# Lecture Notes on Stochastic Calculus (Part I)

Fabrizio Gelsomino, Olivier Lévêque, EPFL

December 17, 2009

## Contents

# 1 Probability "review"

## 1.1 $\sigma$-fields

In probability, the *fundamental set* $\Omega$ describes the set of all possible *outcomes* (or *realizations*) of a given experiment. It might be *any* set, without any particular structure, such as for example $\Omega = \{1, \dots, 6\}$ representing the outcomes of a die roll, or $\Omega = [0, 1]$ representing e.g. the outcomes of a concentration measurement of some chemical product. Notice moreover that the set $\Omega$ need not be composed of numbers exclusively. It is e.g. perfectly valid to consider the set $\Omega = \{\text{banana, apple, orange}\}$.

Given a fundamental set $\Omega$, it is important to describe what *information* does one have on the system, namely on the outcomes of the experiment. This notion of information is well captured by the mathematical notion of *$\sigma$-field*, which is defined below. Notice that in elementary probability courses, it is generally assumed that the information one has about a system is *complete*, so that it becomes useless to introduce the concept below.

**Definition 1.1.** Let $\Omega$ be a set. A *$\sigma$-field* (or *$\sigma$-algebra*) on $\Omega$ is a collection $\mathcal{F}$ of subsets of $\Omega$ (or *events*) satisfying the following three properties or *axioms*:

(i) $\emptyset \in \mathcal{F}$.

(ii) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.

(iii) If $(A_n)_{n=1}^{\infty} \subset \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$. In particular, if $A, B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$.

The following properties can be further deduced from the above axioms (this is left as an exercise):

(iv) $\Omega \in \mathcal{F}$.

(v) If $(A_n)_{n=1}^{\infty} \subset \mathcal{F}$, then $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$. In particular, if $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$.

(vi) If $A, B \in \mathcal{F}$ and $A \subset B$, then $B \backslash A \in \mathcal{F}$.

**Terminology.** The pair $(\Omega, \mathcal{F})$ is called a *measurable space* and the events belonging to $\mathcal{F}$ are said to be *$\mathcal{F}$-measurable*, that is, they are the events that one can decide on whether they happened or not, given the information $\mathcal{F}$. In other words, if one knows the information $\mathcal{F}$, then one is able to tell to which events of $\mathcal{F}$ (= subsets of $\Omega$) does the realization of the experiment $\omega$ belong.

**Example.** For a generic set $\Omega$, the following are always $\sigma$-fields:

$\mathcal{F}_0 = \{\emptyset, \Omega\}$ (= *trivial* $\sigma$-field).
$\mathcal{P}(\Omega) = \{\text{all subsets of } \Omega\}$ (= *complete* $\sigma$-field).

**Example 1.2.** Let $\Omega = \{1, \dots, 6\}$. The following are $\sigma$-fields on $\Omega$:

$\mathcal{F}_1 = \{\emptyset, \{1\}, \{2, \dots, 6\}, \Omega\}$.
$\mathcal{F}_2 = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$.

**Example 1.3.** Let $\Omega = [0, 1]$ and $I_1, \dots, I_n$ be a family of disjoint intervals in $\Omega$ such that $I_1 \cup \dots \cup I_n = \Omega$ ($\{I_1, \dots, I_n\}$ is also called a *partition* of $\Omega$). The following is a $\sigma$-field on $\Omega$:

$$\mathcal{F}_3 = \{\emptyset, I_1, \dots, I_n, I_1 \cup I_2, \dots, I_1 \cup I_2 \cup I_3, \dots, \Omega\} \quad \text{(NB: there are } 2^n \text{ events in total in } \mathcal{F}_3\text{)}.$$

**$\sigma$-field generated by a collection of events.**

An event carries in general more information than itself. As an example, if one knows whether the result of a die roll is odd (corresponding to the event $\{1, 3, 5\}$), then one also knows of course whether the result is even (corresponding to the event $\{2, 4, 6\}$). It is therefore convenient to have a mathematical description of the information generated by a single event, or more generally by a family of events.

**Definition 1.4.** Let $\mathcal{A} = \{A_i,\, i \in I\}$ be a collection of events, where $I$ need not be a countable set. The *$\sigma$-field generated by $\mathcal{A}$* is the smallest $\sigma$-field on $\Omega$ containing all the events $A_i$. It is denoted as $\sigma(\mathcal{A})$.

**Example.** Let $\Omega = \{1, \ldots, 6\}$ (cf. Example 1.2).

Let $\mathcal{A}_1 = \{\{1\}\}$. Then $\sigma(\mathcal{A}_1) = \mathcal{F}_1$.
Let $\mathcal{A}_2 = \{\{1, 3, 5\}\}$. Then $\sigma(\mathcal{A}_2) = \mathcal{F}_2$.
Let $\mathcal{A} = \{\{1\}, \ldots, \{6\}\}$. Then $\sigma(\mathcal{A}) = \mathcal{P}(\Omega)$.

**Exercise.** Let $\mathcal{A} = \{\{1, 2, 3\}, \{1, 3, 5\}\}$. Compute $\sigma(\mathcal{A})$.

**Example.** Let $\Omega = [0, 1]$ and let $\mathcal{A}_3 = \{I_1, \ldots, I_n\}$ (cf. Example 1.3). Then $\sigma(\mathcal{A}_3) = \mathcal{F}_3$.

**Borel $\sigma$-field.** Another important example of generated $\sigma$-field on $\Omega = [0, 1]$ is the following:

$$\mathcal{B}([0, 1]) = \sigma(\{\,]a, b[ :\, a, b \in [0, 1],\, a < b\}),$$

is the *Borel $\sigma$-field* on $[0, 1]$ and elements of $\mathcal{B}([0, 1])$ are called the *Borel subsets* of $[0, 1]$. As surprising as it may be, $\mathcal{B}([0, 1]) \neq \mathcal{P}([0, 1])$, which generates some difficulties from the theoretical point of view. Nevertheless, it is quite difficult to construct explicit examples of subsets of $[0, 1]$ which are *not* in $\mathcal{B}([0, 1])$.

**Sub-$\sigma$-field.**

One may have more or less information about a system. In mathematical terms, this translates into the fact that a $\sigma$-field has more or less elements. It is therefore convenient to introduce a (partial) ordering on the ensemble of existing $\sigma$-fields, in order to establish a *hierarchy* of information. This notion of hierarchy is important and will come back when we will be studying stochastic processes that evolve in time.

**Definition 1.5.** Let $\Omega$ be a set and $\mathcal{F}$ be a $\sigma$-field on $\Omega$. A *sub-$\sigma$-field of $\mathcal{F}$* is a collection $\mathcal{G}$ of events such that:

(i) If $A \in \mathcal{G}$, then $A \in \mathcal{F}$.

(ii) $\mathcal{G}$ is itself a $\sigma$-field.

**Notation.** $\mathcal{G} \subset \mathcal{F}$.

**Remark.** Let $\Omega$ be a generic set. The trivial $\sigma$-field $\mathcal{F}_0 = \{\emptyset, \Omega\}$ is a sub-$\sigma$-field of any other $\sigma$-field on $\Omega$. Likewise, any $\sigma$-field on $\Omega$ is a sub-$\sigma$-field of the complete $\sigma$-field $\mathcal{P}(\Omega)$.

**Example.** Let $\Omega = \{1, \ldots, 6\}$ (cf. Example 1.2). Notice that $\mathcal{F}_1$ is *not* a sub-$\sigma$-field of $\mathcal{F}_2$ (even though $\{1\} \subset \{1, 3, 5\}$), nor is $\mathcal{F}_2$ a sub-$\sigma$-field of $\mathcal{F}_1$. In general, notice that

1) If $A \in \mathcal{G}$ and $\mathcal{G} \subset \mathcal{F}$, then it is true that $A \in \mathcal{F}$.

but

2) $A \subset B$ and $B \in \mathcal{G}$ together do *not* imply that $A \in \mathcal{G}$.

**Example.** Let $\Omega = [0, 1]$ (cf. Example 1.3). Then $\mathcal{F}_3$ is a sub-$\sigma$-field of $\mathcal{B}([0, 1])$.

## 1.2 Random variables

The notion of random variable is usually introduced in elementary probability courses as a vague concept, essentially characterized by its distribution. In mathematical terms however, random variables do exist prior to their distribution: they are functions from the fundamental set $\Omega$ to $\mathbb{R}$. Here is a preliminary definition.

**Definition 1.6.** On the set $\mathbb{R}$, one defines the *Borel $\sigma$-field* as

$$\mathcal{B}(\mathbb{R}) = \sigma(\{\,]a, b[ :\, a, b \in \mathbb{R}, a < b\}).$$

The elements of $\mathcal{B}(\mathbb{R})$ are called *Borel sets*. Again, notice that $\mathcal{B}(\mathbb{R})$ is strictly included in $\mathcal{P}(\mathbb{R})$.

**Definition 1.7.** Let $(\Omega, \mathcal{F})$ be a measurable space. A *random variable* on $(\Omega, \mathcal{F})$ is a map $X : \Omega \to \mathbb{R}$ satisfying

$$\{\omega \in \Omega \,:\, X(\omega) \in B\} \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbb{R}). \tag{1}$$

**Notation.** One often simply denotes the set $\{\omega \in \Omega \,:\, X(\omega) \in B\} = \{X \in B\} = X^{-1}(B)$: it is called the inverse image of the set $B$ through the map $X$ (watch out that $X$ need not be a bijective function in order for this set to be well defined).

**Terminology.** The above random variable $X$ is sometimes called *$\mathcal{F}$-measurable*, in order to emphasize that if one knows the information $\mathcal{F}$, then one knows the value of $X$.

**Example.** If $\mathcal{F} = \mathcal{P}(\Omega)$, then condition (1) is always satisfied, so every map $X : \Omega \to \mathbb{R}$ is an $\mathcal{F}$-measurable random variable. On the contrary, if $\mathcal{F} = \{\emptyset, \Omega\}$, then the only random variables which are $\mathcal{F}$-measurable are the maps $X : \Omega \to \mathbb{R}$ which are constant.

**Remark.** Condition (1) can be shown to be equivalent to the following condition:

$$\{\omega \in \Omega \,:\, X(\omega) \leq t\} \in \mathcal{F}, \quad \forall t \in \mathbb{R},$$

which is significantly easier to check.

**Definition 1.8.** Let $(\Omega, \mathcal{F})$ be a measurable space and $A \in \mathcal{F}$ be an event. Then the map $\Omega \to \mathbb{R}$ defined as

$$\omega \mapsto 1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{otherwise,} \end{cases}$$

is a random variable on $(\Omega, \mathcal{F})$. It is called the *indicator function* of the event $A$.

**Example.** Let $\Omega = \{1, \ldots, 6\}$ and $\mathcal{F} = \mathcal{P}(\Omega)$ (cf. Example 1.2). Then $X_1(\omega) = \omega$ and $X_2(\omega) = 1_{\{1,3,5\}}(\omega)$ are both random variables on $(\Omega, \mathcal{F})$. Moreover, $X_2$ is $\mathcal{F}_2$-measurable, but notice that $X_1$ is neither $\mathcal{F}_1$- nor $\mathcal{F}_2$-measurable.

**Example.** Let $\Omega = [0, 1]$ and $\mathcal{F} = \mathcal{B}([0, 1])$ (cf. Example 1.3). Then $X_3(\omega) = \sum_{j=1}^{n} x_j 1_{I_j}(\omega)$ and $X_4(\omega) = \omega$ are both random variables on $(\Omega, \mathcal{F})$. Notice however that only $X_3$ is $\mathcal{F}_3$-measurable.

We will need to consider not only random variables, but also functions of random variables. This is why we introduce the following definition.

**Definition 1.9.** A map $g : \mathbb{R} \to \mathbb{R}$ such that

$$\{x \in \mathbb{R} \,:\, g(x) \in B\} \in \mathcal{B}(\mathbb{R}), \quad \forall B \in \mathcal{B}(\mathbb{R}),$$

is called a *Borel-measurable function* on $\mathbb{R}$.

**Remark.** A Borel-measurable function on $\mathbb{R}$ is therefore nothing but a random variable on the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

As it is difficult to construct explicitly sets which are not Borel sets, it is equally difficult to construct functions which are not Borel-measurable. Nevertheless, one often needs to check that a given function is Borel-measurable. A useful criterion for this is the following (given here without proof).

**Proposition 1.10.** If $g : \mathbb{R} \to \mathbb{R}$ is continuous, then it is Borel-measurable.

Finally, let us mention this useful property of functions of random variables.

**Proposition 1.11.** If $X$ is an $\mathcal{F}$-measurable random variable and $g : \mathbb{R} \to \mathbb{R}$ is Borel-measurable, then $Y = g(X)$ is also an $\mathcal{F}$-measurable random variable.

*Proof.* Let $B \in \mathcal{B}(\mathbb{R})$. Then

$$\{Y \in B\} = \{g(X) \in B\} = \{X \in g^{-1}(B)\} \in \mathcal{F},$$

since $X$ is an $\mathcal{F}$-measurable random variable and $g^{-1}(B) \in \mathcal{B}(\mathbb{R})$ by assumption. □

**$\sigma$-field generated by a collection of random variables.**

The amount of information contained in a random variable, or more generally in a collection of random variables, is given by the definition below.

**Definition 1.12.** Let $(\Omega, \mathcal{F})$ be a measurable space and $\{X_i,\, i \in I\}$ be a collection of random variables on $(\Omega, \mathcal{F})$. The *$\sigma$-field generated by $X_i$, $i \in I$*, denoted as $\sigma(X_i,\, i \in I)$, is the smallest $\sigma$-field $\mathcal{G}$ on $\Omega$ such that all the random variables $X_i$ are $\mathcal{G}$-measurable.

**Remark.** Notice that

$$\sigma(X_i,\, i \in I) = \sigma(\{\{X_i \in B\},\, i \in I,\, B \in \mathcal{B}(\mathbb{R})\}),$$

where the right-hand side expression refers to Definition 1.4. It turns out that one also has

$$\sigma(X_i,\, i \in I) = \sigma(\{\{X_i \leq t\},\, i \in I,\, t \in \mathbb{R}\}).$$

**Example.** Let $(\Omega, \mathcal{F})$ be a measurable space. If $X_0$ is a constant random variable (i.e. $X_0(\omega) = c \in \mathbb{R},\, \forall \omega \in \Omega$), then $\sigma(X_0) = \{\emptyset, \Omega\}$.

**Example.** Let $\Omega = \{1, \ldots, 6\}$ and $\mathcal{F} = \mathcal{P}(\Omega)$ (cf. Example 1.2). Then $\sigma(X_1) = \mathcal{P}(\Omega)$ and $\sigma(X_2) = \mathcal{F}_2$.

**Example.** Let $\Omega = [0, 1]$ and $\mathcal{F} = \mathcal{B}([0, 1])$ (cf. Example 1.3). Then $\sigma(X_3) = \mathcal{F}_3$ and $\sigma(X_4) = \mathcal{B}([0, 1])$.

Following the proof of Proposition 1.11, the proposition below can be easily shown.

**Proposition 1.13.** If $X$ is a random variable on a measurable space $(\Omega, \mathcal{F})$ and $g : \mathbb{R} \to \mathbb{R}$ is Borel-measurable, then $Y = g(X)$ is a $\sigma(X)$-measurable random variable (this applies in particular to $Y = X$).

As a matter of fact, it turns out that the reciprocal statement is also true: if $Y$ is a $\sigma(X)$-measurable random variable, then there exists a Borel-measurable function $g : \mathbb{R} \to \mathbb{R}$ such that $Y = g(X)$.

## 1.3 Probability measures

**Definition 1.14.** Let $(\Omega, \mathcal{F})$ be a measurable space. A *probability measure* on $(\Omega, \mathcal{F})$ is a map $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfying the following two axioms:

(i) $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$.

(ii) If $(A_n)_{n=1}^{\infty} \subset \mathcal{F}$ is such that $A_n \cap A_m = \emptyset$, $\forall n \neq m$, then $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.
In particular, if $A, B \in \mathcal{F}$ are such that $A \cap B = \emptyset$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

The following properties can be further deduced from the above axioms:

(iii) If $(A_n)_{n=1}^{\infty} \subset \mathcal{F}$, then $\mathbb{P}(\cup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.
In particular, if $A, B \in \mathcal{F}$, then $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.

(iv) If $A, B \in \mathcal{F}$ and $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$ and $\mathbb{P}(B \backslash A) = \mathbb{P}(B) - \mathbb{P}(A)$.
In particular, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(v) If $A, B \in \mathcal{F}$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

(vi) If $(A_n)_{n=1}^{\infty} \subset \mathcal{F}$ is such that $A_n \subset A_{n+1}$, $\forall n$, then $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \lim_{n \to \infty} \mathbb{P}(A_n)$.

(vii) If $(A_n)_{n=1}^{\infty} \subset \mathcal{F}$ is such that $A_n \supset A_{n+1}$, $\forall n$, then $\mathbb{P}(\cap_{n=1}^{\infty} A_n) = \lim_{n \to \infty} \mathbb{P}(A_n)$.

**Terminology.** The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*. Property (ii) is referred to as the *$\sigma$-additivity* (or simply *additivity* in the finite case) of probability measures.

**Example.** Let $\Omega = \{1, .., 6\}$ and $\mathcal{F} = \mathcal{P}(\Omega)$ be the measurable space associated to a die roll. The probability measure associated to a balanced die is defined as

$$\mathbb{P}_1(\{i\}) = \frac{1}{6},\ \forall i \in \{1, \ldots, 6\},$$

and is extended by additivity to all subsets of $\Omega$. E.g.,

$$\mathbb{P}_1(\{1,3,5\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

The probability measure associated to a loaded die is defined as

$$\mathbb{P}_2(\{6\}) = 1 \quad \text{and} \quad \mathbb{P}_2(\{i\}) = 0, \ \forall i \in \{1,\dots,5\},$$

and is extended by additivity to all subsets of $\Omega$.

**Example.** Let $\Omega = [0,1]$ and $\mathcal{F} = \mathcal{B}([0,1])$. One defines the following probability measure on the subintervals of $[0,1]$:

$$\mathbb{P}(\,]a,b[\,) = b - a.$$

**Fact.** $\mathbb{P}$ can be extended by $\sigma$-additivity to all Borel subsets of $[0,1]$. It is called the *Lebesgue measure* on $[0,1]$ and is sometimes denoted as $\mathbb{P}(B) = |B|$.

## 1.4 Distribution of a random variable

**Definition 1.15.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X$ be a random variable defined on this probability space. The *distribution* of $X$ is the map $\mu_X : \mathcal{B}(\mathbb{R}) \to [0,1]$ defined as

$$\mu_X(B) = \mathbb{P}(\{X \in B\}), \quad B \in \mathcal{B}(\mathbb{R}).$$

**Remark.** The triple $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$ forms a new probability space.

**Notation.** If a random variable $X$ has distribution $\mu$, this is denoted as $X \sim \mu$. Likewise, if two random variables $X$ and $Y$ share the same distribution $\mu$, then they are are said to be *identically distributed* and this is denoted as $X \sim Y \sim \mu$.

