such that for all $t \geq 0$ and all $n \geq 0$, $\langle M,N\rangle_{t\wedge U_n} = \langle M^{U_n}, N^{U_n}\rangle_t$. We then set $\langle M\rangle = \langle M,M\rangle$.

2. Take $M = N$ in the previous item.

3. It suffices to proceed as in Remark 4.2.3. Note that $\langle M^{T_n}\rangle = \langle M\rangle^{T_n}$ gives $|\langle M^{T_n}\rangle| \leq n$.

4. We reduce to $M = N$ by polarization. Next, let $(T_n)_{n\geq 0}$ be a localization sequence for M. For all $n \geq 0$, Theorem 4.1.4 used for the square integrable martingale M^{T_n} gives

$$S^{T_n}(\delta) = \sum_i (M_{t_{i+1}}^{T_n} - M_{t_i}^{T_n}) \xrightarrow[|\delta|\to 0]{\mathrm{L}^2} \langle M^{T_n}\rangle_t = \langle M\rangle_{T_n\wedge t}.$$

Now for all $\varepsilon > 0$ and all $n \geq 0$,

$$\begin{aligned}\mathbb{P}(|S(\delta) - \langle M\rangle_t| > \varepsilon) &\leq \mathbb{P}(T_n \leq t) + \mathbb{P}(|S(\delta) - \langle M\rangle_t| > \varepsilon, t < T_n)\\ &\leq \mathbb{P}(T_n \leq t) + \mathbb{P}(|S^{T_n}(\delta) - \langle M\rangle_{T_n\wedge t}| > \varepsilon),\end{aligned}$$

and therefore $\lim_{|\delta|\to 0} \mathbb{P}(|S(\delta) - \langle M\rangle_t| > \varepsilon) = 0$.

■

Lemma 4.2.7. Martingale criterion.

Let M be a continuous local martingale with $M_0 \in \mathrm{L}^2$.

1. The following properties are equivalent:

   (a) M is a martingale which is square integrable

   (b) $\mathbb{E}(\langle M\rangle_t) < \infty$ for all $t \geq 0$.

2. The following properties are equivalent:
   (a) M is a martingale which is bounded in L^2 and $\sup_{t\geq 0}\mathbb{E}(M_t^2)=\mathbb{E}(\langle M\rangle_\infty)$
   (b) $\mathbb{E}(\langle M\rangle_\infty)<\infty$

   Moreover, in this case $M^2-\langle M\rangle$ is a u.i. martingale and $\mathbb{E}(M_\infty^2)=\mathbb{E}(M_0^2)+\mathbb{E}(\langle M\rangle_\infty)$.

The proof of the lemma is rather short but uses many typical martingale ingredients!

Proof. By replacing M with $M-M_0$, we assume without loss of generality that $M_0=0$.

1. If M is a square integrable martingale then $M^2-\langle M\rangle$ is a martingale and in particular $\langle M\rangle_t\in\mathrm{L}^1$ for all $t\geq 0$. Conversely, if M is a continuous local martingale with $\langle M\rangle_t\in\mathrm{L}^1$ for all $t\geq 0$ then since $M^2-\langle M\rangle$ is a continuous local martingale, it follows that there exists a sequence $(T_n)_{n\geq 0}$ of stopping times such that almost surely $T_n\nearrow+\infty$ as $n\to\infty$ and for all $n\geq 0$ the process $(M^{T_n})^2-\langle M\rangle^{T_n}$ is a square integrable continuous martingale issued from 0. Hence, for all $t\geq 0$, using monotone convergence,
$$\mathbb{E}(M_{t\wedge T_n}^2)=\mathbb{E}(\langle M\rangle_{t\wedge T_n})\underset{n\to\infty}{\longrightarrow}\mathbb{E}(\langle M\rangle_t)<\infty.$$
   This implies that $(M_{t\wedge T_n})_{n\geq 0}$ is bounded in L^2. On the other hand, it follows by the Fatou lemma that
$$\mathbb{E}(M_t^2)=\mathbb{E}\Big(\underline{\lim}_{n\to\infty}M_{t\wedge T_n}^2\Big)\leq\underline{\lim}_{n\to\infty}\mathbb{E}(M_{t\wedge T_n}^2)<\infty.$$
   Finally, since for all $t\geq 0$, $(M_{t\wedge T_n})_{n\geq 0}$ is bounded in L^2, it is u.i., and thus, for all $0\leq s\leq t$, since $\lim_{n\to\infty}M_{t\wedge T_n}=M_t$ a.s., this convergence holds in L^1 and we obtain the martingale property via
$$\mathbb{E}(M_t\mid\mathscr{F}_s)=\mathbb{E}\Big(\lim_{n\to\infty}M_{t\wedge T_n}\mid\mathscr{F}_s\Big)=\lim_{n\to\infty}\mathbb{E}(M_{t\wedge T_n}\mid\mathscr{F}_s)=\lim_{n\to\infty}M_{s\wedge T_n}=M_s.$$

2. If M is a martingale bounded in L^2 then, by Corollary 4.1.5, $\langle M\rangle_\infty\in\mathrm{L}^1$. Conversely, if M is a local martingale with $\langle M\rangle_\infty\in\mathrm{L}^1$, then, by monotony and positivity of $\langle M\rangle$, $\langle M\rangle_t\in\mathrm{L}^1$ for all $t\geq 0$, next, by the first part, M is a square integrable martingale, and thus, by Corollary 4.1.5, M is bounded in L^2.

   Finally if M is a martingale bounded in L^2, then the Doob maximal inequality (Theorem 2.5.7) gives
$$\mathbb{E}\Big(\sup_{s\in[0,t]}M_s^2\Big)\leq 4\mathbb{E}(M_t^2)$$
   for all $t\geq 0$, and by sending t to ∞, we get, by monotone convergence,
$$\mathbb{E}\Big(\sup_{t\geq 0}M_t^2\Big)\leq 4\sup_{t\geq 0}\mathbb{E}(M_t^2).$$
   This gives the domination
$$\sup_{t\geq 0}|M_t^2-\langle M\rangle_t|\leq\sup_{t\geq 0}M_t^2+\langle M\rangle_\infty\in\mathrm{L}^1,$$
   which implies that $M^2-\langle M\rangle$ is uniformly integrable.

■

Remark 4.2.8. Vocabulary.

If X and A are continuous adapted processes with $X_0=A_0=0$ with A of finite variation and such that $X-A$ is a local martingale then A is unique and is called the compensator of X. For instance if X is a continuous local martingale with $X_0=0$ then the compensator of X^2 is $\langle X\rangle$.

Remark 4.2.9. Advanced link with Brownian motion.

The Lévy characterization of Brownian motion states that among all continuous local martingales, Brownian motion is characterized by its angle bracket. On the other hand, the Dubins–Schwarz theorem states that all continuous local martingales with angle bracket tending to infinity at infinity are time changed Brownian motion by the angle bracket.

The following result is essential for the Dubins–Schwarz theorem.

Theorem 4.2.10. Simultaneous flatness for M and $\langle M\rangle$.

Let M be a continuous local martingale. Then the processes M and $\langle M\rangle$ are constant on same intervals, in the sense that almost surely, for all $0 \le a < b$,

$$\forall t \in [a,b], M_t = M_a \quad \text{if and only if} \quad \langle M\rangle_b = \langle M\rangle_a.$$

Proof of Theorem 4.2.10. Since M and $\langle M\rangle$ are continuous, it suffices to show that for all $0 \le a \le b$, a.s.

$$\{\forall t \in [a,b] : M_t = M_a\} = \{\langle M\rangle_b = \langle M\rangle_a\}.$$

The inclusion $\subset$ comes from the approximation of the quadratic variation $\langle M\rangle = [M]$. Let us prove the converse. To this end, we consider the continuous local martingale $(N_t)_{t\ge0} = (M_t - M_{t\wedge a})_{t\ge0}$. We have

$$\langle N\rangle = \langle M\rangle - 2\langle M, M^a\rangle + \langle M^a\rangle = \langle M\rangle - 2\langle M\rangle^a + \langle M\rangle^a = \langle M\rangle - \langle M\rangle^a.$$

For all $\varepsilon > 0$, we set the stopping time $T_\varepsilon = \inf\{t \ge 0 : \langle N\rangle_t > \varepsilon\}$. The continuous semi-martingale N^{T_ε} satisfies $N_0^{T_\varepsilon} = 0$ and $\langle N^{T_\varepsilon}\rangle_\infty = \langle N\rangle_{T_\varepsilon} \le \varepsilon$. By Lemma 4.2.7, N^{T_ε} is a martingale bounded in L^2, and for all $t \ge 0$,

$$\mathbb{E}(N^2_{t\wedge T_\varepsilon}) = \mathbb{E}(\langle N\rangle_{t\wedge T_\varepsilon}) \le \varepsilon.$$

Let us define the event $A = \{\langle M\rangle_b = \langle M\rangle_a\}$. Then $A \subset \{T_\varepsilon \ge b\}$ and, for all $t \in [a,b]$,

$$\mathbb{E}(\mathbf{1}_A N_t^2) = \mathbb{E}(\mathbf{1}_A N^2_{t\wedge T_\varepsilon}) \le \mathbb{E}(N^2_{t\wedge T_\varepsilon}) \le \varepsilon.$$

By sending ε to 0 we obtain $\mathbb{E}(\mathbf{1}_A N_t^2) = 0$ and thus $N_t = 0$ almost surely on A. ∎

4.3 Convergence in L^2 and the Hilbert space $\mathbb{M}_0^2$

Let $\mathbb{M}_0^2$ be the set of continuous martingales issued from the origin and bounded in L^2, for some fixed underlying filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\ge0}, \mathbb{P})$.

The elements of $\mathbb{M}_0^2$ are centered: for all $M \in \mathbb{M}_0^2$ and all $t \ge 0$, $\mathbb{E}(M_t) = \mathbb{E}(M_0) = 0$.

For all $M \in \mathbb{M}_0^2$, we have $M_0 = 0$ and $\sup_{t\ge0}\mathbb{E}(M_t^2) < \infty$. By Theorem 2.1.3, we see the elements of $\mathbb{M}_0^2$ as random variables taking values in $(\mathscr{C}(\mathbb{R}_+,\mathbb{R}), \mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R})})$. In particular for all $M, N \in \mathbb{M}_0^2$, we have $M = N$ iff M and N are indistinguishable in other words $\mathbb{P}(\forall t \ge 0 : M_t = N_t) = 1$. Also $M = 0$ iff for all $t \ge 0$, $M_t = 0$.