**Example 1.16.** The probability space describing two independent (and balanced) dice rolls is $\Omega = \{1,\dots,6\} \times \{1,\dots,6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and

$$\mathbb{P}(\{(i,j)\}) = \frac{1}{36}, \quad \forall (i,j) \in \Omega.$$

Let $X_1(i,j) = i$ be the result of the first die, and $Y(i,j) = i + j$ be the sum of the two dice. Then

$$\mu_{X_1}(\{i\}) = \mathbb{P}(\{X_1 = i\}) = \mathbb{P}(\{(i,1),\dots,(i,6)\}) = \frac{6}{36} = \frac{1}{6}, \quad \forall i \in \{1,\dots,6\},$$

and

$$\mu_Y(\{2\}) = \mathbb{P}(\{Y = 2\}) = \mathbb{P}(\{(1,1)\}) = \frac{1}{36}, \quad \mu_Y(\{3\}) = \mathbb{P}(\{Y = 3\}) = \mathbb{P}(\{(1,2),(2,1)\}) = \frac{1}{18}.$$

More generally:

$$\mu_Y(\{i\}) = \frac{6 - |7 - i|}{36}, \quad i \in \{2,\dots,12\}.$$

**Cumulative distribution function.**

**Definition 1.17.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X$ be a random variable defined on this probability space. The *cumulative distribution function (or cdf)* of $X$ is the map $F_X : \mathbb{R} \to [0,1]$ defined as

$$F_X(t) = \mu_X(\,]-\infty, t]) = \mathbb{P}(\{X \le t\}), \quad t \in \mathbb{R}.$$

**Fact.** The knowledge of $F_X$ is equivalent to the knowledge of $\mu_X$.

From the properties of probability measures, one deduces easily that the cdf of a random variable satisfies the following properties:

(i) $\lim_{t\to-\infty} F_X(t) = 0$, $\lim_{t\to+\infty} F_X(t) = 1$.

(ii) $F_X$ is *non-decreasing*, i.e. $F_X(s) \leq F_X(t)$ for all $s < t$.

(iii) $F_X$ is *right-continuous* on $\mathbb{R}$, i.e. $\lim_{\varepsilon\downarrow 0} F_X(t+\varepsilon) = F_X(t)$, for all $t \in \mathbb{R}$.

**Remark.** $F_X$ has at most a countable number of jumps on the real line. If $F_X$ has a jump of size $p \in [0,1]$ at $t \in \mathbb{R}$, this actually means that $\mathbb{P}(\{X = t\}) = F_X(t) - \lim_{\varepsilon\downarrow 0} F_X(t-\varepsilon) = p$.

**Two important classes of random variables.**

**Discrete random variables.**

**Definition 1.18.** $X$ is a *discrete random variable* if it takes values in a countable subset $C$ of $\mathbb{R}$, that is, $\mathbb{P}(\{X \in C\}) = 1$.

The distribution of a discrete random variable is entirely characterized by the numbers $p_x = \mathbb{P}(\{X = x\})$, where $x \in C$. Notice that $p_x \geq 0$ for all $x \in C$ and that $\sum_{x\in C} p_x = \mathbb{P}(\{X \in C\}) = 1$. Moreover,

$$\mu_X(B) = \mathbb{P}(\{X \in B\}) = \sum_{x\in B} p_x, \quad \forall B \in \mathcal{B}(\mathbb{R}),$$

and

$$F_X(t) = \mathbb{P}(\{X \leq t\}) = \sum_{x\leq t} p_x, \quad \forall t \in \mathbb{R},$$

is a step function.

**Example.** A binomial random variable $X$ with parameters $n \geq 1$ and $p \in [0,1]$ (denoted as $X \sim \mathrm{Bi}(n,p)$) takes values in $\{0,\ldots,n\}$ and is characterized by the numbers

$$p_k = \mathbb{P}(\{X = k\}) = \left( \begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}, \quad k \in \{0,\ldots,n\},$$

where $\left( \begin{array}{c} n \\ k \end{array} \right) = \dfrac{n!}{k!(n-k)!}$ are the binomial coefficients.

**Continuous random variables.**

**Definition 1.19.** $X$ is a *continuous random variable* if $\mathbb{P}(\{X \in B\}) = 0$ whenever $B \in \mathcal{B}(\mathbb{R})$ is such that $|B| = 0$ (remember that $|B|$ is the Lebesgue measure of $B$).

In particular, this implies that if $X$ is a continuous random variable, then $\mathbb{P}(\{X = x\}) = 0$ $\forall x \in \mathbb{R}$ (as $|\{x\}| = 0$ $\forall x \in \mathbb{R}$).

**Fact.** If $X$ is a continuous random variable according to the above definition, then there exists a function $f_X : \mathbb{R} \to \mathbb{R}$, called the *probability density function (or pdf)* of $X$, such that $f_X(x) \geq 0$ $\forall x \in \mathbb{R}$, $\int_{\mathbb{R}} f_X(x)\,dx = 1$ and

$$\mu_X(B) = \mathbb{P}(\{X \in B\}) = \int_B f_X(x)\,dx, \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Moreover,

$$F_X(t) = \mathbb{P}(\{X \leq t\}) = \int_{-\infty}^{t} f_X(x)\,dx, \quad \forall t \in \mathbb{R},$$

is a differentiable function (whose derivative is $F_X'(t) = f_X(t)$).

**Example.** A Gaussian random variable $X$ with mean $\mu$ and variance $\sigma^2$ (denoted as $X \sim \mathcal{N}(\mu, \sigma^2)$) takes values in $\mathbb{R}$ and has pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

## 1.5 Independence

The notion of independence is a central notion in probability. It is usually defined for events and random variables in elementary probability courses. Nevertheless, as it will become clear below, the independence between *σ-fields* turns out to be the most natural concept (remembering that a $\sigma$-field is related to the amount of information one has on a system).

In the three paragraphs below, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a generic probability space.

**Independence of events.**

One starts by defining the independence of two events in $\mathcal{F}$.

**Definition 1.20.** Two events $A, B \in \mathcal{F}$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

**Notation.** $A \perp\!\!\!\perp B$.

**Proposition 1.21.** If two events $A, B \in \mathcal{F}$ are independent, then it also holds that

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A)\,\mathbb{P}(B^c), \quad \mathbb{P}(A^c \cap B) = \mathbb{P}(A^c)\,\mathbb{P}(B) \quad \text{and} \quad \mathbb{P}(A^c \cap B^c) = \mathbb{P}(A^c)\,\mathbb{P}(B^c).$$

*Proof.* One shows here the first equality (noticing that the other two can be proved in a similar way):

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A \backslash (A \cap B)) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A)\,\mathbb{P}(B) = \mathbb{P}(A)\,(1 - \mathbb{P}(B)) = \mathbb{P}(A)\,\mathbb{P}(B^c).$$

□

For a collection of more than 2 events, the property $\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n)$ does not suffice to guarantee that the same property holds for complements of the events $A_i$. A slightly more involved definition of independence is therefore required.

**Definition 1.22.** Let $\{A_1, \ldots, A_n\}$ be a collection of events in $\mathcal{F}$. This collection is independent if

$$\mathbb{P}(A_1^* \cap \ldots \cap A_n^*) = \mathbb{P}(A_1^*) \cdots \mathbb{P}(A_n^*)$$

where $A_i^* =$ either $A_i$ or $A_i^c$, $i \in \{1, \ldots, n\}$.

An intuitive reason why complements should be included in the definition of independence is the following. Let us assume that one rolls a balanced die with four faces. Then the events {the outcome is 1 or 2} and {the outcome is even} are clearly independent; more precisely, the different *informations* associated with these events are. So the events {the outcome is 1 or 2} and {the outcome is odd} are also independent. This motivates the extension of the definition of independence to $\sigma$-fields in the next paragraph.

**Fact.** It can be shown that Definition 1.22 is equivalent to saying that

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i), \quad \forall I \subset \{1, \ldots, n\}.$$

From the above fact, one deduces that a collection of events might not be independent, even though its events are two-by-two independent.

**Independence of $\sigma$-fields.**

**Definition 1.23.** Let $\{\mathcal{G}_1, \ldots, \mathcal{G}_n\}$ be a collection of sub-$\sigma$-fields of $\mathcal{F}$. This collection is independent if

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n), \quad \forall A_1 \in \mathcal{G}_1, \ldots, A_n \in \mathcal{G}_n.$$

**Example.** Let again $\{A_1, \ldots, A_n\}$ be a collection of events in $\mathcal{F}$. Then the collection of events $\{A_1, \ldots, A_n\}$ is independent (according to Definition 1.22) if and only if the collection of $\sigma$-fields $\{\sigma(A_1), \ldots, \sigma(A_n)\}$ is independent (according to Definition 1.23). In order to see this, observe that $\sigma(A_i) = \{\emptyset, A_i, A_i^c, \Omega\}$.

**Independence of random variables.**

**Definition 1.24.** Let $\{X_1, \ldots, X_n\}$ be a collection of random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. This collection is independent if the collection of $\sigma$-fields $\{\sigma(X_1), \ldots, \sigma(X_n)\}$ is independent.

Since $\sigma(X_i) = \sigma(\{X_i \in B\}, B \in \mathcal{B}(\mathbb{R}))$, the collection $\{X_1, \ldots, X_n\}$ is independent if and only if

$$\mathbb{P}(\{X_1 \in B_1, \ldots, X_n \in B_n\}) = \mathbb{P}(\{X_1 \in B_1\}) \cdots \mathbb{P}(\{X_n \in B_n\}), \quad \forall B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R}).$$

But one also knows that $\sigma(X_i) = \sigma(\{X_i \leq t\}, t \in \mathbb{R})$, so it turns out that $\{X_1, \ldots, X_n\}$ is independent if and only if

$$\mathbb{P}(\{X_1 \leq t_1, \ldots, X_n \leq t_n\}) = \mathbb{P}(\{X_1 \leq t_1\}) \cdots \mathbb{P}(\{X_n \leq t_n\}), \quad \forall t_1, \ldots, t_n \in \mathbb{R}.$$

For discrete random variables taking values in a countable set $C$, this reduces to

$$\mathbb{P}(\{X_1 = x_1, \ldots, X_n = x_n\}) = \mathbb{P}(\{X_1 = x_1\}) \cdots \mathbb{P}(\{X_n = x_n\}), \quad \forall x_1, \ldots, x_n \in C.$$

And for jointly continuous random variables with joint pdf $f_{X_1,\ldots,X_n}$, this reduces to the classical relation

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n), \quad \forall x_1, \ldots, x_n \in \mathbb{R}.$$

The advantage of the above theoretical definition involving $\sigma$-fields is the following. Assume $\{X_1, \ldots, X_n\}$ is a collection of independent random variables and let $g_1, \ldots, g_n : \mathbb{R} \to \mathbb{R}$ be Borel-measurable functions. Then one directly deduces from the definition (and the fact that $g_i(X_i)$ is $\sigma(X_i)$-measurable) that $\{g_1(X_1), \ldots, g_n(X_n)\}$ is also a collection of independent random variables, which might have been cumbersome to check using any of the other "simpler" definition.

**Example.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a generic probability space and let $X_0(\omega) = c \in \mathbb{R}$, $\forall \omega \in \Omega$ be a constant random variable. As $\sigma(X_0) = \mathcal{F}_0 = \{\emptyset, \Omega\}$, $X_0$ is independent of any other random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$.

**Example.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space describing two independent dice rolls in Example 1.16 and let $X_1(i,j) = i$ and $X_2(i,j) = j$. One verifies below that these two random variables are indeed independent. It was already shown that $\mathbb{P}(\{X_1 = i\}) = \frac{1}{6}$, $\forall i \in \{1, \ldots, 6\}$. Likewise, $\mathbb{P}(\{X_2 = j\}) = \frac{1}{6}$, $\forall j \in \{1, \ldots, 6\}$ and

$$\mathbb{P}(\{X_1 = i, X_2 = j\}) = \mathbb{P}(\{(i,j)\}) = \frac{1}{36} = \mathbb{P}(\{X_1 = i\})\,\mathbb{P}(\{X_2 = j\}), \quad \forall (i,j) \in \Omega,$$

so $X_1$ and $X_2$ are independent.

## 1.6 Expectation

From the point of view of measure theory, random variables are maps from $\Omega$ to $\mathbb{R}$. Correspondingly, the *expectation* (or *mean*) of a random variable $X$ is the *Lebesgue integral* of the map $X$, that is, the "area under the curve $\omega \mapsto X(\omega)$", where the horizontal axis is measured with the probability measure $\mathbb{P}$.

**Definition.**

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X$ be a random variable defined on this probability space. The expectation of $X$, denoted as $\mathbb{E}(X)$, will be defined in three steps.

**Step 1.** Assume first that $X$ is a non-negative discrete random variable, i.e. that $X$ may be written as

$$X(\omega) = \sum_{i=i}^{\infty} x_i \, 1_{A_i}(\omega),$$

where $x_i \geq 0$ and $A_i \in \mathcal{F}$ (notice that if the $x_i$ are all different, then $A_i = \{X = x_i\}$). The expectation of $X$ is then defined as

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i \, \mathbb{P}(A_i),$$

which corresponds to the traditional definition of expectation in elementary probability courses. Notice here that since the sum is infinite, $\mathbb{E}(X)$ may take the value $+\infty$; but because of the assumption that $x_i \geq 0$, $\mathbb{E}(X)$ is always non-negative.

Notice also that in the particular case where $X = 1_A$, with $A \in \mathcal{F}$, one has $\mathbb{E}(X) = \mathbb{P}(A)$.

**Step 2.** Assume now that $X$ is a generic non-negative random variable (i.e. $X(\omega) \geq 0$, $\forall \omega \in \Omega$). Let us define the following sequence of discrete random variables:

$$X_n(\omega) = \sum_{i=1}^{\infty} \frac{i-1}{2^n} \, 1_{\{\frac{i-1}{2^n} < X \leq \frac{i}{2^n}\}}(\omega).$$

Notice that $x_i = \frac{i-1}{2^n} \geq 0$ and that $\{\frac{i-1}{2^n} < X \leq \frac{i}{2^n}\} \in \mathcal{F}$, since $X$ is $\mathcal{F}$-measurable. So according to Step 1, one has for each $n$

$$\mathbb{E}(X_n) = \sum_{i=1}^{\infty} \frac{i-1}{2^n} \, \mathbb{P}\left( \left\{ \frac{i-1}{2^n} < X \leq \frac{i}{2^n} \right\} \right) \in [0, +\infty].$$

It should be observed that $(X_n, n \in \mathbb{N})$ is actually an increasing sequence of non-negative "staircases", that is,

$$0 \leq X_n(\omega) \leq X_{n+1}(\omega), \quad \forall n.$$

As the size of the steps is divided by two from $n$ to $n+1$, the staircase gets refined. Likewise, one easily sees that $\mathbb{E}(X_n) \leq \mathbb{E}(X_{n+1})$ for all $n$, so $(\mathbb{E}(X_n), n \in \mathbb{N})$ is an increasing sequence, that therefore converges (possibly to $+\infty$). One defines

$$\mathbb{E}(X) = \lim_{n \to \infty} \mathbb{E}(X_n) = \lim_{n \to \infty} \sum_{i=1}^{\infty} \frac{i-1}{2^n} \, \mathbb{P}\left( \left\{ \frac{i-1}{2^n} < X \leq \frac{i}{2^n} \right\} \right) \in [0, \infty].$$

**Step 3.** Finally, consider a generic random variable $X$. One defines its *positive and negative parts*:

$$X^+(\omega) = \max(0, X(\omega)), \quad X^-(\omega) = \max(0, -X(\omega))$$

Notice that both $X^+(\omega) \geq 0$ and $X^-(\omega) \geq 0$, and that

$$X^+(\omega) - X^-(\omega) = X(\omega), \quad X^+(\omega) + X^-(\omega) = |X(\omega)|.$$

In measure theory, one does not want to deal with ill defined quantities such as $\infty - \infty$. One therefore defines $\mathbb{E}(X)$ only when $\mathbb{E}(|X|) = \mathbb{E}(X^+) + \mathbb{E}(X^-) < \infty$:

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-).$$

**Two important particular cases.** Let $X$ be a random variable and $g : \mathbb{R} \to \mathbb{R}$ be a Borel-measurable function such that $\mathbb{E}(|g(X)|) < \infty$ (this last condition is verified if for example $g$ is a bounded function).

- If $X$ is a discrete random variable with values in a countable set $C$, then

$$\mathbb{E}(g(X)) = \sum_{x \in C} g(x)\, \mathbb{P}(\{X = x\}).$$

- If $X$ is a continuous random variable with pdf $f_X$, then

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x)\, f_X(x)\, dx.$$

**Terminology.** - If $\mathbb{E}(|X|) < \infty$, then $X$ is said to be an *integrable* random variable.
- If $\mathbb{E}(X^2) < \infty$, then $X$ is said to be a *square-integrable* random variable.
- If there exists $c > 0$ such that $|X(\omega)| \leq c$, $\forall \omega \in \Omega$, then $X$ is said to be a *bounded* random variable.
- If $\mathbb{E}(X) = 0$, then $X$ is said to be a *centered* random variable.

One has the following series of implications:

$$\begin{array}{c} X \text{ is bounded} \quad \Rightarrow \quad X \text{ is square-integrable} \quad \Rightarrow \quad X \text{ is integrable,} \\ X \text{ is integrable and } Y \text{ is bounded} \quad \Rightarrow \quad XY \text{ is integrable,} \\ X, Y \text{ are both square-integrable} \quad \Rightarrow \quad XY \text{ is integrable.} \end{array}$$

**Negligible and almost sure sets.** An event $A \in \mathcal{F}$ is said to be negligible if $\mathbb{P}(A) = 0$. On the contrary, an event $B \in \mathcal{F}$ is said to be almost sure (a.s.) if $\mathbb{P}(B) = 1$. For example, if $\mathbb{P}(\{X \geq c\}) = 1$, one says that "$X \geq c$ almost surely".

**Basic properties of expectation.**

Linearity. If $c \in \mathbb{R}$ and $X$, $Y$ are integrable, then $\mathbb{E}(cX + Y) = c\,\mathbb{E}(X) + \mathbb{E}(Y)$.

Positivity. If $X$ is integrable and $X \geq 0$ a.s., then $\mathbb{E}(X) \geq 0$.

Strict positivity. If $X$ is integrable, $X \geq 0$ a.s. and $\mathbb{E}(X) = 0$, then $X = 0$ a.s.

Monotonicity. If $X$, $Y$ are integrable and $X \geq Y$ a.s., then $\mathbb{E}(X) \geq \mathbb{E}(Y)$.

**Inequalities.**

**Cauchy-Schwarz's inequality.** If $X$, $Y$ are square-integrable random variables, then the product $XY$ is integrable and

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

In particular, considering $Y = 1$ shows that if $X$ is square-integrable, then it is also integrable.

**Jensen's inequality.** If $X$ is a random variable and $\psi : \mathbb{R} \to \mathbb{R}$ is convex and such that $\mathbb{E}(|\psi(X)|) < \infty$, then

$$\psi(\mathbb{E}(X)) \leq \mathbb{E}(\psi(X)).$$

In particular, $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

Also, if $X$ is such that $\mathbb{P}(\{X = a\}) = \mathbb{P}(\{X = b\}) = 1/2$, then the above inequality says that

$$\psi\left(\frac{a+b}{2}\right) \leq \frac{\psi(a) + \psi(b)}{2},$$

which is pretty much the definition of convexity for $\psi$.

**Chebychev's inequality.** If $X$ is a random variable and $\varphi : \mathbb{R} \to \mathbb{R}_+$ is increasing on $\mathbb{R}_+$ and such that $\mathbb{E}(\varphi(X)) < \infty$, then for any $a > 0$, one has

$$\mathbb{P}(\{X \geq a\}) \leq \frac{\mathbb{E}(\varphi(X))}{\varphi(a)}.$$

In particular, if $X$ is square-integrable, then taking $\varphi(x) = x^2$ gives

$$\mathbb{P}(\{X \geq a\}) \leq \frac{\mathbb{E}(X^2)}{a^2}.$$

**Variance, covariance and independence.**

**Definition 1.25.** Let $X, Y$ be two square-integrable random variables. The *variance* of $X$ is defined as

$$\mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0$$

and the *covariance* of $X$ and $Y$ is defined as

$$\mathrm{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))\,(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y).$$

**Terminology.** If $\mathrm{Cov}(X, Y) = 0$, then $X$ and $Y$ are said to be *uncorrelated*.

**Fact.** If $X, Y$ are independent square-integrable random variables, then

a) $\mathrm{Cov}(X, Y) = 0$, i.e. $X$ and $Y$ are uncorrelated (but the reciprocal statement is wrong).

b) $\mathrm{Var}(cX + Y) = c^2\,\mathrm{Var}(X) + \mathrm{Var}(Y)$, for any $c \in \mathbb{R}$.

## 1.7 Convergence of sequences of random variables

For a given sequence of random variables $(X_n,\, n \geq 1)$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, there are several notions of convergence to a limiting random variable $X$. Let us review the most important ones.

**Convergence in probability.** The sequence $(X_n)$ is said to converge in probability to $X$ (and this is denoted as $X_n \xrightarrow{\mathbb{P}} X$) if for all $\varepsilon > 0$,

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

**Almost sure convergence.** The sequence $(X_n)$ is said to converge almost surely to $X$ (and this is denoted as $X_n \to X$ a.s.) if

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

**Fact.** Almost sure convergence implies convergence in probability, but the reverse implication is wrong. Nevertheless, it holds that $X_n \to X$ a.s. if for all $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}(\{|X_n - X| > \varepsilon\}) < \infty.$$

**Quadratic convergence.** Let us moreover assume that the random variables $X_n$ and $X$ are square-integrable. The sequence $(X_n)$ is then said to converge quadratically to $X$ if

$$\lim_{n\to\infty} \mathbb{E}(|X_n - X|^2) = 0.$$

**Fact.** By Chebychev's inequality, quadratic convergence implies convergence in probability (but not almost sure convergence).

**Convergence in distribution.** The sequence $(X_n)$ is said to converge in distribution to $X$ (and this is denoted as $X_n \xrightarrow{d} X$) if

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t),$$

for all $t \in \mathbb{R}$ which are continuity points of $F_X$.

**Remark.** For this last definition, the random variables $X_n$ need not be all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The knowledge of their respective distributions suffices.

**"Examples": limit theorems.**

**Weak law of large numbers (not the standard version).** Let $(\xi_n,\, n \geq 1)$ be a sequence of square-integrable and uncorrelated random variables with a common expectation $\mathbb{E}(\xi_n) = \mu$ and a common variance $\mathrm{Var}(\xi_n) = \sigma^2$. Let also $S_n = \xi_1 + \ldots + \xi_n$. Then

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu.$$

**Remark.** The convergence is also quadratic in this case.

**Strong law of large numbers.** Let $(\xi_n,\, n \geq 1)$ be a sequence independent and identically distributed (i.i.d.) random variables such that $\mathbb{E}(|\xi_1|) < \infty$. Let also $\mu = \mathbb{E}(\xi_1)$ and $S_n = \xi_1 + \ldots + \xi_n$. Then

$$\frac{S_n}{n} \to \mu \quad \text{a.s.}$$

**Example.** Assume that $\mathbb{P}(\{\xi_1 = 1\}) = \mathbb{P}(\{\xi_1 = 0\}) = 1/2$ (so $\mu = 1/2$). Then the above theorem says approximately that as $n$ gets large,

$$S_n \simeq \frac{n}{2} \quad \text{with high probability.}$$

The next question is: for a given $n$, how close is $S_n$ from $n/2$? The answer is given by the following theorem.

**Central limit theorem.** Let $(\xi_n)$ be a sequence of i.i.d. random variables such that $\mathbb{E}(\xi_1^2) < \infty$. Let also $\mu = \mathbb{E}(\xi_1)$, $\sigma^2 = \mathrm{Var}(\xi_1)$ and $S_n = \xi_1 + \ldots + \xi_n$. Then

$$\frac{S_n - n\,\mu}{\sqrt{n}\,\sigma} \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

This more specifically says that

$$\lim_{n\to\infty} \mathbb{P}\left(\left\{\frac{S_n - n\,\mu}{\sqrt{n}\,\sigma} \leq t\right\}\right) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx,$$

for all $t \in \mathbb{R}$ (as the cdf of $\mathcal{N}(0,1)$ is continuous on $\mathbb{R}$).

**Example.** Assume again that $\mathbb{P}(\{\xi_1 = 1\}) = \mathbb{P}(\{\xi_1 = 0\}) = 1/2$ (so $\mu = 1/2$ and $\sigma = 1/2$). Then the above theorem says approximately that as $n$ gets large,

$$S_n \simeq \frac{n}{2} + \frac{\sqrt{n}}{2}\, Z,$$

where $Z$ is a standard Gaussian random variable. So typically, the standard deviation of $S_n$ from its mean $n/2$ is of order $\sqrt{n}$.

## 1.8 Conditional expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, as usual.

**Conditioning with respect to an event $B \in \mathcal{F}$.**

The conditional probability of an event $A \in \mathcal{F}$ given another event $B \in \mathcal{F}$ is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{given that } \mathbb{P}(B) > 0.$$

In a similar way, the conditional expectation of an integrable random variable $X$ given $B$ is defined as

$$\mathbb{E}(X|B) = \frac{\mathbb{E}(X\, 1_B)}{\mathbb{P}(B)}, \quad \text{given that } \mathbb{P}(B) > 0.$$

**Conditioning with respect to a discrete random variable $Y$.**

Let us assume that the random variable $Y$ (is $\mathcal{F}$-measurable and) takes values in a countable set $C$.

$$\begin{aligned} \mathbb{P}(A|Y) &= \varphi(Y), \qquad \text{where } \varphi(y) = \mathbb{P}(A|\{Y = y\}), \quad y \in C. \\ \mathbb{E}(X|Y) &= \psi(Y), \qquad \text{where } \psi(y) = \mathbb{E}(X|\{Y = y\}), \quad y \in C. \end{aligned}$$

If $X$ is also a discrete random variable with values in $C$, then

$$\mathbb{E}(X|Y) = \psi(Y), \text{ where } \psi(y) = \frac{\mathbb{E}(X\, 1_{\{Y=y\}})}{\mathbb{P}(\{Y = y\})} = \sum_{x \in C} x\, \frac{\mathbb{E}(1_{\{X=x\} \cap \{Y=y\}})}{\mathbb{P}(\{Y = y\})} = \sum_{x \in C} x\, \mathbb{P}(\{X = x\}|\{Y = y\}).$$

**Important remark.** $\varphi(y)$ and $\psi(y)$ are regular functions, but $\mathbb{P}(A|Y)$ and $\mathbb{E}(X|Y)$ are *random variables*. They both are functions of the outcome of the random variable $Y$, that is, they are $\sigma(Y)$-measurable random variables.

**Example.** Let $X_1$, $X_2$ be two independent dice rolls and let us compute $\mathbb{E}(X_1 + X_2|X_2) = \psi(X_2)$, where

$$\begin{aligned} \psi(y) &= \mathbb{E}(X_1 + X_2|\{X_2 = y\}) = \frac{\mathbb{E}((X_1 + X_2)\, 1_{\{X_2=y\}})}{\mathbb{P}(\{X_2 = y\})} \\ &= \frac{\mathbb{E}(X_1\, 1_{\{X_2=y\}}) + \mathbb{E}(X_2\, 1_{\{X_2=y\}})}{\mathbb{P}(\{X_2 = y\})} \overset{(a)}{=} \frac{\mathbb{E}(X_1)\, \mathbb{E}(1_{\{X_2=y\}}) + \mathbb{E}(y\, 1_{\{X_2=y\}})}{\mathbb{P}(\{X_2 = y\})} \\ &= \frac{\mathbb{E}(X_1)\, \mathbb{P}(\{X_2 = y\}) + y\, \mathbb{P}(\{X_2 = y\})}{\mathbb{P}(\{X_2 = y\})} = \mathbb{E}(X_1) + y, \end{aligned}$$

where the independence assumption between $X_1$ and $X_2$ has been used in equality (a). So finally (as one would expect), $\mathbb{E}(X_1 + X_2|X_2) = \mathbb{E}(X_1) + X_2$, which can be explained intuitively as follows: the expectation of $X_1$ conditioned on $X_2$ is nothing but the expectation of $X_1$, as the outcome of $X_2$ provides no information on the outcome of $X_1$ ($X_1$ and $X_2$ being independent); on the other side, the expectation of $X_2$ conditioned on $X_2$ is exactly $X_2$, as the outcome of $X_2$ is known.

**Conditioning with respect to a continuous random variable $Y$?**

In this case, one faces the following problem: if $Y$ is a continuous random variable, $\mathbb{P}(\{Y = y\}) = 0$ for all $y \in \mathbb{R}$. So a direct generalization of the above formulas to the continuous case is impossible at first sight. A possible solution to this problem is to replace the event $\{Y = y\}$ by $\{y \leq Y < y + \varepsilon\}$ and take the limit $\varepsilon \to 0$ for the definition of conditional expectation. This actually works, but also leads to a paradox in the multidimensional setting (known as Borel's paradox). In addition, some random variables are neither discrete, nor continuous. It turns out that the cleanest way to define conditional expectation in the general case is through $\sigma$-fields.

**Conditioning with respect to a sub-$\sigma$-field $\mathcal{G}$.**

In order to define the conditional expectation in the general case, one needs the following proposition.

**Proposition 1.26.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G}$ be a sub-$\sigma$-field of $\mathcal{F}$ and $X$ be an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. There exists then an integrable random variable $Z$ such that

(i) $Z$ is $\mathcal{G}$-measurable,

(ii) $\mathbb{E}(ZU) = \mathbb{E}(XU)$ for any random variable $U$ $\mathcal{G}$-measurable and bounded.

Moreover, if $Z_1$, $Z_2$ are two integrable random variables satisfying (i) and (ii), then $Z_1 = Z_2$ a.s.

**Definition 1.27.** The above random variable $Z$ is called the *conditional expectation of $X$ given $\mathcal{G}$*. It is defined up to a negligible set.

**Notation.** $Z$ is denoted as $\mathbb{E}(X|\mathcal{G})$.

One further *defines* $\mathbb{P}(A|\mathcal{G}) = \mathbb{E}(1_A|\mathcal{G})$ for $A \in \mathcal{F}$.

**Remark.** Notice that as before, both $\mathbb{P}(A|\mathcal{G})$ and $\mathbb{E}(X|\mathcal{G})$ are random variables.

**Properties.**

The above definition does not give a computation rule for the conditional expectation; it is only an existence theorem. The properties listed below will therefore be of help for computing conditional expectations.

- Linearity. $\mathbb{E}(c\,X + Y|\mathcal{G}) = c\,\mathbb{E}(X|\mathcal{G}) + \mathbb{E}(Y|\mathcal{G})$ a.s.

- Monotonicity. If $X \geq Y$ a.s., then $\mathbb{E}(X|\mathcal{G}) \geq \mathbb{E}(Y|\mathcal{G})$ a.s.

- $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$.

- If $X$ is independent of $\mathcal{G}$, then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$ a.s.

- If $X$ is $\mathcal{G}$-measurable, then $\mathbb{E}(X|\mathcal{G}) = X$ a.s.

- If $Y$ is $\mathcal{G}$-measurable and bounded, then $\mathbb{E}(XY|\mathcal{G}) = \mathbb{E}(X|\mathcal{G})\,Y$ a.s.

- If $\mathcal{H}$ is a sub-$\sigma$-field of $\mathcal{G}$, then $\mathbb{E}(\mathbb{E}(X|\mathcal{H})|\mathcal{G}) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(X|\mathcal{H})$ a.s.

Some of these properties are illustrated below with an example.

**Example.** Let $\Omega = \{1, \ldots, 6\}$, $\mathcal{F} = \mathbb{P}(\Omega)$ and $\mathbb{P}(\{\omega\}) = \frac{1}{6}$ for $\omega = 1, \ldots, 6$ (the probability space of the die roll). Let also $X(\omega) = \omega$ be the outcome of the die roll and consider the two sub-$\sigma$-fields:

$$\mathcal{G} = \sigma(\{1,3\}, \{2\}, \{5\}, \{4,6\}) \quad \text{and} \quad \mathcal{H} = \sigma(\{1,3,5\}, \{2,4,6\}).$$

Then $\mathbb{E}(X) = 3.5$,

$$\mathbb{E}(X|\mathcal{G})(\omega) = \begin{cases} 2 & \text{if } \omega \in \{1,3\} \text{ or } \omega = 2 \\ 5 & \text{if } \omega \in \{4,6\} \text{ or } \omega = 5 \end{cases} \quad \text{and} \quad \mathbb{E}(X|\mathcal{H})(\omega) = \begin{cases} 3 & \text{if } \omega \in \{1,3,5\} \\ 4 & \text{if } \omega \in \{2,4,6\} \end{cases}$$

So $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(\mathbb{E}(X|\mathcal{H})) = \mathbb{E}(X)$. Moreover,

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H})(\omega) = \begin{cases} \frac{1}{3}(2+2+5) = 3 & \text{if } \omega \in \{1,3,5\} \\ \frac{1}{3}(2+5+5) = 4 & \text{if } \omega \in \{2,4,6\} \end{cases} = \mathbb{E}(X|\mathcal{H})(\omega)$$

and

$$\mathbb{E}(\mathbb{E}(X|\mathcal{H})|\mathcal{G})(\omega) = \begin{cases} 3 & \text{if } \omega \in \{1,3\} \text{ or } \omega = 5 \\ 4 & \text{if } \omega \in \{4,6\} \text{ or } \omega = 2 \end{cases} = \mathbb{E}(X|\mathcal{H})(\omega).$$

On other words, the smallest $\sigma$-field always "wins".

**Proposition 1.28.** Let $\mathcal{G}$ be a sub-$\sigma$-field of $\mathcal{F}$, $X, Y$ be two random variables such that $X$ is independent of $\mathcal{G}$ and $Y$ is $\mathcal{G}$-measurable, an let $\varphi : \mathbb{R}^2 \to \mathbb{R}$ be a Borel-measurable function such that $\mathbb{E}(|\varphi(X,Y)|) < \infty$. Then

$$\mathbb{E}(\varphi(X,Y)|\mathcal{G}) = \psi(Y) \quad a.s., \quad \text{where } \psi(y) = \mathbb{E}(\varphi(X,y)).$$

This proposition has the following consequence: when computing the expectation of a function $\varphi$ of two independent random variables $X$ and $Y$, one can always divide the computation in two steps by writing

$$\mathbb{E}(\varphi(X,Y)) = \mathbb{E}(\mathbb{E}(\varphi(X,Y)|Y)) = \mathbb{E}(\psi(Y))$$

where $\psi(y) = \mathbb{E}(\varphi(X,y))$ (this is actually nothing but Fubini's theorem).

Finally, the proposition below shows that Jensen's inequality also holds for conditional expectation.

**Proposition 1.29.** Let $X$ be a random variable, $\mathcal{G}$ be a sub-$\sigma$-field of $\mathcal{F}$ and $\psi : \mathbb{R} \to \mathbb{R}$ be convex and such that $\mathbb{E}(|\psi(X)|) < \infty$. Then

$$\psi(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}(\psi(X)|\mathcal{G}) \quad a.s.$$

In particular, $|\mathbb{E}(X|\mathcal{G})| \leq \mathbb{E}(|X||\mathcal{G})$ a.s.

**Conditioning with respect to a generic random variable $Y$.**

Once the definition of conditional expectation with respect to a $\sigma$-field is set, it is natural to define for a generic random variable $Y$:

$$\mathbb{E}(X|Y) = \mathbb{E}(X|\sigma(Y)) \quad \text{and} \quad \mathbb{P}(A|\mathcal{G}) = \mathbb{P}(A|\sigma(Y)).$$

**Remark.** Since any $\sigma(Y)$-measurable random variable may be written as $g(Y)$, where $g$ is a Borel-measurable function, the definition of $\mathbb{E}(X|Y)$ may be rephrased as follows.

**Definition 1.30.** $E(X|Y) = \psi(Y)$, where $\psi : \mathbb{R} \to \mathbb{R}$ is the unique Borel-measurable function such that $\mathbb{E}(\psi(Y)\, g(Y)) = \mathbb{E}(X\, g(Y))$ for any function $g : \mathbb{R} \to \mathbb{R}$ Borel-measurable and bounded.

In two particular cases, the function $\psi$ can be made explicit.

- As already seen above, if $X$, $Y$ are two discrete random variables with values in a countable set $C$, then

$$E(X|Y) = \psi(Y), \quad \text{where} \quad \psi(y) = \sum_{x \in D} x\, \mathbb{P}(\{X = x\}|\{Y = y\}), \quad y \in C.$$

- If $X, Y$ are two jointly continuous random variables with joint pdf $f_{X,Y}$, then

$$E(X|Y) = \psi(Y), \quad \text{where} \quad \psi(y) = \int_{\mathbb{R}} x\, \frac{f_{X,Y}(x,y)}{f_Y(y)}\, dy, \quad y \in \mathbb{R},$$

and $f_Y$ is the marginal pdf of $Y$ given by $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\, dy$, assumed here to be strictly positive. Let us check that the random variable $\psi(Y)$ is indeed the conditional expectation of $X$ given $Y$ according to Definition 1.30: for any function $g : \mathbb{R} \to \mathbb{R}$ Borel-measurable and bounded, one has

$$\begin{aligned}
\mathbb{E}(\psi(Y)\, g(Y)) &= \int_{\mathbb{R}} \psi(y)\, g(y)\, f_Y(y)\, dy \\
&= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} x\, \frac{f_{X,Y}(x,y)}{f_Y(y)}\, dy \right) g(y)\, f_Y(y)\, dy \\
&= \iint_{\mathbb{R}^2} x\, g(y)\, f_{X,Y}(x,y)\, dx\, dy = \mathbb{E}(X\, g(Y)).
\end{aligned}$$

## 1.9 Random vectors

**Preliminary.** - The Borel $\sigma$-field on $\mathbb{R}^n$ is defined as

$$\mathcal{B}(\mathbb{R}^n) = \sigma\left(\{]a_1, b_1[\times \ldots \times ]a_n, b_n[ :\ a_i < b_i,\ \forall i\}\right)$$

It contains nearly all possible subsets of $\mathbb{R}^n$ (not only rectangles!).

- The Lebesgue measure on $\mathbb{R}^n$ is defined as

$$|\,]a_1, b_1[\times \ldots \times ]a_n, b_n[\,| = \prod_{i=1}^{n}(b_i - a_i)$$

and can be extended by $\sigma$-additivity to any Borel subset of $\mathbb{R}^n$.

Let now $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

**Definition 1.31.** A *random vector of dimension* $n \geq 1$ is a map $X : \begin{cases} \Omega \to \mathbb{R}^n \\ \omega \mapsto X(\omega) = (X_1(\omega), \ldots, X_n(\omega)) \end{cases}$ such that

$$\{\omega \in \Omega\ :\ X(\omega) \in B\} \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbb{R}^n).$$

**Proposition 1.32.** - $X = (X_1, \ldots, X_n)$ is a random vector if and only if

$$\{\omega \in \Omega\ :\ X_1(\omega) \leq t_1, \ldots, X_n(\omega) \leq t_n\} \in \mathcal{F}, \quad \forall t_1, \ldots, t_n \in \mathbb{R}.$$

- If $X$ is a random vector, then each component $X_i$ is a random variable, but the reciprocal statement is wrong.

**Two important classes of random vectors.**

**Discrete random vectors.**

**Definition 1.33.** $X$ is a discrete random vector if it takes values in a countable subset $C$ of $\mathbb{R}^n$.

If $g : \mathbb{R}^n \to \mathbb{R}$ is a Borel-measurable and bounded function, then

$$\mathbb{E}(g(X_1, \ldots, X_n)) = \sum_{x_1, \ldots, x_n \in C} g(x_1, \ldots, x_n)\, \mathbb{P}(\{X_1 = x_1, \ldots, X_n = x_n\}).$$

**Proposition 1.34.** $X$ is a discrete random vector if and only if each $X_i$, $1 \leq i \leq n$, is a discrete random variable.

**Continuous random vectors.**

**Definition 1.35.** $X$ is a *continuous random vector* if $\mathbb{P}(\{X \in B\}) = 0$ for every $B \in \mathcal{B}(\mathbb{R}^n)$ such that $|B| = 0$.