Theorem 4.3.1. Hilbert structure on $\mathbb{M}_0^2$.

The set $\mathbb{M}_0^2$ is a Hilbert space with scalar product $\langle M, N\rangle_{\mathbb{M}_0^2} = \mathbb{E}(\langle M, N\rangle_\infty)$.
Moreover, for all $M \in \mathbb{M}_0^2$, we have $\|M\|^2_{\mathbb{M}_0^2} = \mathbb{E}(\langle M\rangle_\infty) = \sup_{t\ge0}\mathbb{E}(\langle M\rangle_t) = \sup_{t\ge0}\mathbb{E}(M_t^2)$.

More generally, it can be shown similarly that for all fixed $T > 0$, the set $\mathbb{M}^2_{0,T}$ of square integrable continuous martingales $(M_t)_{t\in[0,T]}$ such that $M_0 = 0$ is a Hilbert space for the scalar product $\langle M, N\rangle_{\mathbb{M}^2_{0,T}} = \mathbb{E}(\langle M, N\rangle_T)$. In this case, for all $M \in \mathbb{M}^2_{0,T}$, we have $\|M\|^2_{\mathbb{M}^2_{0,T}} = \sup_{t\in[0,T]}\mathbb{E}(\langle M\rangle_t) = \sup_{t\in[0,T]}\mathbb{E}(M_t^2)$.

Proof. The facts that $\mathbb{M}_0^2$ is a vector space and that $\langle\cdot\rangle$ is bilinear, symmetric, and non-negative on the diagonal are almost immediate. For the positivity, if $M \in \mathbb{M}_0^2$ with $\mathbb{E}(\langle M\rangle_\infty) = 0$ then we have $\langle M\rangle_t = 0$ for all $t \geq 0$, hence $\mathbb{E}(M_t^2) = 0$ for all $t \geq 0$, thus $M_t = 0$ for all $t \geq 0$. To prove completeness, let $(M^{(n)})_{n\geq 1}$ be a Cauchy sequence in $\mathbb{M}_0^2$. Then for all $\varepsilon > 0$, there exists $r \geq 1$ such that for all $m, n \geq r$, $\|M^{(n)} - M^{(m)}\|_{\mathbb{M}_0^2} \leq \varepsilon$. Thus

$$\sup_{t\geq 0} \mathbb{E}(|M_t^{(n)} - M_t^{(m)}|^2) \leq \varepsilon^2.$$

This implies that for all $t \geq 0$, $(M_t^{(n)})_{t\geq 0}$ is a Cauchy sequence in $\mathbb{L}^2$, and thus converges to an element $M_t \in \mathrm{L}^2$. It follows that $M = (M_t)_{t\geq 0}$ is a square integrable martingale, issued from the origin. It remains to prove that M is continuous. To this end, the idea is to use uniform convergence on finite time intervals. Namely, let us fix $t > 0$. From the L^2 convergence, there exists a sub-sequence $(n_k)_{k\geq 1}$ such that for all $k \geq 1$,

$$\mathbb{E}(|M_t^{(n_k)} - M_t^{(n_k+1)}|^2) \leq 2^{-k}.$$

Now the Doob maximal inequality (Theorem 2.5.7) for the martingale $(M_t^{(n)} - M_t^{(n+1)})_{t\geq 0}$ gives

$$\mathbb{E}(\sup_{s\in[0,t]} |M_s^{(n_k)} - M_s^{(n_k+1)}|^2) \leq 4\mathbb{E}(|M_t^{(n_k)} - M_t^{(n_k+1)}|^2) \leq 2^{-k+2},$$

and thus, by monotone convergence or the Fubini–Tonelli theorem,

$$\mathbb{E}\Big(\sum_{k\geq 1} \sup_{s\in[0,t]} |M_s^{(n_k)} - M_s^{(n_k+1)}|^2\Big) = \sum_{k\geq 1} \mathbb{E}\Big(\sup_{s\in[0,t]} |M_s^{(n_k)} - M_s^{(n_k+1)}|^2\Big) < \infty.$$

Therefore for all $t > 0$, almost surely

$$\sum_{k\geq 1} \sup_{s\in[0,t]} |M_s^{(n_k)} - M_s^{(n_k+1)}| < \infty.$$

Lemma 4.3.2. Criterion.

In a Banach space if $\sum_{n=1}^{\infty} \|u_n - u_{n+1}\| < \infty$ then $(u_n)_{n\geq 1}$ converges.

The converse is false, for instance $u_n = \frac{(-1)^n}{n} \underset{n\to\infty}{\longrightarrow} 0$ but $|u_n - u_{n+1}| \underset{n\to\infty}{\sim} \frac{2}{n}$ and thus $\sum_{n=1}^{\infty} |u_n - u_{n+1}| = \infty$.