**Fact.** If $X$ is a continuous random vector, then there exists a Borel-measurable function $f_X : \mathbb{R}^n \to \mathbb{R}$ such that

$$f_X(x_1, \ldots, x_n) \geq 0, \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n, \quad \int_{\mathbb{R}^n} f_X(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1$$

and

$$\mathbb{P}\left(\{X \in B\}\right) = \int_B f_X(x_1, \ldots, x_n)\, dx_1 \cdots dx_n, \quad \forall B \in \mathcal{B}(\mathbb{R}^n).$$

**Terminology.** $f_X$ is called the joint probability density function of the random vector $X$.

If $g : \mathbb{R}^n \to \mathbb{R}$ is a Borel-measurable and bounded function then

$$\mathbb{E}(g(X_1, \ldots, X_n)) = \int_{\mathbb{R}^n} g(x_1, \ldots, x_n)\, f_X(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.$$

**Proposition 1.36.** If $X$ is a continuous random vector, then each $X_i$, $1 \leq i \leq n$, is a continuous random variable, but the reciprocal statement is wrong.

Here is a counter-example. Let $Y$ be a continuous random variable and $X = (Y, Y)$; then $X$ is not a continuous random vector (indeed, let $\Delta = \{(x, y) \in \mathbb{R}^2 \ : \ x = y\}$: $|\Delta| = 0$, but $\mathbb{P}(\{X \in \Delta\}) = 1$).

**Expectation and covariance of random vector.**

Let $X$ be an $n$-dimensional random vector such that each component $X_i$, $1 \leq i \leq n$, is square-integrable.

**Expectation (or mean) of $X$ :** $\mathbb{E}(X) = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_n))$, $n$-variate vector.

**Covariance (matrix) of $X$ :** $\mathrm{Cov}(X) = K$, $n \times n$ matrix, where

$$K_{ij} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\,\mathbb{E}(X_j).$$

**Properties.** - $K$ is symmetric, i.e. $K_{ij} = K_{ji}$.

- $K$ is positive semi-definite, i.e. $\forall c_1, \ldots c_n \in \mathbb{R}$, $\sum_{i,j=1}^n c_i\, c_j\, K_{ij} \geq 0$. Indeed,

$$\sum_{i,j=1}^n c_i\, c_j\, K_{ij} = \sum_{i,j=1}^n c_i\, c_j\, \mathrm{Cov}(X_i, X_j) = \mathrm{Cov}\left(\sum_{i=1}^n c_i\, X_i, \sum_{j=1}^n c_j\, X_j\right) = \mathrm{Var}\left(\sum_{i=1}^n c_i\, X_i\right) \geq 0.$$

**Remark.** If $K$ is a symmetric $n \times n$ matrix, then $K$ is positive semi-definite if and only if all its eigenvalues are non-negative.

**Proposition 1.37.** If $X_1, \ldots, X_n$ are independent and square-integrable then $X = (X_1, \ldots, X_n)$ is a random vector and $\mathrm{Cov}(X)$ is a diagonal matrix (i.e. $\mathrm{Cov}(X_i, X_j) = 0$, $\forall i \neq j$).

**Gaussian random vectors.**

**Convention.** If $Y(\omega) = c$, $\forall \omega \in \Omega$, then $Y$ is said to be a Gaussian random variable with mean $c$ and variance 0 ($Y \sim \mathcal{N}(c, 0)$).

**Definition 1.38.** A random vector $X = (X_1, \ldots, X_n)$ is Gaussian if $\forall c_1, \ldots, c_n \in \mathbb{R}$, $c_1 X_1 + \ldots + c_n X_n$ is a Gaussian random variable (possibly with variance 0).

**Remark.** This is more than saying that every $X_i$, $1 \leq i \leq n$, is Gaussian (see below)!

**Proposition 1.39.** If $X_1, \ldots, X_n$ are independent Gaussian random variables, then $X = (X_1, \ldots X_n)$ is a Gaussian random vector.

**Proposition 1.40.** Let $X = (X_1, \ldots, X_n)$ be a Gaussian random vector. Then the random variables $X_1, \ldots, X_n$ are independent if and only if $\mathrm{Cov}(X)$ is a diagonal matrix.

**But.** If $X_1, \ldots, X_n$ are Gaussian random variables, then it is not necessarily true that $X = (X_1, \ldots, X_n)$ is a Gaussian random vector. Also $\mathrm{Cov}(X_i, X_j) = 0$, $\forall i \neq j$ does not imply in general that $X_1, \ldots, X_n$ are independent!

**Remark.** If $X$ is a Gaussian random vector with mean $m$ and covariance $K$, this is denoted as $X \sim \mathcal{N}(m, K)$. Moreover, $X$ is entirely characterized by its mean $m$ and its covariance $K$.

Let $X$ be an $n$-dimensional Gaussian random vector with mean $m = \mathbb{E}(X)$ and covariance $K = \mathrm{Cov}(X)$.

**Definition 1.41.** $X$ is *non-degenerate* if $\mathrm{rank}(K) = n$.

**Reminder.** Let $K$ be an $n \times n$ symmetric matrix.

- $\mathrm{rank}(K) = n$ if and only if $K$ is invertible if and only if $\det(K) \neq 0$ if and only if all its eigenvalues are non-zero.

- More generally, $\mathrm{rank}(K)$ = number of non-zero eigenvalues of $K$.

**Proposition 1.42.** Let $X$ be a non-degenerate $n$-dimensional Gaussian random vector, with mean $m$ and covariance $K$. Then $X$ is a continuous random vector with joint pdf

$$f_X(x_1,\dots,x_n) = \frac{1}{\sqrt{(2\pi)^n \det(K)}} \exp\left(-\frac{1}{2}\sum_{i,j=1}^{n}(x_i-m_i)\,(K^{-1})_{ij}\,(x_j-m_j)\right), \quad (x_1,\dots,x_n)\in\mathbb{R}^n.$$

**Proposition 1.43.** Let $X$ be an $n$-dimensional Gaussian random vector with mean 0 and covariance $K$. Let also $k = \mathrm{rank}(K) \in \{0,\dots,n\}$. Then there exist $k$ i.i.d. random variables $U_1,\dots,U_k \sim \mathcal{N}(0,1)$ and $\alpha_{ij}\in\mathbb{R}$ $(1\le i\le n,\ 1\le j\le k)$ such that $X_i = \sum_{j=1}^{k}\alpha_{ij}\,U_j$, $i=1,\dots,n$.

**Remark.** In matrix form, $X = AU$ and $AA^T = K$.

**Example.** Let $X=(Y,Y)$, where $Y\sim\mathcal{N}(0,1)$. In this simple case, we have

$$\mathrm{Cov}(X) = K = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

so $k=1$, and $U_1 = Y$, $\alpha_{11}=\alpha_{21}=1$.

# 2 Discrete-time stochastic processes

A discrete-time stochastic process can be viewed

a) either as a collection of random variables $(X_n,\ n\in\mathbb{N})$

b) or as a random sequence $X : \left\{ \begin{array}{ccc} \Omega & \to & \mathbb{R}^{\mathbb{N}} \\ \omega & \mapsto & (X_n(\omega),\ n\in\mathbb{N}) \end{array} \right.$

**Canonical example: the random walk.**

Let $(\xi_n,\ n\ge 1)$ be a collection of i.i.d. random variables such that $\mathbb{P}(\{\xi_1=+1\}) = \mathbb{P}(\{\xi_1=-1\}) = 1/2$. Let $S_0=0$, $S_n = \xi_1+\ldots+\xi_n$, $n\ge 1$: the process $(S_n,\ n\in\mathbb{N})$ is called the simple symmetric random walk. As simple as it may be, this process exhibits already many fascinating properties.

**Remarks.** - $\mathbb{E}(\xi_1)=0$, so $\mathbb{E}(S_n) = \mathbb{E}(\xi_1)+\ldots+\mathbb{E}(\xi_n) = 0$.

- $\mathrm{Var}(\xi_1)=1$, so by independence, $\mathrm{Var}(S_n) = \mathrm{Var}(\xi_1)+\ldots+\mathrm{Var}(\xi_n) = n$.

- By the (strong) law of large numbers, we know that $\dfrac{S_n}{n} \underset{n\to\infty}{\to} 0$ a.s.

- By the central limit theorem, we also know that $\dfrac{S_n}{\sqrt{n}} \xrightarrow{d} Z\sim\mathcal{N}(0,1)$, i.e. $S_n \simeq \sqrt{n}\,Z$.

**Variations.**

- simple asymmetric random walk: $S_n = \xi_1+\ldots+\xi_n$, with $\mathbb{P}(\{\xi_1=+1\}) = p = 1-\mathbb{P}(\{\xi_1=-1\})$ and $p\neq 1/2$.

- random walk with values in $\mathbb{R}$ : $S_n = \xi_1+\ldots+\xi_n$ with $\xi_1\sim\mathcal{N}(0,1)$ e.g.

- continuous-time random walk = Brownian motion (see next chapter).

## 2.1 Martingales

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

**Definition 2.1.** A *filtration* is a sequence $(\mathcal{F}_n,\, n \in \mathbb{N})$ of sub-$\sigma$-fields of $\mathcal{F}$ such that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, $\forall n \in \mathbb{N}$.

**Example.** Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $X_n(\omega) = n^{th}$ decimal of $\omega$, for $n \geq 1$. Let also $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Then $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, $\forall n \in \mathbb{N}$.

**Definitions 2.2.** - A discrete-time process $(X_n,\, n \in \mathbb{N})$ is said to be *adapted* to the filtration $(\mathcal{F}_n,\, n \in \mathbb{N})$ if $X_n$ is $\mathcal{F}_n$-measurable $\forall n \in \mathbb{N}$.

- The *natural filtration* of a process $(X_n,\, n \in \mathbb{N})$ is defined as $\mathcal{F}_n^X = \sigma(X_0, \ldots, X_n)$, $n \in \mathbb{N}$. It represents the available amount of information about the process at time $n$.

**Remark.** A process is adapted to its natural filtration, by definition.

Let now $(\mathcal{F}_n,\, n \in \mathbb{N})$ be a given filtration.

**Definition 2.3.** A discrete-time process $(M_n,\, n \in \mathbb{N})$ is a *martingale* with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ if

(i) $\mathbb{E}(|M_n|) < \infty$, $\forall n \in \mathbb{N}$.

(ii) $M_n$ is $\mathcal{F}_n$-measurable, $\forall n \in \mathbb{N}$ (i.e., $(M_n,\, n \in \mathbb{N})$ is adapted to $(\mathcal{F}_n,\, n \in \mathbb{N})$).

(iii) $\mathbb{E}(M_{n+1}|\mathcal{F}_n) = M_n$ a.s., $\forall n \in \mathbb{N}$.

A martingale is therefore a fair game: the expectation of the process at time $n+1$ given the information at time $n$ is equal to the value of the process at time $n$.

**Remark.** Conditions (ii) and (iii) are actually redundant, as (iii) implies (ii).

**Properties.** If $(M_n,\, n \in \mathbb{N})$ is a martingale, then

- $\mathbb{E}(M_{n+1}) = \mathbb{E}(M_n)\, (= \ldots = \mathbb{E}(M_0))$, $\forall n \in \mathbb{N}$.

- $\mathbb{E}(M_{n+1} - M_n|\mathcal{F}_n) = 0$ a.s.

- $\mathbb{E}(M_{n+m}|\mathcal{F}_n) = M_n$ a.s., $\forall n, m \in \mathbb{N}$.

This last property is important, as it says that the martingale property propagates over time.

**Example: the simple symmetric random walk.**

Let $(S_n,\, n \in \mathbb{N})$ be the simple symmetric random walk : $S_0 = 0$, $S_n = \xi_1 + \ldots + \xi_n$, where the $\xi_n$ are i.i.d. and $\mathbb{P}(\{\xi_1 = +1\}) = \mathbb{P}(\{\xi_1 = -1\}) = 1/2$.

Let us define the following filtration: $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma(\{\xi_1, \ldots, \xi_n\})$, $n \geq 1$. Then $(S_n,\, n \in \mathbb{N})$ is a martingale with respect to $(\mathcal{F}_n, , n \in \mathbb{N})$. Indeed:

(i) $\mathbb{E}(|S_n|) \leq \mathbb{E}(|\xi_1|) + \ldots + \mathbb{E}(|\xi_n|) = 1 + \ldots + 1 = n < \infty$, $\forall n \in \mathbb{N}$.

(ii) $S_n = \xi_1 + \ldots + \xi_n$ is a function of $(\xi_1, \ldots, \xi_n)$, i.e., is $\sigma(\xi_1, \ldots, \xi_n) = \mathcal{F}_n$-measurable.

(iii) We have

$$\begin{aligned} \mathbb{E}(S_{n+1}|\mathcal{F}_n) &= \mathbb{E}(S_n + \xi_{n+1}|\mathcal{F}_n) = \mathbb{E}(S_n|\mathcal{F}_n) + \mathbb{E}(\xi_{n+1}|\mathcal{F}_n) \\ &= S_n + \mathbb{E}(\xi_{n+1}) = S_n + 0 = S_n \quad \text{a.s.} \end{aligned}$$

The first equality on the second line follows from the fact that $S_n$ is $\mathcal{F}_n$-measurable and that $\xi_{n+1}$ is independent of $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$.

**Generalization.** If the random variables $\xi_n$ are i.i.d. and such that $\mathbb{E}(|\xi_1|) < \infty$ and $\mathbb{E}(\xi_1) = 0$, then $(S_n,\, n \in \mathbb{N})$ is also a martingale (in particular, $\xi_1 \sim \mathcal{N}(0,1)$ works).

**Definition 2.4.** Let $(\mathcal{F}_n,\, n \in \mathbb{N})$ be a filtration. A process $(M_n,\, n \in \mathbb{N})$ is a *submartingale* (resp. a *supermartingale*) with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ if

(i) $\mathbb{E}(|M_n|) < \infty$, $\forall n \in \mathbb{N}$.

(ii) $M_n$ is $\mathcal{F}_n$-measurable, $\forall n \in \mathbb{N}$.

(iii) $\mathbb{E}(M_{n+1}|\mathcal{F}_n) \geq M_n$ a.s., $\forall n \in \mathbb{N}$ (resp. $\mathbb{E}(M_{n+1}|\mathcal{F}_n) \leq M_n$ a.s., $\forall n \in \mathbb{N}$).

**Remarks.** - Not every process is either a sub- or a supermartingale!

- The appellations sub- and supermartingale are counter-intuitive. They are due to historical reasons.

- Condition (ii) is now necessary in itself, as (iii) does not imply it.

- If $(M_n,\, n \in \mathbb{N})$ is both a submartingale and a supermartingale, then it is a martingale.

**Example: the simple asymmetric random walk.**

- If $\mathbb{P}(\{\xi_1 = +1\}) = p = 1 - \mathbb{P}(\{\xi_1 = -1\})$ with $p \geq 1/2$, then $S_n = \xi_1 + \ldots + \xi_n$ is a submartingale.

- More generally, $S_n = \xi_1 + \ldots + \xi_n$ is a submartingale if $\mathbb{E}(\xi_1) \geq 0$.

**Proposition 2.5.** If $(M_n,\, n \in \mathbb{N})$ is a martingale with respect to a filtration $(\mathcal{F}_n,\, n \in \mathbb{N})$ and $\varphi : \mathbb{R} \to \mathbb{R}$ is a Borel-measurable and convex function such that $\mathbb{E}(|\varphi(M_n)|) < \infty$, $\forall n \in \mathbb{N}$, then $(\varphi(M_n),\, n \in \mathbb{N})$ is a submartingale.

*Proof.* (i) $\mathbb{E}(|\varphi(M_n)|) < \infty$ by assumption.

(ii) $\varphi(M_n)$ is $\mathcal{F}_n$-measurable as $M_n$ is (and $\varphi$ is Borel-measurable).

(iii) $\mathbb{E}(\varphi(M_{n+1})|\mathcal{F}_n) \geq \varphi(\mathbb{E}(M_{n+1}|\mathcal{F}_n)) = \varphi(M_n)$ a.s.

In (iii), the first inequality follows from Jensen's inequality and the second follows from the fact that $M$ is a martingale. □

**Example.** If $(M_n,\, n \in \mathbb{N})$ is a square-integrable martingale (i.e., $\mathbb{E}(M_n^2) < \infty$, $\forall n \in \mathbb{N}$), then the process $(M_n^2,\, n \in \mathbb{N})$ is a submartingale (as $x \mapsto x^2$ is convex).

**Doob's Decomposition theorem.**

**Definition 2.6.** A process $(A_n,\, n \in \mathbb{N})$ is said to be *predictable* with respect to a filtration $(\mathcal{F}_n,\, n \in \mathbb{N})$ if $A_0 = 0$ and $A_n$ is $\mathcal{F}_{n-1}$-measurable $\forall n \geq 1$.

**Remark.** If a process is predictable, then it is adapted.

**Theorem 2.7.** Let $(X_n,\, n \in \mathbb{N})$ be a submartingale with respect to a filtration $(\mathcal{F}_n,\, n \in \mathbb{N})$. Then there exists a martingale $(M_n,\, n \in \mathbb{N})$ with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ and a process $(A_n,\, n \in \mathbb{N})$ predictable with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ and increasing (i.e, $A_n \leq A_{n+1}$ $\forall n \in \mathbb{N}$) such that $A_0 = 0$ and $X_n = M_n + A_n$, $\forall n \in \mathbb{N}$. Moreover, this decomposition of the process $X$ is unique.

*Proof.* (main idea)
$\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$, so a natural candidate for the process $A$ is to set $A_0 = 0$ and $A_{n+1} = A_n + \mathbb{E}(X_{n+1}|\mathcal{F}_n) - X_n\, (\geq A_n)$, which is a predictable and increasing process. Then, $M_0 = X_0$ and $M_{n+1} - M_n = X_{n+1} - X_n - (A_{n+1} - A_n) = X_{n+1} - \mathbb{E}(X_{n+1}|\mathcal{F}_n)$ is indeed a martingale, as $\mathbb{E}(M_{n+1} - M_n|\mathcal{F}_n) = 0$. □

## 2.2 Stopping times

**Definitions 2.8.** - A *random time* is a random variable $T$ with values in $\mathbb{N} \cup \{+\infty\}$.

- Given a process $(X_n,\, n \in \mathbb{N})$, one defines $X_T(\omega) = X_{T(\omega)}(\omega) = \sum_{n\in\mathbb{N}} X_n(\omega)\, 1_{\{T=n\}}(\omega)$.

- A *stopping time* with respect to a filtration $(\mathcal{F}_n, n \in \mathbb{N})$ is a random time $T$ such that $\{T \leq n\} \in \mathcal{F}_n$, $\forall n \in \mathbb{N}$.

**Proposition 2.9.** $T$ is a stopping time with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ if and only if $\{T = n\} \in \mathcal{F}_n$, $\forall n \in \mathbb{N}$.

**Example.** Let $(X_n,\, n \in \mathbb{N})$ be a process adapted to $(\mathcal{F}_n,\, n \in \mathbb{N})$ and $a > 0$. Then, $T_a = \inf\{n \in \mathbb{N} : |X_n| \geq a\}$ is a stopping time with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$. Indeed:

$$\begin{aligned}\{T_a = n\} &= \{|X_i| < a,\ \forall 0 \leq i \leq n-1 \text{ and } |X_n| \geq a\} \\ &= \bigcap_{i=0}^{n-1} \underbrace{\{|X_i| < a\}}_{\in \mathcal{F}_i \subset \mathcal{F}_n\ \forall i=0,\ldots,n-1} \quad \cap \quad \{|X_n| \geq a\} \in \mathcal{F}_n, \quad \forall n \in \mathbb{N}.\end{aligned}$$

**Definition 2.10.** Let $T$ be a stopping time with respect to a filtration $(\mathcal{F}_n,\, n \in \mathbb{N})$. One defines the information one possesses at time $T$ as the following $\sigma$-field:

$$\mathcal{F}_T = \{A \in \mathcal{F} \,:\, A \cap \{T = n\} \in \mathcal{F}_n,\ \forall n \in \mathbb{N}\}.$$

**Facts.**

- If $T(\omega) = N$, $\forall \omega \in \Omega$, then $\mathcal{F}_T = \mathcal{F}_N$.