Proof of Lemma 4.3.2. The sequence $(u_n)_{n\geq 1}$ is Cauchy since for all $n \geq 1$ and $m \geq 1$ we have

$$\|u_{n+m} - u_n\| \leq \sum_{k=n}^{n+m-1} \|u_{k+1} - u_k\| \leq \sum_{k\geq n} \|u_{k+1} - u_k\| \underset{n\to\infty}{\longrightarrow} 0.$$

■

By using Lemma 4.3.2 with the Banach space $(\mathscr{C}([0,t],\mathbb{R}), \|\cdot\| = \sup_{[0,t]} |\cdot|)$, this implies that for all $t > 0$, almost surely, the sequence of continuous functions $(s \in [0,t] \mapsto M_s^{(n_k)})_{k\geq 1}$ converges uniformly towards a limit denoted $(M'_s)_{s\geq[0,t]}$ which is continuous thanks the uniform convergence. This almost sure event can be chosen independent of t for instance by taking integer values for t. Now for all $t \geq 0$, $(M_t^{(n_k)})_{k\geq 1}$ converges to M_t in L^2 and to M'_t almost surely, and therefore $M_t = M'_t$. ■

Theorem 4.3.3. Convergence of martingales bounded in L^2.

Let M be a square integrable martingale bounded in L^2. Then there exists $M_\infty \in \mathrm{L}^2$ such that

$$\lim_{t\to\infty} M_t = M_\infty \text{ almost surely and in } \mathrm{L}^2.$$

Note that M is uniformly integrable because it is bounded in L^p with $p = 2 > 1$.

Proof. Let us show that M satisfies the L^2 Cauchy criterion. Recall that for all $0 \le s \le t$, we have

$$\mathbb{E}((M_t - M_s)^2) = \mathbb{E}(M_t^2 - 2M_s\mathbb{E}(M_t \mid \mathscr{F}_s) + M_s^2) = \mathbb{E}(M_t^2 - M_s^2).$$

But M^2 is a sub-martingale and $t \mapsto \mathbb{E}(M_t^2)$ grows and is bounded above by $\sup_{t\ge 0}\mathbb{E}(M_t^2) < \infty$. Thus $\lim_{t\to\infty}\mathbb{E}(M_t^2)$ exists. Hence $(M_t)_{t\ge 0}$ is Cauchy in L^2, and therefore it converges in L^2 towards some $M_\infty \in \mathrm{L}^2$. It remains to establish the almost sure convergence. Now, by the Markov inequality, for all $s \ge 0$ and all $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}\Big(\sup_{t\ge s}|M_t - M_\infty| \ge \varepsilon\Big) &\le \frac{1}{\varepsilon^2}\mathbb{E}\Big(\sup_{t\ge s}(M_t - M_\infty)^2\Big)\\ &\le \frac{2}{\varepsilon^2}\Big(\mathbb{E}\Big((M_s - M_\infty)^2\Big) + \mathbb{E}\Big(\sup_{t\ge s}(M_t - M_s)^2\Big)\Big).\end{aligned}$$

Now the monotone convergence theorem gives

$$\mathbb{E}\Big(\sup_{t\ge s}(M_t - M_s)^2\Big) = \lim_{T\to\infty}\mathbb{E}\Big(\sup_{t\in[s,T]}(M_t - M_s)^2\Big).$$

On the other hand, for all $s \ge 0$, the process $(|M_t - M_s|)_{t\ge s}$ is a continous non-negative sub-martingale, for which the Doob maximal inequality of Theorem 2.5.7 gives

$$\mathbb{E}\Big(\sup_{t\ge s}(M_t - M_s)^2\Big) \le \lim_{T\to\infty} 4\mathbb{E}((M_T - M_s)^2) = 4\mathbb{E}((M_\infty - M_s)^2).$$

Therefore we obtain

$$\mathbb{P}\Big(\sup_{t\ge s}|M_t - M_\infty| \ge \varepsilon\Big) \le \frac{10}{\varepsilon^2}\mathbb{E}((M_s - M_\infty)^2) \underset{s\to\infty}{\longrightarrow} 0.$$

Since the right hand side decreases as s grows, we get, for all $\varepsilon > 0$,

$$\mathbb{P}\Big(\cap_{s\in\mathbb{Q}_+}\{\sup_{t\ge s}|M_t - M_\infty| \ge \varepsilon\}\Big) = \lim_{s\to\infty}\mathbb{P}\Big(\sup_{t\ge s}|M_t - M_\infty| \ge \varepsilon\Big) = 0,$$

Similarly, the right hand side decreases as ε grows, and then

$$\mathbb{P}\Big(\cup_{\varepsilon\in\mathbb{Q}_+}\cap_{s\in\mathbb{Q}_+}\{\sup_{t\ge s}|M_t - M_\infty| \ge \varepsilon\}\Big) = \lim_{\varepsilon\to 0}\lim_{s\to\infty}\mathbb{P}\Big(\sup_{t\ge s}|M_t - M_\infty| \ge \varepsilon\Big) = 0,$$

which means that $\lim_{t\to\infty} M_t = M_\infty$ almost surely! ■

4.4 Convergence in L^1, closedness, uniform integrability

As for the sum of independent and identically distributed random variables, there is, for martingales, in a way, an L^2 theory and an L^1 theory. The L^2 theory is in a sense simpler due to the Hilbert structure.

Theorem 4.4.1. Doob convergence theorem for martingales bounded in L^1.

Let M be a continuous martingale bounded in L^1. Then there exists $M_\infty \in \mathrm{L}^1$ such that

$$\lim_{n\to\infty} M_t = M_\infty \text{ almost surely.}$$

Moreover the convergence holds in L^1 if and only if M is uniformly integrable.

If M is a non-negative martingale, then it is always bounded in L^1.
If M is martingale bounded in L^1 but not u.i., then $\mathbb{E}(M_t) = \mathbb{E}(M_0)$ for all $t \ge 0$ but $\mathbb{E}(M_\infty) \ne \mathbb{E}(M_0)$.

This can be skipped at first reading.

Proof. We can assume that $M_0 = 0$, otherwise consider the martingale $M - M_0 = (M_t - M_0)_{t\ge 0}$ which is also bounded in L^1, making $M_t \to M_0 + (M - M_0)_\infty$ a.s. We proceed by truncation and reduction to the square integrable case. By the Doob maximal inequality (Theorem 2.5.7) with $p=1$ and all $r>0$,

$$\mathbb{P}\Big(\sup_{s\in[0,t]}|M_s|\geq r\Big)\leq\frac{\mathbb{E}(|M_t|)}{r}.$$

By monotone convergence, with $C=\sup_{t\geq 0}\mathbb{E}(|M_t|)<\infty$, for all $r>0$,

$$\mathbb{P}\Big(\sup_{t\geq 0}|M_t|\geq r\Big)\leq\frac{C}{r}.$$

It follows that

$$\mathbb{P}\Big(\sup_{t\geq 0}|M_t|=\infty\Big)\leq\lim_{r\to\infty}\mathbb{P}\Big(\sup_{t\geq 0}|M_t|\geq r\Big)=0,$$

in other words almost surely $(M_t)_{t\geq 0}$ is bounded: $\sup_{t\geq 0}|M_t|<\infty$. Thus, there exists an almost sure event, say Ω', on which for all $n\geq\sup_{t\geq 0}|M_t|$ (beware that this threshold on n is random),

$$T_n=\inf\{t\geq 0:|M_t|\geq n\}=\infty.$$

Next, by Doob stopping (Theorem 2.5.1), for all $n\geq 0$, $(M_{t\wedge T_n})_{t\geq 0}$ is a martingale, bounded since $\sup_{t\geq 0}|M_{t\wedge T_n}|\leq|M_0|\vee n=n$ (M is continuous and $M_0=0$). Now, since $(M_{t\wedge T_n})_{t\geq 0}$ is bounded in L^2, by Theorem 4.3.3, there exists $M_\infty^{(n)}\in\mathrm{L}^2$ such that $\lim_{t\to\infty}M_{t\wedge T_n}=M_\infty^{(n)}$ almost surely (and in L^2 but this is useless). Let us denote by Ω_n the almost sure event on which this holds. Then, on the almost sure event $\Omega'\cap(\cap_n\Omega_n)$, for all $t\geq 0$ and $n\geq\sup_{s\geq 0}|M_s|$, we have $M_{t\wedge T_n}=M_t$, thus the sequence $(M_\infty^{(n)})_n$ is stationary in the sense that $M_\infty^{(n)}$ is constant when $n\geq\sup_{s\geq 0}|M_s|$, hence, if M_∞ is its limit,

$$\lim_{t\to\infty}M_t=M_\infty.$$

Contrary to $M_\infty^{(n)}$, the limit M_∞ has not reason to belong to L^2. However $M_\infty\in\mathrm{L}^1$ since from the almost sure convergence, the boundedness in L^1 of $(M_t)_{t\geq 0}$, and by using the Fatou lemma, we have

$$\mathbb{E}(|M_\infty|)=\mathbb{E}(\varliminf_{t\to\infty}|M_t|)\leq\varliminf_{t\to\infty}\mathbb{E}(|M_t|)\leq C<\infty.$$

Finally an almost sure convergence to an L^1 limit holds in L^1 if and only if the sequence is u.i. ■

The result remains valid for super-martingales.

Theorem 4.4.2: Doob convergence theorem for super-martingales bounded in L^1

Let M be a continuous super-martingale bounded in L^1. Then there exists $M_\infty\in\mathrm{L}^1$ such that

$$\lim_{n\to\infty}M_t=M_\infty\text{ almost surely.}$$

Note that a non-negative super-martingale is automatically bounded in L^1.

Proof. See for instance [14, Theorem 3.19 page 58–59] for a classical proof using oscillations. ■

Remark 4.4.3. Non-negative local martingales are super-martingales.

If $(M_t)_{t\geq 0}$ is a non-negative continuous local martingale, then it is a non-negative super-martingale and by Theorem 4.4.2 it converges almost surely to an integrable random variable. Indeed, if $(T_n)_n$ is a localizing sequence then for all $t\geq 0$ and $s\in[0,t]$, by the Fatou Lemma,

$$\mathbb{E}(M_t\mid\mathcal{F}_s)=\mathbb{E}(\lim_{n\to\infty}M_{t\wedge T_n}\mid\mathcal{F}_s)\leq\lim_{n\to\infty}\mathbb{E}(M_{t\wedge T_n}\mid\mathcal{F}_s)=\lim_{n\to\infty}M_{s\wedge T_n}=M_s.$$

Note that the conditional expectations are well defined in $[0,+\infty]$ because M is non-negative.