- If $T_1(\omega) \leq T_2(\omega)$, $\forall \omega \in \Omega$, then $\mathcal{F}_{T_1} \subset \mathcal{F}_{T_2}$.

[Here is an example of stopping times $T_1$, $T_2$ such that $T_1 \leq T_2$:
let $0 < a < b$ and consider $T_1 = \inf\{n \in \mathbb{N} : |X_n| \geq a\}$ and $T_2 = \inf\{n \in \mathbb{N} : |X_n| \geq b\}$.]

- A random variable $Y$ is $\mathcal{F}_T$-measurable if and only if $Y\, 1_{\{T=n\}}$ is $\mathcal{F}_n$-measurable, $\forall n \in \mathbb{N}$.

As a consequence:

- If $(X_n,\, n \in \mathbb{N})$ is adapted to $(\mathcal{F}_n,\, n \in \mathbb{N})$, then $X_T$ is $\mathcal{F}_T$-measurable.

**Doob's optional sampling theorem.**
Let $(M_n,\, n \in \mathbb{N})$ be a martingale with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ and $T_1$, $T_2$ be two stopping times such that $0 \leq T_1(\omega) \leq T_2(\omega) \leq N < \infty$, $\forall \omega \in \Omega$. Then

$$\mathbb{E}(M_{T_2} | \mathcal{F}_{T_1}) = M_{T_1} \text{ a.s.}$$

In particular, $\mathbb{E}(M_{T_2}) = \mathbb{E}(M_{T_1})$ (this consequence is referred to as the *optional stopping* theorem).

In particular, if $T$ is a stopping time such that $0 \leq T(\omega) \leq N < \infty$, $\forall \omega \in \Omega$, then

$$\mathbb{E}(M_T) = \mathbb{E}(M_0).$$

**Remarks.** - The above theorem says that the martingale property holds even if one is given the option to stop at any (bounded) stopping time.

- The theorem also holds for sub- and supermartingales.

*Proof.* - We first show that if $T$ is a stopping time such that $0 \leq T(\omega) \leq N$, then $\mathbb{E}(M_N | \mathcal{F}_T) = M_T$ $(*)$:

Indeed, let $Z = M_T = \sum_{n=0}^{N} M_n\, 1_{\{T=n\}}$. We check below that $Z$ is the conditional expectation of $M_N$ given $\mathcal{F}_T$:

(i) $Z$ is $\mathcal{F}_T$-measurable: $Z\,1_{\{T=n\}} = M_n\,1_{\{T=n\}}$, so by the above mentioned fact, $Z$ is $\mathcal{F}_T$-measurable.

(ii) $\mathbb{E}(ZU) = \mathbb{E}(M_N U)$, $\forall U$ $\mathcal{F}_T$-measurable and bounded:

$$\mathbb{E}(ZU) = \sum_{n=0}^{N} \mathbb{E}(M_n 1_{\{T=n\}} U) = \sum_{n=0}^{N} \mathbb{E}(\mathbb{E}(M_N|\mathcal{F}_n) \underbrace{1_{\{T=n\}} U}_{\mathcal{F}_n-\text{measurable}}) = \sum_{n=0}^{N} \mathbb{E}(M_N 1_{\{T=n\}} U) = \mathbb{E}(M_N U).$$

- Second, let us check that $\mathbb{E}(M_{T_2}|\mathcal{F}_{T_1}) = M_{T_1}$:

$$M_{T_1} \underset{(*) \text{ with } T=T_1}{=} \mathbb{E}(M_N|\mathcal{F}_{T_1}) \underset{\mathcal{F}_{T_1} \subset \mathcal{F}_{T_2}}{=} \mathbb{E}(\mathbb{E}(M_N|\mathcal{F}_{T_2})|\mathcal{F}_{T_1}) \underset{(*) \text{ with } T=T_2}{=} \mathbb{E}(M_{T_2}|\mathcal{F}_{T_1}).$$

This concludes the proof of the theorem. □

## 2.3 Martingale transforms

Let $(\mathcal{F}_n,\, n \in \mathbb{N})$ be a filtration, $(H_n,\, n \in \mathbb{N})$ be a predictable process with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ and $(M_n,\, n \in \mathbb{N})$ be a martingale with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$.

**Definition 2.11.** The process $G$ defined as

$$G_0 = 0, \quad G_n = (H \cdot M)_n = \sum_{i=1}^{n} H_i(M_i - M_{i-1}), \quad n \geq 1,$$

is called the *martingale transform* of $M$ through $H$.

**Remark.** This process is the discrete version of the stochastic integral. It represents the gain obtained by applying the strategy $H$ to the game $M$:

- $H_i$ = amount bet on day $i$ ($\mathcal{F}_{i-1}$-measurable)

- $M_i - M_{i-1}$ = increment of the process $M$ on day $i$.

- $G_n$ = gain on day $n$.

**Proposition 2.12.** If $H_n$ is a bounded random variable for each $n$ (i.e, $|H_n(\omega)| \leq K_n$ $\forall \omega \in \Omega$), then the process $G$ is a martingale with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$.

In other words, one cannot win on a martingale!

*Proof.* (i) $\mathbb{E}(|G_n|) \leq \sum_{i=1}^{n} \mathbb{E}(|H_i|\,|M_i - M_{i-1}|) \leq \sum_{i=1}^{n} K_i\,(\mathbb{E}(|M_i|) + \mathbb{E}(|M_{i-1}|)) < \infty$.

(ii) $G_n$ is $\mathcal{F}_n$-measurable by construction.

(iii) $\mathbb{E}(G_{n+1}|\mathcal{F}_n) = \mathbb{E}(G_n + H_{n+1}\,(M_{n+1} - M_n)|\mathcal{F}_n) = G_n + H_{n+1}\,\mathbb{E}(M_{n+1} - M_n|\mathcal{F}_n) = G_n + 0 = G_n$. □

**Example: "the" martingale.**

Let $(M_n,\, n \in \mathbb{N})$ be the simple symmetric random walk ($M_n = \xi_1 + \ldots + \xi_n$) and consider the following strategy:

$$H_0 = 0,\, H_1 = 1,\, H_{n+1} = \begin{cases} 2H_n, & \text{if } \xi_1 = \ldots = \xi_n = -1, \\ 0. & \text{otherwise.} \end{cases}$$

Notice that all the $H_n$ are bounded random variables. Then by the above proposition, the process $G$ defined as

$$G_0 = 0, \quad G_n = \sum_{i=1}^{n} H_i\,(M_i - M_{i-1}) = \sum_{i=1}^{n} H_i\,\xi_i, \quad n \geq 1,$$

is a martingale. So $\mathbb{E}(G_n) = \mathbb{E}(G_0) = 0$, $\forall n \in \mathbb{N}$. Let now

$$T = \inf\{n \geq 1 \,:\, \xi_n = +1\}.$$

$T$ is a stopping time and it is easily seen that $G_T = +1$. But then $\mathbb{E}(G_T) = 1 \neq 0 = \mathbb{E}(G_0)$? Is there a contradiction? Actually no. The optional stopping theorem does not apply here, because the time $T$ is unbounded: $\mathbb{P}(T = n) = 2^{-n}$, $\forall n \in \mathbb{N}$, i.e., there does not exist $N$ fixed such that $T(\omega) \leq N$, $\forall \omega \in \Omega$.

## 2.4 Markov processes

Let $(X_n,\, n \in \mathbb{N})$ be a discrete-time process adapted to a filtration $(\mathcal{F}_n,\, n \in \mathbb{N})$.

**Definition 2.13.** $(X_n,\, n \in \mathbb{N})$ is said to be a *Markov process* with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ if

$$\mathbb{P}(X_{n+1} \in B \,|\, \mathcal{F}_n) = \mathbb{P}(X_{n+1} \in B \,|\, X_n), \quad \forall n \in \mathbb{N},\, B \in \mathcal{B}(\mathbb{R}).$$

(remember that $\mathbb{P}(X_{n+1} \in B \,|\, X_n) = \mathbb{P}(X_{n+1} \in B \,|\, \sigma(X_n))$ by definition.)

**Proposition 2.14.** $(X_n,\, n \in \mathbb{N})$ is a Markov process with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ if and only if

$$\mathbb{E}(g(X_{n+1})|\mathcal{F}_n) = \mathbb{E}(g(X_{n+1})|X_n), \quad \forall n \in \mathbb{N}$$

and $\forall g : \mathbb{R} \to \mathbb{R}$ Borel-measurable and bounded.

**Particular class of Markov process.** Let $(\xi_n,\, n \geq 1)$ be a sequence of independent random variables (not necessarily i.i.d.) and

$$\mathcal{F}_0 = \{\emptyset,\, \Omega\}, \quad \mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n).$$

Then any process defined recursively as

$$X_0 = \text{cst}, \quad X_{n+1} = f(X_n, \xi_{n+1}), \quad n \geq 0,$$

with $f : \mathbb{R}^2 \to \mathbb{R}$ a Borel-measurable function, is a Markov process with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$.

*Proof.* As $X_n$ is both and $\mathcal{F}_n$-measurable and $\sigma(X_n)$-measurable, and $\xi_{n+1}$ is independent of $\mathcal{F}_n$, and therefore of $X_n$, we have by Proposition 1.28:

$$\mathbb{E}(g(X_{n+1})|\mathcal{F}_n) = \mathbb{E}(g(f(X_n, \xi_{n+1}))|\mathcal{F}_n) = \psi(X_n), \quad \text{where } \psi(y) = \mathbb{E}(g(f(y, \xi_{n+1})))$$

and also

$$\mathbb{E}(g(X_{n+1})|X_n) = \mathbb{E}(g(f(X_n, \xi_{n+1}))|X_n) = \psi(X_n).$$

□

**Example.** Let $S$ be a "generalized" random walk with independent increments, i.e. $S_0 = 0$, $S_n = \xi_1 + \ldots + \xi_n$, $n \geq 1$, where the random variables $\xi_i$ are independent. Then $S_{n+1} = S_n + \xi_{n+1}$ is indeed a Borel-measurable function of $S_n$ and $\xi_{n+1}$; it falls therefore into the above-mentioned class of Markov processes.

**Remarks.**

- It is important not to mix martingale property ($\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$) with the Markov property ($\mathbb{E}(g(X_{n+1})|\mathcal{F}_n) = \mathbb{E}(g(X_{n+1})|X_n)$, $\forall g$). None of the two implies the other one.

- Markov chains form a particular class of Markov processes (those with discrete state space).

- Stationarity is not included in the above definition of a Markov process.

# 3 Continuous-time stochastic processes

**Definition 3.1.** A *continuous-time stochastic process* is a collection of random variables $(X_t,\, t \in \mathbb{R}_+)$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Alternatively, a stochastic process may be seen as a random function

$$X : \left\{ \begin{array}{ccc} \Omega & \to & \{f : \mathbb{R}_+ \to \mathbb{R}\} \\ \omega & \mapsto & \{t \mapsto X_t(\omega)\} \end{array} \right.$$

**Remark.** In order to describe a continuous-time stochastic process, one generally needs a LARGE probability space $\Omega$!

**Question.** For a single random variable $X$, the knowledge of its cdf $\mathbb{P}(X \leq x)$, $\forall x \in \mathbb{R}$ characterizes entirely the random variable. In the case of a stochastic process $(X_t,\, t \in \mathbb{R}_+)$, what is needed in order to characterize the process entirely?

**First answer.** Specify $\mathbb{P}(X_t \leq x)$, $\forall t \in \mathbb{R}_+$, $\forall x \in \mathbb{R}$ ? This is insufficient. Here is why: assume we only know that $X_t \sim \mathcal{N}(0,t)$, $\forall t \in \mathbb{R}_+$. Let us then define

- $X_t^{(1)} = \sqrt{t}\, Y$, where $Y \sim \mathcal{N}(0,1)$.

- $X_t^{(2)}$ = standard Brownian motion (defined below).

It turns out that these two processes satisfy both $X_t^{(1)} \sim \mathcal{N}(0,t)$ and $X_t^{(2)} \sim \mathcal{N}(0,t)$, $\forall t \in \mathbb{R}_+$, even though they have little to do with each other!

**Second answer.** Specify $\mathbb{P}(X_{t_1} \leq x_1, X_{t_2} \leq x_2)$, $\forall t_1, t_2 \in \mathbb{R}_+$, $x_1, x_2 \in \mathbb{R}$? This is better, but still insufficient! (Actually, it is sufficient for Gaussian processes: see below).

$n^{th}$ **answer.** Specify $\mathbb{P}(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n)$, $\forall t_1, \ldots, t_n \in \mathbb{R}_+$ *and* $n \geq 1$! This is the correct answer. Specifying all these joint distributions is cumbersome in general, but we will focus our attention on specific classes of processes for which a simpler description is possible.

### Processes with independent and stationary increments

**Definition 3.2.** The random variables $X_t - X_s$, for $t \geq s \geq 0$, are called the *increments* of the process $X = (X_t,\, t \in \mathbb{R}_+)$.

**Definition 3.3.** A process $X = (X_t,\, t \in \mathbb{R}_+)$ is said to have *independent and stationary increments* if

- $X_t - X_s \perp\!\!\!\perp \mathcal{F}_s^X = \sigma(X_r,\, 0 \leq r \leq s)$, $\forall t \geq s \geq 0$ (independence).

- $X_t - X_s \sim X_{t-s} - X_0$, $\forall t \geq s \geq 0$ (stationarity).

(Remember that $X \sim Y$ means "$X$ has the same distribution as $Y$").

For such process, it is sufficient to specify the distribution of $X_0$ and $X_t - X_0$, $\forall t \in \mathbb{R}_+$, in order to fully characterize the process. So in some sense in this case, the first answer above is valid. But having independent and stationary increments is a strong requirement for a continuous-time process.

### Processes with continuous trajectories

**Definition 3.4.** A process $X = (X_t,\, t \in \mathbb{R}_+)$ is said to have *continuous trajectories* if

$$\mathbb{P}(\{\omega \in \Omega :\ \text{the function } t \mapsto X_t(\omega) \text{ is continuous }\}) = 1.$$

We now have all the concepts in our hands in order to define the standard Brownian motion, which exhibits many interesting properties and plays a central role in the theory of stochastic calculus.

## 3.1 Standard Brownian motion

**Definition 3.5.** (first version) A *standard Brownian motion* is a continuous-time stochastic process $B = (B_t,\, t \in \mathbb{R}_+)$ such that

- $B_0 = 0$ a.s.

- $B$ has independent and stationary increments.

- $B_t \sim \mathcal{N}(0,t)$, $\forall t \in \mathbb{R}_+$.

- $B$ has continuous trajectories.

**Basic properties.** - $\mathbb{E}(B_t) = 0$, $\mathbb{E}(B_t^2) = t$, $\forall t \in \mathbb{R}_+$.

- $B_t - B_s \sim B_{t-s} - B_0 = B_{t-s} \sim \mathcal{N}(0, t-s)$, so $\mathbb{E}(B_t - B_s) = 0$, $\mathbb{E}((B_t - B_s)^2) = t - s$, $\forall t \geq s \geq 0$.

- By the law of large numbers, $\lim_{t\to\infty} \frac{B_t}{t} = 0$ a.s.

- Moreover, $\frac{B_t}{\sqrt{t}} \sim \mathcal{N}(0,1)$, $\forall t \geq 0$, so the central limit theorem applies trivially here: $\frac{B_t}{\sqrt{t}} \underset{n\to\infty}{\overset{d}{\to}} Z \sim \mathcal{N}(0,1)$, i.e.

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{B_t}{\sqrt{t}} \leq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2)\, dy, \quad \forall x \in \mathbb{R}.$$

**Remarks.** - These properties are reminiscent from those of the random walk.

- The existence of a process $B$ that satisfies all the above properties is ensured by a deep and important theorem of Kolmogorov, but we shall not state it explicitly here.

**Construction from the random walk.**

- Let $(S_n,\, n \in \mathbb{N})$ be the simple symmetric random walk (i.e $S_0 = 0$, $S_n = \xi_1 + \ldots + \xi_n$ with $\xi_i$ i.i.d., $\mathbb{P}(\xi_1 = 1) = \mathbb{P}(\xi_1 = -1) = \frac{1}{2}$). Remember that by the central limit theorem, $\frac{S_n}{\sqrt{n}} \underset{n\to\infty}{\overset{d}{\to}} Z \sim \mathcal{N}(0,1)$.

- Let now

$$Y_t = S_{[t]} + (t - [t])\, \xi_{[t]+1},\, t \in \mathbb{R}_+, \quad \text{i.e., if } t = n + \varepsilon,\, \varepsilon \in [0,1], \quad \text{then } Y_t = S_n + \varepsilon\, \xi_{n+1}.$$

This process is known as the *broken line process*.

**Remark.** $Y$ is not a process with independent increments, nor is it a standard Brownian motion!

- Let us define $B_t^{(n)} = \frac{Y_{nt}}{\sqrt{n}}$, $t \in \mathbb{R}_+$: this amounts to looking at the process $Y$ from far away, rescaling the $x$-axis by a factor $n$, while rescaling the $y$-axis by a factor $\sqrt{n}$. Assume now for simplicity that $nt \in \mathbb{N}$. Then

$$B_t^{(n)} = \frac{S_{nt}}{\sqrt{n}} = \sqrt{t}\, \frac{S_{nt}}{\sqrt{nt}} \underset{n\to\infty}{\overset{d}{\to}} \sqrt{t}\, Z \sim \mathcal{N}(0,t) \quad \text{i.e.} \quad \mathbb{P}(B_t^{(n)} \leq x) \underset{n\to\infty}{\to} \mathbb{P}(B_t \leq x).$$

as $B_t \sim \mathcal{N}(0,t)$.

- Similarly, one can show that

$$\mathbb{P}(B_{t_1}^{(n)} \leq x_1, \ldots, B_{t_m}^{(n)} \leq x_m) \underset{n\to\infty}{\to} \mathbb{P}(B_{t_1} \leq x_1, \ldots, B_{t_m} \leq x_m),$$

$\forall t_1, \ldots, t_m \in \mathbb{R}_+$, $x_1, \ldots, x_m \in \mathbb{R}$ and $m \geq 1$. This shows that the sequence of processes $B^{(n)}$ converges in distribution to the process $B$.

**Remark.** From this, we deduce that even though the limiting process $B$ has continuous trajectories, these are nowhere differentiable. Indeed, the slope of $B_t^{(n)}$ is $\pm\sqrt{n}$, so the "slope" of $B_t$ is $\pm\infty$. The derivative of $B_t$ is formally called the *white noise* process (although this process does not exist!).

## 3.2 Mean and covariance

Let $X = (X_t,\, t \in \mathbb{R}_+)$ be a square integrable (i.e. $\mathbb{E}(X_t^2) < \infty,\, \forall t \in \mathbb{R}_+$) continuous-time process.

**Definition 3.6.** - The *mean* of the process $X$ is the function $m : \mathbb{R}_+ \to \mathbb{R}$ given by $m(t) = \mathbb{E}(X_t)$, $t \in \mathbb{R}_+$.

- The *covariance* of the process $X$ is the function $K : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ given by $K(t,s) = \mathrm{Cov}(X_t, X_s)$, $t, s \in \mathbb{R}_+$.

**Properties.** - $K$ is symmetric, i.e. $K(t,s) = K(s,t)$.

- $K$ is positive semi-definite, i.e.

$$\sum_{i,j=1}^{n} c_i\, c_j\, K(t_i, t_j) \geq 0, \quad \forall c_1, \ldots, c_n \in \mathbb{R}, \quad t_1, \ldots t_n \in \mathbb{R}_+ \quad \text{and} \quad n \geq 1.$$

The proof of this follows the same lines as the proof for the covariance of a random vector.

In general, the mean $m$ and the covariance $K$ alone do not fully characterize a process $X$ (as it is the case for random variables and random vectors). The only exception is given in the following paragraph.