Corollary 4.4.4. Convergence of martingales bounded in L^p, $p > 1$.

If M is a continuous martingale bounded in L^p with $p > 1$ then there exists $M_\infty \in \mathrm{L}^p$ such that

$$\lim_{t\to\infty} M_t = M_\infty \text{ almost surely and in } \mathrm{L}^p.$$

In particular, for $p = 2$ this gives an alternative to Theorem 4.3.3.

This can be skipped at first reading.

Proof. Since M is a super-martingale bounded in L^1, Theorem 4.4.1 gives $M_\infty \in \mathrm{L}^1$ such that $\lim_{t\to\infty} M_t = M_\infty$ almost surely. But since M is bounded in L^p with $p > 1$, it follows that M is uniformly integrable, and therefore $\lim_{t\to\infty} M_t = M_\infty$ in L^1. We have $M_\infty \in \mathrm{L}^p$ since by the Fatou lemma,

$$\mathbb{E}(|M_\infty|^p) = \mathbb{E}(\varliminf_{t\to\infty} |M_t|^p) \le \varliminf_{t\to\infty} \mathbb{E}(|M_t|^p) < \infty.$$

On the other hand, by the Doob maximal inequality (Theorem 2.5.7), since M is bonded in L^p, for all $t \ge 0$, $\sup_{s\in[0,t]} |M_s|^p \in \mathrm{L}^1$ and $\mathbb{E}(\sup_{s\in[0,t]} |M_s|^p) \le c_p \mathbb{E}(|M_t|^p)$. Therefore, by monotone convergence,

$$\mathbb{E}\Big(\sup_{t\ge0} |M_t|^p\Big) \le \sup_{t\ge0} \mathbb{E}(|M_t|^p) < \infty.$$

Hence $\sup_{t\ge0} |M_t|^p \in \mathrm{L}^1$, and thus, by dominated convergence, $\lim_{t\to\infty} M_t = M_\infty$ in L^p. ■

Corollary 4.4.5. Doob theorem on closed martingales or Doob martingale convergence theorem.

Let M be a continuous martingale. The following properties are equivalent:

1. (convergence) M_t converges in L^1 as $t \to \infty$
2. (closedness) there exists $M_\infty \in \mathrm{L}^1$ such that for all $t \ge 0$, $M_t = \mathbb{E}(M_\infty \mid \mathscr{F}_t)$
3. (integrability) the family $\{M_t : t \ge 0\}$ is uniformly integrable.

In this case, for all $t \ge 0$, $M_t = \mathbb{E}(M_\infty \mid \mathscr{F}_t)$, and $\lim_{t\to\infty} M_t = M_\infty$ a.s. and in L^1, and $\mathbb{E}(M_0) = \mathbb{E}(M_\infty)$.

If M is a martingale then for all fixed $a \ge 0$, the stopped martingale $M^a = (M_{t\wedge a})_{t\ge0}$ is closed by M_a since $\mathbb{E}(M_a \mid \mathscr{F}_t) = M_a \mathbf{1}_{a\le t} + M_t \mathbf{1}_{a>t} = M_{t\wedge a}$. Hence M^a is uniformly integrable. Note that $\lim_{t\to\infty} M_{t\wedge a} = M_a$.

Note that in the proof below, Theorem 4.4.1 is used in every implication of the equivalence.

This can be skipped at first reading.

Proof. 1. ⇒ 2. If M converges in L^1, then it is bounded in L^1, and by Theorem 4.4.1, its converges a.s. to $M_\infty \in \mathrm{L}^1$ (the convergence holds also in L^1 but we do not use this fact now). For all $t \ge 0$ and $s \in [0, t]$ and all $A \in \mathscr{F}_s$, the martingale property for M gives $\mathbb{E}(M_t \mathbf{1}_A) = \mathbb{E}(M_s \mathbf{1}_A)$. By dominated convergence as $t \to \infty$, we get $\mathbb{E}(M_\infty \mathbf{1}_A) = \mathbb{E}(M_s \mathbf{1}_A)$ therefore $M_s = \mathbb{E}(M_\infty \mid \mathscr{F}_s)$ for all $s \ge 0$.
2. ⇒ 3. Let us assume that for some $M_\infty \in \mathrm{L}^1$ we have $M_t = \mathbb{E}(M_\infty \mid \mathscr{F}_t)$ for all $t \ge 0$. Then $\sup_{t\ge0} \mathbb{E}(|M_t|) \le \mathbb{E}(|M_\infty|) < \infty$ and thus, by Theorem 4.4.1, M_t converges a.s. as $t \to \infty$. It follows that almost surely $M_* = \sup_{t\ge0} |M_t| < \infty$. Now $\lim_{R\to\infty} \mathbf{1}_{M_*\ge R} = 0$ almost surely, and for all $R \ge 0$,

$$\sup_{t\ge0} \mathbb{E}(|M_t| \mathbf{1}_{|M_t|\ge R}) = \sup_{t\ge0} \mathbb{E}(|\mathbb{E}(M_\infty \mid \mathscr{F}_t)| \mathbf{1}_{|M_t|\ge R}) \le \sup_{t\ge0} \mathbb{E}(|M_\infty| \mathbf{1}_{|M_t|\ge R}) \le \mathbb{E}(|M_\infty| \mathbf{1}_{M_*\ge R}) \xrightarrow[R\to\infty]{} 0$$

where the convergence follows by dominated convergence. Therefore M is u.i.
3. ⇒ 1. If M is u.i. then it is bounded in L^1, and from Theorem 4.4.1, there exists $M_\infty \in \mathrm{L}^1$ such that $\lim_{t\to\infty} M_t = M_\infty$ a.s. Since M is u.i., the convergence holds in L^1. ■

The following generalizes the Doob stopping theorem (Theorem 2.5.1).