## 3.3 Gaussian processes

**Definition 3.7.** A *Gaussian process* is a process $(X_t,\, t \in \mathbb{R}_+)$ such that $c_1 X_{t_1} + \ldots + c_n X_{t_n}$ is a Gaussian random variable $\forall c_1, \ldots c_n \in \mathbb{R}$, $t_1, \ldots, t_n \in \mathbb{R}_+$ and $n \geq 1$.

In other words, the process $X$ is a Gaussian process if and only if each sample $(X_{t_1}, \ldots, X_{t_n})$ is a Gaussian vector.

**Theorem 3.8.** (Kolmogorov) Given $m : \mathbb{R}_+ \to \mathbb{R}$ and $K : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ symmetric and positive semi-definite, there exists a Gaussian process $X$ with mean $m$ and covariance $K$. In addition, $m$ and $K$ characterize entirely the process $X$.

**Proposition 3.9.** (second possible definition of the standard Brownian motion)
The standard Brownian motion $B = (B_t,\, t \in \mathbb{R}_+)$ is a Gaussian process with continuous trajectories, with mean $m(t) = 0$ and covariance $K(t,s) = t \wedge s$ $(= \min(t,s))$.

*Proof.* (that the first definition implies the second)
- One should first check that $c_1 B_{t_1} + \ldots + c_n B_{t_n}$ is a Gaussian random variable $\forall c_1, \ldots, c_n \in \mathbb{R}$, $t_1, \ldots, t_n \in \mathbb{R}_+$ and $n \geq 1$. Let us simply check that $B_t + B_s$ is Gaussian $\forall t \geq s \geq 0$:

$$B_t + B_s = B_t - B_s + 2B_s \quad \text{is Gaussian,}$$

as $B_t - B_s$ and $2B_s$ are independent and Gaussian. The proof in the general case follows the same idea.

- $m(t) = \mathbb{E}(B_t) = 0$.

- Let $t \geq s \geq 0$ :

$$\begin{aligned} K(t,s) &= \mathbb{E}(B_t B_s) = \mathbb{E}((B_t - B_s + B_s)B_s) = \mathbb{E}((B_t - B_s)B_s) + \mathbb{E}(B_s^2) \\ &= \mathbb{E}(B_t - B_s)\,\mathbb{E}(B_s) + \mathbb{E}(B_s^2) = 0 + \mathbb{E}(B_s^2) = s = \min(t,s). \end{aligned}$$

□

## 3.4 Markov processes

**Definition 3.10.** A (continuous-time) *Markov process* with respect to a filtration $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$ is a process $(X_t,\, t \in \mathbb{R}_+)$ such that

$$\mathbb{P}(X_t \in B \,|\, \mathcal{F}_s) = \mathbb{P}(X_t \in B \,|\, X_s) \qquad \forall t \geq s \geq 0,\, \forall B \in \mathcal{B}(\mathbb{R}).$$

Equivalently,

$$\mathbb{E}(g(X_t) \,|\, \mathcal{F}_s) = \mathbb{E}(g(X_t) \,|\, X_s) \qquad \forall t \geq s \geq 0$$

and $g : \mathbb{R} \to \mathbb{R}$ Borel-measurable and bounded.

**Proposition 3.11.** The standard Brownian motion is a Markov process with respect to its natural filtration $\mathcal{F}_s^B = \sigma(B_r,\, 0 \leq r \leq s)$.

*Proof.* - $\mathbb{E}(g(B_t) \,|\, \mathcal{F}_s^B) = \mathbb{E}(g(B_t - B_s + B_s) \,|\, \mathcal{F}_s) = \psi(B_s)$, where $\psi(y) = \mathbb{E}(g(B_t - B_s + y))$ (this follows from the fact that $B_t - B_s \perp\!\!\!\perp \mathcal{F}_s$ and that $B_s$ is $\mathcal{F}_s$-measurable).

- Similarly, $\mathbb{E}(g(B_t) \,|\, B_s) = \mathbb{E}(g(B_t - B_s + B_s) \,|\, B_s) = \psi(B_s)$ given above. □

**Remark.** More generally, any process with independent increments (but not necessarily stationary) is a Markov process with respect to its natural filtration.

## 3.5 Martingales

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

**Definitions 3.12.** - A (continuous-time) *filtration* is a collection $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$ of sub-$\sigma$-fields of $\mathcal{F}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$, $\forall t \geq s \geq 0$.

- A process $(X_t,\, t \in \mathbb{R}_+)$ is said to be *adapted* to the filtration $(\mathcal{F}_t, t \in \mathbb{R}_+)$ if $X_t$ is $\mathcal{F}_t$-measurable $\forall t \in \mathbb{R}_+$.

- The *natural filtration* of a process $(X_t,\, t \in \mathbb{R}_+)$ is defined as $\mathcal{F}_t^X = \sigma(X_s,\, 0 \leq s \leq t)$, $t \in \mathbb{R}_+$.

**Remark.** Every process is adapted to its natural filtration.

**Definition 3.13.** A process $(M_t,\, t \in \mathbb{R}_+)$ is said to be a (continuous-time) *martingale* with respect to a filtration $(\mathcal{F}_t, t \in \mathbb{R}_+)$ if

(i) $\mathbb{E}(|M_t|) < \infty$, $\forall t \in \mathbb{R}_+$.

(ii) $M_t$ is $\mathcal{F}_t$-measurable, $\forall t \in \mathbb{R}_+$.

(iii) $\mathbb{E}(M_t \,|\, \mathcal{F}_s) = M_s$, $\forall t \geq s \geq 0$.

**Generalization.** The process $M$ is said to be a *submartingale* (respectively a *supermartingale*) if condition (iii) is replaced by $\mathbb{E}(M_t \,|\, \mathcal{F}_s) \geq M_s$ (respectively $\mathbb{E}(M_t \,|\, \mathcal{F}_s) \leq M_s$), $\forall t \geq s \geq 0$.

**Proposition 3.14.** If $(M_t,\, t \in \mathbb{R}_+)$ is a martingale and $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and such that $\mathbb{E}(|\varphi(M_t)|) < \infty$ for all $t \in \mathbb{R}_+$, then the process $(\varphi(M_t),\, t \in \mathbb{R}_+)$ is a submartingale.

**Proposition 3.15.** The standard Brownian motion $(B_t,\, t \in \mathbb{R}_+)$ is a martingale with respect to its natural filtration $(\mathcal{F}_t^B,\, t \in \mathbb{R}_+)$.

*Proof.* (i) $\mathbb{E}(|B_t|) \leq \sqrt{\mathbb{E}(B_t^2)} = \sqrt{t} < \infty$, $\forall t \in \mathbb{R}_+$.

(ii) $B_t$ is $FC_t^B$-measurable, by definition, $\forall t \in \mathbb{R}_+$.

(iii) Let $t \geq s \geq 0$:

$$\mathbb{E}(B_t \,|\, \mathcal{F}_s^B) = \mathbb{E}(B_t - B_s + B_s \,|\, \mathcal{F}_s^B) = \mathbb{E}(B_t - B_s \,|\, \mathcal{F}_s^B) + \mathbb{E}(B_s \,|\, \mathcal{F}_s^B) = \mathbb{E}(B_t - B_s) + B_s = 0 + B_s = B_s.$$

□

**Proposition 3.16.** The following processes are also martingales with respect to $(\mathcal{F}_t^B,\, t \in \mathbb{R}_+)$:

- $(M_t = B_t^2 - t,\, t \in \mathbb{R}_+)$.

- $(N_t = \exp(B_t - \frac{t}{2}),\, t \in \mathbb{R}_+)$.

**Theorem 3.17.** (Lévy) (third possible definition of the standard Brownian motion)

Let $(X_t,\, t \in \mathbb{R}_+)$ be a process with continuous trajectories, adapted to a filtration $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$ and such that $X_0 = 0$ a.s. and

(i) $(X_t,\, t \in \mathbb{R}_+)$ is a martingale with respect to $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$.

(ii) $(X_t^2 - t,\, t \in \mathbb{R}_+)$ is also a martingale with respect to $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$.

Then $(X_t,\, t \in \mathbb{R}_+)$ is a standard Brownian motion.

**Definitions 3.18.** - A *stopping time* with respect to a filtration $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$ is a random time $T$ with values in $\mathbb{R}_+ \cup \{+\infty\}$ such that $\{T \leq t\} \in \mathcal{F}_t,\, \forall t \in \mathbb{R}_+$.

- If $X$ is a process, then $X_T(\omega) = X_{T(\omega)}(\omega),\, \omega \in \Omega$ (process evaluated at time $T$).

- $\mathcal{F}_T = \{A \in \mathcal{F} \,:\, A \cap \{T \leq t\} \in \mathcal{F}_t,\ \forall t \in \mathbb{R}_+\}$ (information one possesses at time $T$).

**Doob's optional sampling theorem.**
Let $(M_t,\, t \in \mathbb{R}_+)$ be a martingale with respect to a filtration $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$, with continuous trajectories (from now on, we will also say equivalently: a *continuous* martingale). Let $T_1$, $T_2$ be two stopping times such that $0 \leq T_1(\omega) \leq T_2(\omega) \leq K < \infty,\, \forall \omega \in \Omega$. Then $\mathbb{E}(M_{T_2} \,|\, \mathcal{F}_{T_1}) = M_{T_1}$ a.s. In particular, $\mathbb{E}(M_{T_2}) = \mathbb{E}(M_{T_1})$ (optional stopping).

**Remarks.** - The proof of the theorem is much more involved than in the discrete-time setting.

- The theorem remains valid for sub- and supermartingales (with corresponding inequalities).

**Doob's inequalities.**
Let $(M_t,\, t \in \mathbb{R}_+)$ be continuous square-integrable martingale with respect to $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$ such that $M_0 = 0$ a.s. Then

a) $\mathbb{P}(\sup_{0 \leq s \leq t} |M_s| \geq \lambda) \leq \frac{\mathbb{E}(|M_t|)}{\lambda},\, \forall t > 0,\, \lambda > 0.$

b) $\mathbb{E}(\sup_{0 \leq s \leq t} |M_s|^2) \leq 4\mathbb{E}(|M_t|^2),\, \forall t > 0.$

**Doob's decomposition theorem.**
Let $(X_t,\, t \in \mathbb{R}_+)$ be a continuous submartingale with respect to a filtration $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$. Then there exists a unique process $(A_t,\, t \in \mathbb{R}_+)$ which is increasing (i.e. $A_s \leq A_t$ if $s \leq t$), continuous and adapted to $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$ such that $A_0 = 0$ and $(X_t - A_t,\, t \in \mathbb{R}_+)$ is a martingale with respect to $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$.

**Application.** Let $(M_t,\, t \in \mathbb{R}_+)$ be a continuous square-integrable martingale with respect to $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$. Then there exists a unique process $(A_t,\, t \in \mathbb{R}_+)$ which is increasing, continuous and adapted to $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$ such that $A_0 = 0$ and $(M_t^2 - A_t,\, t \in \mathbb{R}_+)$ is a martingale with respect to $(\mathcal{F}_t,\, t \in \mathbb{R}_+)$.

This process will play a particular role in the following.

**Examples.** - If $M_t = B_t$, then $A_t = t$ (indeed, $B_t^2 - t$ =martingale)

- If $M$ has independent increments, then $A_t = \mathbb{E}(M_t^2) - \mathbb{E}(M_0^2)$.

# 4 Stochastic integral

## 4.1 Functions with bounded variation

**Definition 4.1.** A function $g : \mathbb{R}_+ \to \mathbb{R}$ is said to have *bounded variation* if $\forall t > 0$,

$$\sup \sum_{i=1}^{n} |g(t_i) - g(t_{i-1})| < \infty,$$

where the supremum is taken over all partitions $0 = t_0 < t_1 < \ldots < t_n = t$ of $[0, t]$ (and $n$ is arbitrary).

**Examples.** - If $g$ is increasing (or decreasing), then $g$ has bounded variation. Indeed, in this case:

$$\sum_{i=1}^{n} |g(t_i) - g(t_{i-1})| = \sum_{i=1}^{n} g(t_i) - g(t_{i-1}) = g(t_n) - g(t_0) = g(t) - g(0)$$

for all partitions of $[0, t]$, so

$$\sup \sum_{i=1}^{n} |g(t_i) - g(t_{i-1})| = g(t) - g(0) < \infty.$$

- If $g = g_1 - g_2$, where $g_1$ and $g_2$ are both increasing, then $g$ also has bounded variation.

- If $g$ is continuously differentiable, then $g$ has bounded variation. Indeed,

$$\sum_{i=1}^{n} |g(t_i) - g(t_{i-1})| = \sum_{i=1}^{n} \left| \int_{t_{i-1}}^{t_i} g'(s)\, ds \right| \leq \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} |g'(s)|\, ds = \int_0^t |g'(s)|\, ds < \infty.$$

Again, this expression does not depend on the chosen partition, so

$$\sup \sum_{i=1}^{n} |g(t_i) - g(t_{i-1})| \leq \int_0^t |g'(s)|\, ds < \infty.$$

**Generalization to processes.**

**Definition 4.2.** A continuous-time stochastic process $(X_t,\, t \in \mathbb{R}_+)$ is said to have bounded variation if its trajectories have bounded variation a.s.

We will see that the trajectories of the standard Brownian motion have *unbounded variation*, a.s.

## 4.2 Quadratic variation

**Quadratic variation of the standard Brownian motion.**

Let $(B_t,\, t \in \mathbb{R}_+)$ be a standard Brownian motion. For $t > 0$ and $n \geq 1$ fixed, let

$$\langle B \rangle_t^{(n)} = \sum_{i=1}^{2^n} \left( B\left(\frac{it}{2^n}\right) - B\left(\frac{(i-1)t}{2^n}\right) \right)^2.$$

**Notation.** We use indifferently the notation $B_t \equiv B(t)$.

**Definition 4.3.** The (almost sure) limit $\langle B \rangle_t = \lim_{n\to\infty} \langle B \rangle_t^{(n)}$ is called the *quadratic variation* of the Brownian motion. We show below that it exists and is equal to $t$.

**Proposition 4.4.** For every fixed $t \geq 0$, $\langle B \rangle_t = t$, a.s.

*Proof.* Recall that in order to show that $Z_n \to Z$ a.s., it is sufficient to check that

$$\sum_{n \geq 1} \mathbb{P}(\{|Z_N - Z| > \varepsilon\}) < \infty, \quad \forall \varepsilon > 0.$$

Here, $Z_n = \langle B \rangle_t^{(n)}$ and $Z = t$, which is fixed. Let us first compute $\mathbb{E}(\langle B \rangle_t^{(n)})$ and $\mathrm{Var}(\langle B \rangle_t^{(n)})$:

$$\mathbb{E}(\langle B \rangle_t^{(n)}) = \sum_{i=1}^{2^n} \mathbb{E}\Big( \underbrace{\Big( B\left(\tfrac{it}{2^n}\right) - B\left(\tfrac{(i-1)t}{2^n}\right) \Big)}_{\sim \mathcal{N}(0, \frac{t}{2^n})}^2 \Big) = \sum_{i=1}^{2^n} \frac{t}{2^n} = t$$

and

$$\mathrm{Var}(\langle B \rangle_t^{(n)}) = \sum_{i=1}^{2^n} \mathrm{Var}\Big( \Big( B\left(\tfrac{it}{2^n}\right) - B\left(\tfrac{(i-1)t}{2^n}\right) \Big)^2 \Big)$$

by independence of the increments of $B$. Moreover, if $X \sim \mathcal{N}(0, \sigma^2)$, then

$$\mathrm{Var}(X^2) = \mathbb{E}(X^4) - \mathbb{E}(X^2)^2 = 3\sigma^4 - \sigma^4 = 2\sigma^4, \quad \text{so} \quad \mathrm{Var}(\langle B \rangle_t^{(n)}) = \sum_{i=1}^{2^n} 2\left(\tfrac{t}{2^n}\right)^2 = \frac{t^2}{2^{n-1}}.$$

Therefore, by Chebychev's inequality,

$$\mathbb{P}(\{|\langle B \rangle_t^{(n)} - t| > \varepsilon\}) \leq \frac{1}{\varepsilon^2}\, \mathbb{E}((\langle B \rangle_t^{(n)} - t)^2) = \frac{1}{\varepsilon^2}\, \mathrm{Var}(\langle B \rangle_t^{(n)}) = \frac{t^2}{\varepsilon^2\, 2^{n-1}}$$

and

$$\sum_{n \geq 1} \mathbb{P}(\{|\langle B \rangle_t^{(n)} - t| > \varepsilon\}) \leq \frac{t^2}{\varepsilon^2} \underbrace{\sum_{n \geq 1} \frac{1}{2^{n-1}}}_{=1} < \infty, \quad \forall \varepsilon > 0,$$

so the proposition is proved. □

**Corollary 4.5.** For all $t > 0$, we have

$$\lim_{n \to \infty} \sum_{i=1}^{2^n} \left| B\left(\frac{it}{2^n}\right) - B\left(\frac{(i-1)t}{2^n}\right) \right| = \infty \quad \text{a.s.}$$

Consequently, the process $(B_t,\, t \in \mathbb{R}_+)$ has unbounded variation, a.s.

*Proof.* Let us first check that if $g : \mathbb{R}_+ \to \mathbb{R}$ is a continuous function such that

$$\lim_{n \to \infty} \sum_{i=1}^{2^n} \left| g\left(\frac{it}{2^n}\right) - g\left(\frac{(i-1)t}{2^n}\right) \right| < \infty, \quad \text{then} \quad \langle g \rangle_t = \lim_{n \to \infty} \sum_{i=1}^{2^n} \left( g\left(\frac{it}{2^n}\right) - g\left(\frac{(i-1)t}{2^n}\right) \right)^2 = 0.$$

Indeed,

$$\begin{aligned} &\lim_{n \to \infty} \sum_{i=1}^{2^n} \left( g\left(\frac{it}{2^n}\right) - g\left(\frac{(i-1)t}{2^n}\right) \right)^2 \\ &\quad \leq \underbrace{\lim_{n \to \infty} \max_{1 \leq i \leq 2^n} \left| g\left(\frac{it}{2^n}\right) - g\left(\frac{(i-1)t}{2^n}\right) \right|}_{=0} \cdot \underbrace{\lim_{n \to \infty} \sum_{i=1}^{2^n} \left| g\left(\frac{it}{2^n}\right) - g\left(\frac{(i-1)t}{2^n}\right) \right|}_{<\infty} = 0. \end{aligned}$$

So, as we know that the Brownian motion $B$ has continuous trajectories, if it was the case that

$$\mathbb{P}\left(\lim_{n\to\infty}\sum_{i=1}^{2^n}\left|B\left(\frac{it}{2^n}\right)-B\left(\frac{(i-1)t}{2^n}\right)\right|<\infty\right)>0,$$

then this would imply that $\mathbb{P}(\langle B\rangle_t=0)>0$, which is in contradiction with the previous result ($\langle B\rangle_t=t$ a.s.). In conclusion,

$$\mathbb{P}\left(\lim_{n\to\infty}\sum_{i=1}^{2^n}\left|B\left(\frac{it}{2^n}\right)-B\left(\frac{(i-1)t}{2^n}\right)\right|=\infty\right)=1.$$

□

**Final remark.** Notice that $B_t^2-\langle B\rangle_t=B_t^2-t$ is a martingale. This is not a coincidence.

**Quadratic variation of a martingale.**

**Reminder**(from Doob's decomposition theorem). If $(M_t,\, t\in\mathbb{R}_+)$ is a continuous square-integrable martingale, then there exists a unique process $(A_t,\, t\in\mathbb{R}_+)$ which is increasing, continuous and adapted to the same filtration as $(M_t,\, t\in\mathbb{R}_+)$, such that $A_0=0$ and $(M_t^2-A_t,\, t\in\mathbb{R}_+)$ is a martingale.

**Definition 4.6.** The process $A$ is called the *quadratic variation* of the martingale $M$ and is denoted as $A_t=\langle M\rangle_t$, $t\in\mathbb{R}_+$.

**Proposition 4.7.** If $(M_t,\, t\in\mathbb{R}_+)$ is a continuous square-integrable martingale, then

$$\langle M\rangle_t^{(n)}=\sum_{i=1}^{2^n}\left(M\left(\frac{it}{2^n}\right)-M\left(\frac{(i-1)t}{2^n}\right)\right)^2\underset{n\to\infty}{\overset{\mathbb{P}}{\to}}\langle M\rangle_t,\quad\forall t>0,$$

where $(\langle M\rangle_t,\, t\in\mathbb{R}_+)$ is the process defined above.

**Remarks.** - By the above definition, $\mathbb{E}(\langle M\rangle_t)=\mathbb{E}(M_t^2)-\mathbb{E}(M_0^2)$.

- The process $\langle M\rangle$ is increasing : it therefore has bounded variation itself.

- The only martingales with quadratic variation equal to zero are constant processes! So all non-constant martingales have unbounded variation!

**Quadratic covariation.**

Let $M, N$ be two continuous square-integrable martingales (adapted to the same filtration $(\mathcal{F}_t,\, t\in\mathbb{R}_+)$).

**Definition 4.8.** The *quadratic covariation* of $M$ and $N$ is the unique process $\langle M,N\rangle$ which is continuous, adapted, has bounded variation and is such that $\langle M,N\rangle_0=0$ and $(M_tN_t-\langle M,N\rangle_t,\, t\in\mathbb{R}_+)$ is a martingale.

**Remark.** $\langle M,M\rangle_t=\langle M\rangle_t$.

**Proposition 4.9.**

$$\langle M,N\rangle_t^{(n)}=\sum_{i=1}^{2^n}\left(M\left(\frac{it}{2^n}\right)-M\left(\frac{(i-1)t}{2^n}\right)\right)\left(N\left(\frac{it}{2^n}\right)-N\left(\frac{(i-1)t}{2^n}\right)\right)\underset{n\to\infty}{\overset{\mathbb{P}}{\to}}\langle M,N\rangle_t,\quad\forall t>0.$$

**Proposition 4.10.** If $c\in\mathbb{R}$ and $M$, $N$ are independent, then for all $t\in\mathbb{R}_+$,

$$\langle M,N\rangle_t=0\quad\text{and}\quad\langle cM+N\rangle_t=c^2\langle M\rangle_t+\langle N\rangle_t.$$

**Remark.** From the above two propositions, we see that the quadratic variation of a martingale plays the same role as the variance of a random variable. Likewise, the quadratic covariation of two martingales plays the same role as the covariance of two random variables.

## 4.3 Riemann-Stieltjes' integral

Let $T > 0$, $f : \mathbb{R}_+ \to \mathbb{R}$ be a continuous function and $g : \mathbb{R}_+ \to \mathbb{R}$ be a function with bounded variation.

**Definition 4.11.** The *Riemann-Stieltjes integral* of $f$ with respect to $g$ is defined as

$$\int_0^T f(s)\, dg(s) = \lim_{n\to\infty} \sum_{i=1}^{2^n} f\left(s_i^{(n)}\right) \left(g\left(\frac{iT}{2^n}\right) - g\left(\frac{(i-1)T}{2^n}\right)\right),$$

where $s_i^{(n)}$ is a sequence of numbers such that $\frac{(i-1)T}{2^n} \le s_i^{(n)} \le \frac{iT}{2^n}$.

**Proposition 4.12.** If moreover $g$ is continuously differentiable, then $\int_0^T f(s)\, dg(s) = \int_0^T f(s)\, g'(s)\, ds$.

**Proposition 4.13.** (classical integration by parts formula)
If $f, g$ are both continuous and with bounded variation, then

$$\int_0^T f(s)\, dg(s) = f(T)\, g(T) - f(0)\, g(0) - \int_0^T g(s)\, df(s).$$

In particular,

$$\int_0^T f(s)\, df(s) = \frac{1}{2}\left(f(T)^2 - f(0)^2\right).$$

**Generalization to continuous-time stochastic processes.**

**Definition 4.14.** If $(H_t,\, t \in \mathbb{R}_+)$ is a process with continuous trajectories and $(V_t,\, t \in \mathbb{R}_+)$ is a process with bounded variation, then

$$\left(\int_0^T H_s\, dV_s\right)(\omega) = \int_0^T H_s(\omega)\, dV_s(\omega).$$

**Remark.** If $V$ represents the evolution of a stock price and $H_s$ represents the amount of stock $V$ owned at time $s$, then $\int_0^T H_s\, dV_s$ represents the gain made on the period $[0, T]$ by investing with the strategy $H$ on the stock $V$.

Now, what if the stock price evolution is a standard Brownian motion, or more generally a martingale?

**Problem.** The trajectories of the Brownian motion or a martingale have unbounded variation, almost surely, so how to define $\int_0^T H_s\, dB_s$?

**Partial solution.** If $H$ has continuous trajectories with bounded variation, then using formally the above integration by parts formula, we could define

$$\int_0^T H_s\, dB_s = H_T\, B_T - H_0\, B_0 - \int_0^T B_s\, dH_s,$$

as $B$ has continuous trajectories and $H$ has trajectories with bounded variation. This is fine, but then, how to define $\int_0^T B_s\, dB_s$?

## 4.4 Ito's stochastic integral

Our aim in the following is to define the process $((H \cdot B)_t \equiv \int_0^t H_s\, dB_s,\ t \in [0,T])$, where

- $T > 0$ is a fixed time horizon,

- $(B_t,\ t \in \mathbb{R}_+)$ is a standard Brownian motion with respect to a filtration $(\mathcal{F}_t,\ t \in \mathbb{R}_+)$, that is:

(i) $B_t$ is $\mathcal{F}_t$-measurable, $\forall t \in \mathbb{R}_+$,
(ii) $B_t - B_s \perp\!\!\!\perp \mathcal{F}_s$, $\forall t \geq s \geq 0$,

- $(H_t,\ t \in \mathbb{R}_+)$ is a continuous process, adapted to $(\mathcal{F}_t,\ t \in \mathbb{R}_+)$ and such that $\mathbb{E}\left(\int_0^T H_s^2\, ds\right) < \infty$.

**First step.**

For $n \geq 1$ fixed, let us define

$$H_t^{(n)} = \sum_{i=1}^{2^n} H\left(\frac{(i-1)T}{2^n}\right) 1_{]\frac{(i-1)T}{2^n}, \frac{iT}{2^n}]}(t), \quad t \in [0,T].$$

**Preliminary fact.** Under the above assumption made on $H$, we have

$$\mathbb{E}\left(\int_0^T (H_s^{(n)} - H_s)^2\, ds\right) \underset{n\to\infty}{\to} 0.$$

Let us now define, for a fixed $n \geq 1$,

$$(H^{(n)} \cdot B)_T \equiv \int_0^T H_s^{(n)}\, dB_s = \sum_{i=1}^{2^n} H\left(\frac{(i-1)T}{2^n}\right)\left(B\left(\frac{iT}{2^n}\right) - B\left(\frac{(i-1)T}{2^n}\right)\right).$$

For ease of notation, let us write $t_i = \frac{iT}{2^n}$, so

$$(H^{(n)} \cdot B)_T = \sum_{i=1}^{2^n} H_{t_{i-1}}(B_{t_i} - B_{t_{i-1}}).$$

**Properties.**

- Linearity: $((cH^{(n)} + K^{(n)}) \cdot B)_T = c\,(H^{(n)} \cdot B)_T + (K^{(n)} \cdot B)_T$.

- $\mathbb{E}((H^{(n)} \cdot B)_T) = 0$. Indeed:

$$\mathbb{E}((H^{(n)} \cdot B)_T) = \sum_{i=1}^{2^n} \mathbb{E}(H_{t_{i-1}}\,(B_{t_i} - B_{t_{i-1}})) = \sum_{i=1}^{2^n} \mathbb{E}(H_{t_{i-1}})\,\mathbb{E}(B_{t_i} - B_{t_{i-1}}) = 0,$$

as $H_{t_{i-1}}$ is $\mathcal{F}_{t_i}$-measurable and $B_{t_i} - B_{t_{i-1}} \perp\!\!\!\perp \mathcal{F}_{t_{i-1}}$.

- Isometry property: $\mathrm{Var}((H^{(n)} \cdot B)_T) = \mathbb{E}((H^{(n)} \cdot B)_T^2) = \mathbb{E}\left(\int_0^T (H_s^{(n)})^2\, ds\right)$. Indeed:

$$\begin{aligned}
\mathbb{E}((H^{(n)} \cdot B)_T^2) &= \sum_{i,j=1}^{2^n} \mathbb{E}(H_{t_{i-1}}\,(B_{t_i} - B_{t_{i-1}})\,H_{t_{j-1}}\,(B_{t_j} - B_{t_{j-1}})) \\
&= \sum_{i=1}^{2^n} \mathbb{E}(\mathbb{E}(H_{t_{i-1}}^2\,(B_{t_i} - B_{t_{i-1}})^2|\mathcal{F}_{t_{i-1}})) + 2\sum_{i<j} \mathbb{E}(\mathbb{E}(H_{t_{i-1}}\,(B_{t_i} - B_{t_{i-1}})\,H_{t_{j-1}}(B_{t_j} - B_{t_{j-1}})|\mathcal{F}_{t_{j-1}})) \\
&= \sum_{i=1}^{2^n} \mathbb{E}(H_{t_{i-1}}^2\,\mathbb{E}((B_{t_i} - B_{t_{i-1}})^2|\mathcal{F}_{t_{i-1}})) + 2\sum_{i<j} \mathbb{E}(H_{t_{i-1}}\,(B_{t_i} - B_{t_{i-1}})\,H_{t_{j-1}}\mathbb{E}(B_{t_j} - B_{t_{j-1}}|\mathcal{F}_{t_{j-1}}))
\end{aligned}$$

as $H_{t_{i-1}}$ is $\mathcal{F}_{t_{i-1}}$-measurable and $H_{t_{i-1}}\,(B_{t_i}-B_{t_{i-1}})\,H_{t_{j-1}}$ is $\mathcal{F}_{t_{j-1}}$-measurable for $i<j$. Therefore,

$$\begin{aligned}\mathbb{E}((H^{(n)}\cdot B)_T^2) &= \sum_{i=1}^{2^n}\mathbb{E}(H_{t_{i-1}}^2\,\underbrace{\mathbb{E}((B_{t_i}-B_{t_{i-1}})^2)}_{=t_i-t_{i-1}}) + 2\sum_{i<j}\mathbb{E}(H_{t_{i-1}}\,(B_{t_i}-B_{t_{i-1}})\,H_{t_{j-1}}\,\underbrace{\mathbb{E}(B_{t_j}-B_{t_{j-1}})}_{=0})\\ &= \mathbb{E}\left(\sum_{i=1}^{2^n}H_{t_{i-1}}^2\,(t_i-t_{i-1})\right)=\mathbb{E}\left(\int_0^T(H_s^{(n)})^2\,ds\right).\end{aligned}$$

Similarly, if $K$ is another continuous and adapted process such that $\mathbb{E}\left(\int_0^T K_s^2\,ds\right)<\infty$, then

- $\mathbb{E}((H^{(n)}\cdot B)_T\,(K^{(n)}\cdot B)_T)=\mathbb{E}\left(\int_0^T H_s^{(n)}\,K_s^{(n)}\,ds\right)$.

**Second step.**

Let us now define, for $t\in[0,T]$ (and keeping the notation $t_i=\frac{iT}{2^n}$):

$$(H^{(n)}\cdot B)_t=\sum_{i=1}^{k-1}H_{t_{i-1}}\,(B_{t_i}-B_{t_{i-1}})+H_{t_{k-1}}\,(B_t-B_{t_{k-1}}),\quad\text{if } t_{k-1}<t\le t_k.$$

**Proposition 4.15.** For all $n\ge 1$ fixed, the process $((H^{(n)}\cdot B)_t,\,t\in[0,T])$ is a continuous square-integrable martingale with respect to $(\mathcal{F}_t,\,t\in[0,T])$.

*Proof.* The fact that the process is continuous is clear from the definition. It is also easily checked that it is square-integrable (and actually that $\mathbb{E}((H^{(n)}\cdot B)_t^2)=\mathbb{E}\left(\int_0^t(H_s^{(n)})^2\,ds\right)$, $\forall t\in[0,T]$). Let us now prove that it is a martingale:

(i) $\mathbb{E}(|(H^{(n)}\cdot B)_t|)<\infty$, as $\mathbb{E}((H^{(n)}\cdot B)_t^2)<\infty$.

(ii) $(H^{(n)}\cdot B)_t$ is clearly $\mathcal{F}_t$-measurable, for all $t\in[0,T]$.

(iii) Let us now check that $\mathbb{E}((H^{(n)}\cdot B)_T|\mathcal{F}_t)=(H^{(n)}\cdot B)_t$, $\forall t\in[0,T]$, assuming that $t_{k-1}<t\le t_k$:

$$\begin{aligned}&\mathbb{E}((H^{(n)}\cdot B)_T|\mathcal{F}_t)\\ &= \mathbb{E}\left(\sum_{i=1}^{k-1}H_{t_{i-1}}\,(B_{t_i}-B_{t_{i-1}})\Big|\mathcal{F}_t\right)+\mathbb{E}(H_{t_{k-1}}\,(B_{t_k}-B_{t_{k-1}})|\mathcal{F}_t)+\mathbb{E}\left(\sum_{i=k+1}^{2^n}H_{t_{i-1}}\,(B_{t_i}-B_{t_{i-1}})\Big|\mathcal{F}_t\right)\\ &= \sum_{i=1}^{k-1}H_{t_{i-1}}\,(B_{t_i}-B_{t_{i-1}})+H_{t_{k-1}}\,\mathbb{E}(B_{t_k}-B_{t_{k-1}}|\mathcal{F}_t)+\sum_{i=k+1}^{2^n}\mathbb{E}(\mathbb{E}(H_{t_{i-1}}\,(B_{t_i}-B_{t_{i-1}})|\mathcal{F}_{t_{i-1}})|\mathcal{F}_t)\end{aligned}$$

as the first sum is $\mathcal{F}_t$-measurable, $H_{t_{k-1}}$ is $\mathcal{F}_t$-measurable and $\mathcal{F}_t\subset\mathcal{F}_{t_{i-1}}$ for all $i\ge k+1$. Therefore,

$$\begin{aligned}\mathbb{E}((H^{(n)}\cdot B)_T|\mathcal{F}_t) &= \sum_{i=1}^{k-1}H_{t_{i-1}}\,(B_{t_i}-B_{t_{i-1}})+H_{t_{k-1}}\,(B_t-B_{t_{k-1}})+\sum_{i=k-1}^{2^n}\mathbb{E}(H_{t_{i-1}}\,\underbrace{\mathbb{E}(B_{t_i}-B_{t_{i-1}})}_{=0}|\mathcal{F}_t)\\ &= (H^{(n)}\cdot B)_t.\end{aligned}$$

From this, we deduce that $\forall 0\le s<t\le T$,

$$\mathbb{E}((H^{(n)}\cdot B)_t|\mathcal{F}_s)=\mathbb{E}(\mathbb{E}((H^{(n)}\cdot B)_T|\mathcal{F}_t)|\mathcal{F}_s)=\mathbb{E}((H^{(n)}\cdot B)_T|\mathcal{F}_s)=(H^{(n)}\cdot B)_s,$$

which proves the martingale property.

□

**Third step.**

**Important fact.** For every continuous and adapted process $H$ such that $\mathbb{E}\left(\int_0^T H_s^2\,ds\right) < \infty$, there exists a process $((H\cdot B)_t,\, t\in[0,T])$ such that

$$\mathbb{E}\left(\sup_{0\le t\le T}((H^{(n)}\cdot B)_t - (H\cdot B)_t)^2\right) \underset{n\to\infty}{\to} 0$$

In particular, $(H^{(n)}\cdot B)_t \overset{\mathbb{P}}{\to} (H\cdot B)_t$, $\forall t\in[0,T]$. In addition, the process $(H\cdot B)$ possesses all the properties mentioned above, namely :

- $((cH+K)\cdot B)_t = c\,(H\cdot B)_t + (K\cdot B)_t$.

- $\mathbb{E}((H\cdot B)_t) = 0$, $\forall t\in[0,T]$.

- $\mathbb{E}((H\cdot B)_t^2) = \mathbb{E}\left(\int_0^T H_s^2\,ds\right)$, $\forall t\in[0,T]$.

- $\mathbb{E}((H\cdot B)_t(K\cdot B)_s) = \mathbb{E}\left(\int_0^{t\wedge s} H_r\,K_r\,dr\right)$, $\forall 0\le s\le t\le T$.

- $((H\cdot B)_t,\, t\in[0,T])$ is a continuous square-integrable martingale.

**Alternate notation.** $(H\cdot B)_t \equiv \int_0^t H_s\,dB_s$.

**Remark.** In the definition of the Riemann-Stieltjes integral, the choice of the point $s_i^{(n)} \in [\frac{(i-1)T}{2^n}, \frac{iT}{2^n}]$ is arbitrary. But in the definition of Ito's integral, the choice $s_i^{(n)} = \frac{(i-1)T}{2^n}$ is crucial in order to preserve the martingale property of the integral. Another choice leads to a different integral.

**Example: Fisk-Stratonovič's integral.**

$$\begin{aligned}(H\circ B)_T &\equiv \int_0^T H_s\circ dB_s = \lim_{n\to\infty}\sum_{i=1}^{2^n}\left(\frac{H\left(\frac{(i-1)T}{2^n}\right)+H\left(\frac{iT}{2^n}\right)}{2}\right)\left(B\left(\frac{iT}{2^n}\right)-B\left(\frac{(i-1)T}{2^n}\right)\right)\\ &= (H\cdot B)_T + \frac{1}{2}\,\langle H,B\rangle_T,\end{aligned}$$

where $\langle H,B\rangle_T$ is the quadratic covariation of $H$ and $B$ defined as

$$\langle H,B\rangle_T = \lim_{n\to\infty}\sum_{i=1}^{2^n}\left(H\left(\frac{iT}{2^n}\right)-H\left(\frac{(i-1)T}{2^n}\right)\right)\left(B\left(\frac{iT}{2^n}\right)-B\left(\frac{(i-1)T}{2^n}\right)\right).$$

As soon as this quadratic covariation is non-zero, the process $(H\circ B)$ is not a martingale. Notice that in the case where $H$ is a continuous process with bounded variation, the quadratic covariation is zero, so the two definitions of integral coincide. It was already suggested at the end of Section 4.3 that in this case, an alternate definition of the integral $\int_0^T H_s\,dB_s$ is possible.

**Quadratic variation of the stochastic integral.**

**Proposition 4.16.** Let $H,K$ be continuous and adapted processes such that $\mathbb{E}\left(\int_0^T H_s^2\,ds\right) < \infty$ and $\mathbb{E}\left(\int_0^T K_s^2\,ds\right) < \infty$. Then

$$\langle (H\cdot B)\rangle_t = \int_0^t H_s^2\,ds \quad\text{and}\quad \langle (H\cdot B),(K\cdot B)\rangle_t = \int_0^t H_s\,K_s\,ds, \quad t\in[0,T].$$

That is, the processes $\left((H\cdot B)_t^2 - \int_0^t H_s^2\,ds,\, t\in[0,T]\right)$ and $\left((H\cdot B)_t\,(K\cdot B)_t - \int_0^t H_s\,K_s\,ds,\, t\in[0,T]\right)$ are martingales.

**Remarks.** - The process $\left(\langle (H\cdot B)\rangle_t = \int_0^t H_s^2\,ds,\, t\in[0,T]\right)$ is continuous, adapted and increasing (as $H_s^2\geq 0$). It therefore has bounded variation itself.

- The process $\left(\langle (H\cdot B),(K\cdot B)\rangle_t = \int_0^t H_sK_s\,ds,\, t\in[0,T]\right)$ is continuous, adapted and has bounded variation.