Corollary 4.4.6. Doob stopping for uniformly integrable martingales.

Let M be a u.i. continuous martingale and let T be a stopping time (not necessarily bounded or finite). We set $M_T = M_\infty$ on $\{T=\infty\}$ where $M_\infty = \lim_{t\to\infty} M_t$ is as in Corollary 4.4.5. Then:

1. $(M_{t\wedge T})_{t\geq 0}$ is a uniformly integrable martingale, $M_T \in \mathrm{L}^1$, and for all $t \geq 0$, $M_{t\wedge T} = \mathbb{E}(M_T \mid \mathcal{F}_t)$.
   In particular, for all $t\geq 0$, $\mathbb{E}(M_0) = \mathbb{E}(M_{t\wedge T}) = \mathbb{E}(M_T)$.

2. Moreover if S is another stopping time with $S \leq T$ then $M_S = \mathbb{E}(M_T \mid \mathcal{F}_S)$.
   In particular, for all stopping time S, $M_S = \mathbb{E}(M_\infty \mid \mathcal{F}_S)$ and $\mathbb{E}(M_S) = \mathbb{E}(M_\infty) = \mathbb{E}(M_0)$.

This can be skipped at first reading.

Proof. We will prove the first property by using the second property.

1. For all $t \geq 0$, both $t\wedge T$ and T are stopping times. By the second property of the present theorem, $M_{t\wedge T} \in \mathrm{L}^1$ and $M_T \in \mathrm{L}^1$. Moreover $M_{t\wedge T}$ is measurable for $\mathcal{F}_{t\wedge T}$, and thus for $\mathcal{F}_t$ since $t \leq t\wedge T$. Now, in order to prove that $\mathbb{E}(M_T \mid \mathcal{F}_t) = M_{t\wedge T}$, it suffices to show that for all $A \in \mathcal{F}_t$,
$$\mathbb{E}(\mathbf{1}_A M_T) = \mathbb{E}(\mathbf{1}_A M_{t\wedge T}).$$
   But for all $A \in \mathcal{F}_t$, we have immediately from $T = t\wedge T$ on $\{T \leq t\}$ that
$$\mathbb{E}(\mathbf{1}_{A\cap\{T\leq t\}} M_T) = \mathbb{E}(\mathbf{1}_{A\cap\{T\leq t\}} M_{t\wedge T}).$$
   The second property of the present theorem for the stopping times $S = t\wedge T$ and T gives
$$M_{t\wedge T} = \mathbb{E}(M_T \mid \mathcal{F}_{t\wedge T}).$$
   Now since $A\cap\{T>t\} \in \mathcal{F}_t$ and $A\cap\{T>t\}\in\mathcal{F}_T$, we get $A\cap\{T>t\}\in\mathcal{F}_t\cap\mathcal{F}_T = \mathcal{F}_{t\wedge T}$, and
$$\mathbb{E}(\mathbf{1}_{A\cap\{T>t\}} M_T) = \mathbb{E}(\mathbf{1}_{A\cap\{T>t\}} M_{t\wedge T}).$$
   By adding this to a previous formula we get the desired result $\mathbb{E}(\mathbf{1}_A M_T) = \mathbb{E}(\mathbf{1}_A M_{t\wedge T})$.
   Finally, the fact that $M^T = (M_{t\wedge T})_{t\geq 0}$ is a martingale follows from what precedes used with the u.i. martingale $M^a = (M_{t\wedge a})_{t\geq 0}$ for all $a \geq 0$, which gives $M^a_{s\wedge T} = \mathbb{E}(M^a_T \mid \mathcal{F}_s)$ for all $s \geq 0$, in other words $M_{s\wedge a\wedge T} = \mathbb{E}(M_{a\wedge T} \mid \mathcal{F}_s)$. Taking $a = t \geq s$ gives the martingale property for M^T.

2. Following for instance [14, Theorem 3.22 page 59], we discretize as in the proof of Theorem 2.5.1 or Theorem 3.5.1. Namely, for all $n \geq 0$, let us define the stopping times
$$S_n = \sum_{k=0}^\infty \frac{k+1}{2^n}\mathbf{1}_{k2^{-n}<S\leq(k+1)2^{-n}} + \infty\mathbf{1}_{S=\infty} \quad\text{and}\quad T_n = \sum_{k=0}^\infty \frac{k+1}{2^n}\mathbf{1}_{k2^{-n}<T\leq(k+1)2^{-n}} + \infty\mathbf{1}_{T=\infty}.$$
   We have $S_n \searrow S$ and $T_n \searrow T$ as $n\to\infty$, and $S_n \leq T_n$ for all $n\geq 0$. Next, for all $n \geq 0$, $2^n S_n$ and $2^n T_n$ are integer valued stopping times for the discrete time filtration $(\mathcal{F}^{(n)}_k)_{k\geq 0} = (\mathcal{F}_{k2^{-n}})_{k\geq 0}$, while $M^{(n)} = (M_{k2^{-n}})_{k\geq 0}$ is a uniformly integrable discrete time martingale with respect to this filtration. By using the Doob stopping theorem for u.i. discrete time martingales, we get
$$M_{S_n} = M^{(n)}_{2^n S_n} = \mathbb{E}(M^{(n)}_{2^n T_n} \mid \mathcal{F}^{(n)}_{2^n S_n}) = \mathbb{E}(M_{T_n} \mid \mathcal{F}_{S_n}).$$
   Now, for all $A \in \mathcal{F}_S \subset \mathcal{F}_{S_n}$, we have $\mathbb{E}(\mathbf{1}_A M_{S_n}) = \mathbb{E}(\mathbf{1}_A M_{T_n})$. Since M is (right) continuous, a.s.
$$M_S = \lim_{n\to\infty} M_{S_n} \quad\text{and}\quad M_T = \lim_{n\to\infty} M_{T_n}.$$
   For the L^1 convergence, the Doob stopping theorem for u.i. discrete time martingales gives $M_{S_n} = \mathbb{E}(M_\infty \mid \mathcal{F}_{S_n})$ for all $n \geq 0$ and thus $(M_{S_n})_{n\geq 0}$ and $(M_{T_n})_{n\geq 0}$ are u.i. This also gives that $M_S \in \mathrm{L}^1$ and $M_T \in \mathbb{L}^1$. This also allows to pass to the limit in $\mathbb{E}(\mathbf{1}_A M_{S_n}) = \mathbb{E}(\mathbf{1}_A M_{T_n})$ to get $\mathbb{E}(\mathbf{1}_A M_S) = \mathbb{E}(\mathbf{1}_A M_T)$. This holds for all $A \in \mathcal{F}_S$, thus $M_S = \mathbb{E}(M_T \mid \mathcal{F}_S)$.

■

Bibliography

[1] F. Baudoin – *Diffusion processes and stochastic calculus*, EMS Textbooks in Mathematics, EMS, 2014. iii, 12, 16

[2] R. Bhattacharya & E. C. Waymire – *A basic course in probability theory*, second ed., Universitext, Springer, Cham, 2016. iii

[3] V. S. Borkar – *Probability theory*, Universitext, Springer-Verlag, New York, 1995, An advanced course. iii

[4] B. Bru & M. Yor – "Comments on the life and mathematical legacy of W. Doeblin", *Finance Stoch.* **6** (2002), no. 1, p. 3–47. vi

[5] C. Dellacherie & P.-A. Meyer – *Probabilities and potential*, North-Holland Mathematics Studies, vol. 29, North-Holland, 1978. iii, 16, 18

[6] — , *Probabilities and potential. B*, North-Holland Mathematics Studies, vol. 72, North-Holland Publishing, 1982, Theory of martingales, Translated from the French by J. P. Wilson. iii

[7] — , *Probabilities and potential. C*, North-Holland Mathematics Studies, vol. 151, North-Holland, 1988, Potential theory for discrete and continuous semigroups, Translated from the French by J. Norris. iii

[8] B. Duplantier – "Brownian motion, "diverse and undulating"", in *Einstein, 1905–2005*, Prog. Math. Phys., vol. 47, Birkhäuser, 2006, Translated from the French by Emily Parks, p. 201–293. vi, 40

[9] R. Durrett – *Probability—theory and examples*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 49, Cambridge University Press, Cambridge, 2019, Fifth edition of [MR1068527]. 1

[10] A. Gut – *Probability: a graduate course*, second ed., Springer Texts in Statistics, Springer, New York, 2013. iii, 1

[11] J. Jacod & P. Protter – *Probability essentials*, second ed., Universitext, Springer-Verlag, Berlin, 2003. iii

[12] J.-P. Kahane – "Le mouvement brownien: un essai sur les origines de la théorie mathématique", in *Matériaux pour l'histoire des mathématiques au XXe siècle (Nice, 1996)*, Sémin. Congr., vol. 3, Soc. Math. France, Paris, 1998, p. 123–155. vi

[13] H.-H. Kuo – *Introduction to stochastic integration*, Universitext, Springer, New York, 2006. 32

[14] J.-F. Le Gall – *Brownian motion, martingales, and stochastic calculus*, french ed., Graduate Texts in Mathematics, vol. 274, Springer, 2016. iii, 23, 56, 58

[15] E. Nelson – *Dynamical theories of Brownian motion*, Princeton University Press, 1967. vi

[16] J. Perrin – *Les atomes (1913)*, CNRS Éditions, 2014. 28

[17] D. Revuz & M. Yor – *Continuous martingales and Brownian motion*, third ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 293, Springer, 1999. iii

[18] M. S. Taqqu – "Bachelier and his times: a conversation with Bernard Bru", *Finance Stoch.* **5** (2001), no. 1, p. 3–32. vi, 40