- If $H_t\equiv 1$, then $\langle (H\cdot B)\rangle_t = \langle B\rangle_t = t$, as we already knew.

- $\mathbb{E}(\langle (H\cdot B)\rangle_t) = \mathbb{E}\left(\int_0^t H_s^2\,ds\right) = \mathbb{E}((H\cdot B)_t^2)$, as we already knew also.

**Particular case of stochastic integral: the Wiener integral.**

If $H_t(\omega) = f(t)$ is a deterministic function, then the stochastic integral $(f\cdot B)$ has the following additional properties :

- $((f\cdot B)_t,\, t\in[0,T])$ is a Gaussian process with mean $m(t) = \mathbb{E}((f\cdot B)_t) = 0$ and covariance

$$K(t,s) = \mathbb{E}((f\cdot B)_t\,(f\cdot B)_s) = \int_0^{t\wedge s} f(r)^2\,dr.$$

- $((f\cdot B)_t,\, t\in[0,T])$ has independent increments.

- $\langle (f\cdot B)\rangle_t = \int_0^t f(s)^2\,ds$ is a deterministic process.

# 5 Stochastic calculus

## 5.1 Ito-Doeblin's formula(s)

**Preliminary.** If $V$ is a process with bounded variation and $f:\mathbb{R}\to\mathbb{R}$ is differentiable, then

$$f(V_t) - f(V_0) = \int_0^t f'(V_s)\,dV_s.$$

This rule changes if the process $V$ does not have bounded variation: this is Ito-Doeblin's formula.

**First version.**

**Theorem 5.1.** Let $(B_t,\, t\in\mathbb{R}_+)$ be a standard Brownian motion and $f:\mathbb{R}\to\mathbb{R}$ be twice continuously differentiable and such that $\mathbb{E}(\int_0^t (f'(B_s))^2\,ds) < \infty$, $\forall t\geq 0$. Then

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)\,dB_s + \frac{1}{2}\int_0^t f''(B_s)\,ds, \quad a.s. \quad \forall t\geq 0.$$

**Remarks.**

- $\int_0^t f'(B_s)\,dB_s$ is a well-defined stochastic integral, because of the assumptions made.

- $\frac{1}{2}\int_0^t f''(B_s)\,ds$ is a correction term that arises because of the non-zero quadratic variation of $B$.

*Proof.* (main idea) Let $t_i = \frac{it}{2^n}$ be a partition of $[0,t]$. Then

$$f(B_t) - f(B_0) = \sum_{i=1}^{2^n} (f(B_{t_i}) - f(B_{t_{i-1}})).$$

Using the classical Taylor expansion

$$f(y) - f(x) = f''(x)\,(y-x) + \frac{1}{2}\,f''(x)\,(y-x)^2 + o((y-x)^2),$$

where by definition, $\lim_{h\to 0} \frac{o(h^2)}{h^2} = 0$, we obtain

$$\begin{aligned} f(B_t) - f(B_0) \quad &= \quad \sum_{i=1}^{2^n} f'(B_{t_i})\,(B_{t_i} - B_{t_{i-1}}) + \frac{1}{2}\sum_{i=1}^{2^n} f''(B_{t_{i-1}})\,(B_{t_i} - B_{t_{i-1}})^2 + \sum_{i=1}^{2^n} o((B_{t_i} - B_{t_{i-1}})^2) \\ &\underset{n\to\infty}{\overset{\mathbb{P}}{\to}} \quad \underbrace{\int_0^t f'(B_s)\,dB_S}_{\text{by definition}} + \frac{1}{2}\int_0^t f''(B_s)\,\underbrace{d\langle B\rangle_s}_{=ds} + 0. \end{aligned}$$

In order to get an intuition as to why the second term converges to $\frac{1}{2}\int_0^t f''(B_s)\,ds$, remember that in the case where $f''(x) \equiv 1$, we have

$$\sum_{i=1}^{2^n} (B_{t_i} - B_{t_{i-1}})^2 \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} \langle B\rangle_t = t.$$

□

**Examples.** - Let $f(x) = x$: then $f'(x) = 1$, $f''(x) = 0$ and $\mathbb{E}\left(\int_0^t f'(B_s)^2\,ds\right) = t < \infty$, so

$$B_t - B_0 = \int_0^t 1\,dB_s + 0.$$

Here, the classical rule applies.

- Let $f(x) = x^2$: then $f'(x) = 2x$, $f''(x) = 2$ and

$$\mathbb{E}\left(\int_0^t f'(B_s)^2\,ds\right) = \mathbb{E}\left(\int_0^t 4B_s^2\,ds\right) = 4\int_0^t \mathbb{E}(B_s^2)\,ds = 4\int_0^t s\,ds = 2t^2 < \infty.$$

So

$$B_t^2 - B_0^2 = B_t^2 = \int_0^t 2B_s\,dB_s + \frac{1}{2}\int_0^t 2\,ds = 2\int_0^t B_s\,dB_s + t$$

This is actually Doob's decomposition of the submartingale $B_t^2$: $2\int_0^t B_s\,dB_s$ is a martingale and $t$ is an increasing deterministic process, which is the quadratic variation of the standard Brownian motion. Notice that the above formula may be rewritten as

$$\int_0^t B_s\,dB_s = \frac{1}{2}\,(B_t^2 - t)$$

and that the two processes on both sides are martingales (this by the way shows why the formula $\int_0^t B_s\,dB_s = \frac{1}{2}\,B_t^2$ cannot hold, as $\frac{1}{2}\,B_t^2$ is a submartingale, but not a martingale).

- Let $f(x) = e^x$: then $f'(x) = f''(x) = e^x$ and

$$\mathbb{E}\left(\int_0^t (f'(B_s))^2\,ds\right) = \mathbb{E}\left(\int_0^t e^{2B_s}\,ds\right) = \int_0^t e^{\frac{4s}{2}}\,ds = \frac{e^{2t}-1}{2} < \infty.$$

So

$$e^{B_t} - e^{B_0} = e^{B_t} - 1 = \int_0^t e^{B_s}\, dB_s + \frac{1}{2}\int_0^t e^{B_s}\, ds.$$

This is again Doob's decomposition of the submartingale $e^{B_t} - 1$: the stochastic integral is a martingale, while the Riemann integral is an increasing process (as $e^{B_s} > 0$).

**Remark.** Let $X_t = e^{B_t}$. The above formula then reads

$$X_t - 1 = \int_0^t X_s\, dB_s + \frac{1}{2}\int_0^t X_s\, ds.$$

This is our first example of *stochastic differential equation* (in integral form!).

**Second version.**

**Notation.** For a function $f$ of two variables $t, x$, we adopt the following notation for partial derivatives: $f'_t$,$f'_x$, $f''_{xx}$ and so on.

**Theorem 5.2.** Let $(B_t,\, t \in \mathbb{R}_+)$ be a standard Brownian motion and $f : \left\{ \begin{array}{ccc} \mathbb{R}_+ \times \mathbb{R} & \to & \mathbb{R} \\ (t,x) & \mapsto & f(t,x) \end{array} \right.$ be continuously differentiable in $t$, twice continuously differentiable in $x$ and such that

$$\mathbb{E}\left(\int_0^t (f'_x(s, B_s))^2\, ds\right) < \infty, \quad \forall t \geq 0.$$

Then

$$f(t, B_t) - f(0, B_0) = \int_0^t f'_t(s, B_s)\, ds + \int_0^t f'_x(s, B_s)\, dB_s + \frac{1}{2}\int_0^t f''_{xx}(s, B_s)\, ds, \quad a.s. \quad \forall t \geq 0.$$

The proof follows the same lines as for the first version, this time using the Taylor expansion :

$$f(s,y) - f(t,x) = f'_t(t,x)\,(s-t) + f'_x(t,x)\,(y-x) + \frac{1}{2}\, f''_{xx}(t,x)\,(y-x)^2 + o((s-t),(y-x)^2).$$

**Examples.** - Let $f(t,x) = x^2 - t$: then $f'_t(t,x) = -1$, $f'_x(t,x) = 2x$, $f''_{xx}(t,x) = 2$, so

$$B_t^2 - t = \int_0^t (-1)\, ds + \int_0^t 2B_s\, dB_s + \frac{1}{2}\int_0^t 2\, ds = 2\int_0^t B_s\, dB_s.$$

This is the same formula as above, actually.

- Let $f(t,x) = e^{x - t/2}$: then $f'_t(t,x) = -\frac{1}{2}\, f(t,x)$, $f'_x = (t,x) = f''_{xx}(t,x) = f(t,x)$, so

$$e^{B_t - t/2} - 1 = -\frac{1}{2}\int_0^t e^{B_s - s/2}\, ds + \int_0^t e^{B_s - s/2}\, dB_s + \frac{1}{2}\int_0^t e^{B_s - s/2\, ds} = \int_0^t e^{B_s - s/2}\, dB_s.$$

Again, notice that the two processes on both sides are martingales.

**Remark.** Let $Y_t = e^{B_t - t/2}$. The above formula then reads

$$Y_t - 1 = \int_0^t Y_s\, dB_s.$$

This is our second example of stochastic differential equation (again in integral form).

**Remark.** In the above two examples, $f'_t(t,x) + \frac{1}{2}\, f''_{xx}(t,x) = 0$. From the Ito-Doeblin formula, we see that if $f$ satisfies this *partial differential equation*, then $(f(t, B_t),\, t \in \mathbb{R}_+)$ is a martingale.

**Third version.**

This version is yet a generalization of the previous one. Let $(B_t,\, t \in \mathbb{R}_+)$ be a standard Brownian motion, $(H_t,\, t \in \mathbb{R}_+)$ be a continuous and adapted process such that

$$\mathbb{E}\left(\int_0^t H_s^2\, ds\right) < \infty, \quad \forall t \geq 0,$$

and $(K_t,\, t \in \mathbb{R}_+)$ be a continuous and adapted process. Let now $M_t = \int_0^t H_s\, dB_s$ ($M$ is a continuous square-integrable martingale) and $V_t = \int_0^t K_s\, ds$ ($V$ is a continuous process with bounded variation). Let finally $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be continuously differentiable in $t$, twice continuously differentiable in $x$ and such that

$$\mathbb{E}\left(\int_0^t (f'_x(V_s, M_s))^2\, H_s^2\, ds\right) < \infty, \quad \forall t \geq 0.$$

Then for all $t \geq 0$,

$$f(V_t, M_t) - f(V_0, M_0) = \int_0^t f'_t(V_s, M_s)\, K_s\, ds + \int_0^t f'_x(V_s, M_s)\, H_s\, dB_s + \frac{1}{2}\int_0^t f''_{xx}(V_s, M_s)\, H_s^2\, ds, \quad a.s.$$

**Examples.** - Let $f(t,x) = t + x$ : then $f'_t(t,x) = f'_x(t,x) = 1$, $f''_{xx} = 0$, so

$$(V_t + M_t) - (V_0 + M_0) = \int_0^t K_s\, ds + \int_0^t H_s\, dB_s.$$

- Let $f(t,x) = tx$: then $f'_t(t,x) = x$, $f'_x(t,x) = t$, $f''_{xx}(t,x) = 0$, so

$$V_t M_t - V_0 M_0 = \int_0^t M_s\, K_s\, ds + \int_0^t V_s\, H_s\, dB_s.$$

## 5.2 Stochastic differential equations: a first approach through examples

A (time-homogeneous) stochastic differential equation (SDE) reads :

$$X_t = x_0 + \int_0^t f(X_s)\, ds + \int_0^t g(X_s)\, dB_s, \quad \text{in integral form}$$

and

$$\begin{cases} dX_t = f(X_t)\, dt + g(X_t)\, dB_t, \\ X_0 = x_0, \end{cases} \quad \text{in differential form.}$$

Here $x_0 \in \mathbb{R}$, $B$ is a standard Brownian motion and $f, g : \mathbb{R} \to \mathbb{R}$ are deterministic functions. Notice that the differential form is deduced from the integral form, but has no meaning *per se*, as the Brownian motion is not differentiable.

**Remark.** Solving an SDE is in general much more difficult than solving an ordinary differential equation (ODE) of the form $dX_t/dt = f(X_t)$. Beyond the issue of knowing whether a solution exists and is unique, there are actually not that many functions $f$ and $g$ for which the solution is known explicitly. The behaviour of the solution of an SDE is also very different from that of an ODE.

**Terminology.**

- The term with the function $f$ is called the *drift term*.

- The term with the function $g$ is called the *diffusive term*.

We are now going to see various examples of SDE's.

**Black-Scholes equation.**

$$\begin{cases} dX_t = \mu\, X_t\, dt + \sigma\, X_t\, dB_t, \\ \\ X_0 = x_0 > 0, \end{cases}$$

where $\mu \in \mathbb{R}$ is the drift, $\sigma \geq 0$ is the volatility of the process $X$.

In the case where $\sigma = 0$, the equation becomes an ODE:

$$\frac{dX_t}{X_t} = \mu\, dt, \quad \text{so} \quad \ln X_t - \ln x_0 = \mu t, \quad X_t = x_0\, \exp(\mu t).$$

In the case where $\sigma > 0$, on could be tempted to solve the equation with the following WRONG reasoning:

$$\frac{dX_t}{X_t} = \mu\, dt + \sigma X_t\, dB_t, \quad \text{so} \quad \ln X_t - \ln x_0 = \mu t + \sigma B_t, \quad X_t = x_0 \exp(\mu t + \sigma B_t)?$$

But actually, the correct solution is $X_t = x_0\, \exp((\mu - \frac{\sigma^2}{2})t + \sigma B_t)$. This can be verified using Ito-Doeblin's formula (we have already seen this for $\mu = \frac{1}{2}$, $\sigma = 1$ and $\mu = 0$, $\sigma = 1$, by the way). The rules of standard calculus do not apply here! The mistake in the above reasoning is to "integrate" $dX_t/X_t$ classically to $\ln X_t$.

The process $X$ describes (relatively) well the evolution of a stock price. Here are some important features of $X$ :

- $X_t > 0$, $\forall t \geq 0$, as it is an exponential.

- $X_t = \exp(\text{Gaussian})$, i.e. $X_t$ has a log-normal distribution.

- $\ln(X_t)$ is a process with independent increments.

Notice also that

$$\mathbb{E}(X_t) = \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) t\right)\, \mathbb{E}(\exp(\sigma B_t)) = \exp(\mu t),$$

whereas

$$\mathbb{E}(\ln X_t) = \left(\mu - \frac{\sigma^2}{2}\right) t + \mathbb{E}(\sigma B_t) = \left(\mu - \frac{\sigma^2}{2}\right) t < \mu t.$$

Finally, $X$ is a martingale if and only if $\mu = 0$ (as in this case, $X_t = x_0 + \int_0^t \sigma X_s\, dB_s$).

**Ornstein-Uhlenbeck's process.**
(also known as *Langevin's* or *autoregressive* or *mean-reverting* process!)

$$\begin{cases} dX_t = -aX_t\, dt + \sigma\, dB_t, \\ \\ X_0 = x_0 \in \mathbb{R}, \end{cases}$$

where $a \geq 0$ and $\sigma \geq 0$. The solution to this equation reads:

$$X_t = x_0\, e^{-at} + \sigma \int_0^t e^{-a(t-s)}\, dB_s.$$

Notice that even though $X_t$ may be written as a Wiener integral, it is *not* a martingale (even in the case $x_0 = 0$), as the integrand itself $e^{-a(t-s)}$ depends on $t$.

Here, $\mathbb{E}(X_t) = x_0 e^{-at}$ and

$$\text{Var}(X_t) = \sigma^2 \int_0^t e^{-2a(t-s)}\, ds = \sigma^2 \left(\frac{1 - e^{-2at}}{2a}\right) \underset{t\to\infty}{\to} \frac{\sigma^2}{2a},$$

whereas when $\sigma = 0$, $X(t) = x_0\, e^{-at} \underset{t\to\infty}{\to} 0$ and when $a = 0$, $X_t = \sigma B_t$ diverges when $t \to \infty$.

**Brownian bridge.**

$$\begin{cases} dX_t = -\dfrac{X_t}{T-t}\,dt + \sigma\,dB_t, \\ \\ X_0 = x_0 \in \mathbb{R}, \end{cases}$$

where $T > 0$ and $\sigma \geq 0$. The solution reads:

$$X_t = (T-t)\,x_0 + \sigma \int_0^t \left(\frac{T-t}{T-s}\right)\,dB_s.$$

Here, $\mathbb{E}(X_t) = (T-t)\,x_0$ and

$$\mathrm{Var}(X_t) = \sigma^2 \int_0^t \left(\frac{T-t}{T-s}\right)^2\,ds = \sigma^2\,t(T-t) \underset{t\to T}{\to} 0$$

so $X_T = 0$, in all cases!

**Bessel's process.**

$$\begin{cases} dX_t = \dfrac{a}{X_t}\,dt + dB_t, \\ \\ X_0 = 0, \end{cases}$$

where $a \geq \frac{1}{2}$. An explicit solution is known when $a = \frac{n-1}{2}$, for some integer $n \geq 2$; then $X_t = \|\underline{W}_t\|$, where $\underline{W}_t$ is an $n$-dimensional standard Brownian motion.

## 5.3 Numerical simulation of stochastic differential equations

As stochastic differential equations are typically hard to solve, it is good to know how to obtain solutions by means of numerical simulation. The aim here is to obtain approximations for expressions such as $\mathbb{E}(h(X_T))$ where $T > 0$, $h : \mathbb{R} \to \mathbb{R}$ is some function and $X$ is the solution of the SDE:

$$\begin{cases} dX_t = f(X_t)\,dt + g(X_t)\,dB_t, \\ \\ X_0 = x_0 \in \mathbb{R}. \end{cases}$$

A natural procedure is to perform $M$ independent runs $X^{(1)}, \ldots, X^{(M)}$ of the process $X$ and to approximate the above expectation by the empirical mean :

$$\mathbb{E}(h(X_T)) \approx \frac{1}{M}\sum_{j=1}^{M} h(X_T^{(j)}),$$

which is known to be correct up to a term of order $O(\frac{1}{\sqrt{M}})$, by the central limit theorem. How to run each process $X^{(j)}$ up to time $T$ is a more delicate question that is addressed in the following.

**Euler-Maruyama's scheme.**

First remember that the above SDE reads in integral form:

$$X_t = x_0 + \int_0^t f(X_r)\,dr + \int_0^t g(X_t)\,dB_r, \quad t \geq 0. \tag{2}$$

From this, we deduce that for $t \geq s \geq 0$,

$$X_t = X_s + \int_s^t f(X_r)\,dr + \int_s^t g(X_r)\,dB_r.$$

If the difference $t-s$ is small, this gives us an idea as how to approximate the value of $X_t$ given the value of $X_s$, by noticing that

$$X_t \approx X_s + f(X_s)\,(t-s) + g(X_s)\,\underbrace{(B_t - B_s)}_{\sim\mathcal{N}(0,t-s)}. \tag{3}$$

From there, we define the following numerical scheme: let $N \geq 1$, $\Delta = \frac{T}{N}$ and

$$\begin{cases} X_0^{(N)} = x_0, \\ \\ X_{(n+1)\Delta}^{(N)} = X_{n\Delta}^{(N)} + f(X_{n\Delta}^{(N)})\,\Delta + g(X_{n\Delta}^{(N)})\,\sqrt{\Delta}\,\xi_{n+1}, \quad 0 \leq n \leq N-1, \end{cases}$$

where $(\xi_n,\, n \geq 1)$ is a sequence of i.i.d.$\sim \mathcal{N}(0,1)$ random variables. The process $X$ at time $T$ is then approximated by $X_T \approx X_T^{(N)} = X_{N\Delta}^{(N)}$ (and this has to be done for each independent run of the process).

Now, how close is $X_T^{(N)}$ to $X_T$? Without entering in to much detail here, let us simply say approximately that if $h$ is a continuous function and $f$ and $g$ are such that equation (2) admits a unique solution, then

$$|\mathbb{E}(h(X_T^{(N)})) - \mathbb{E}(h(X_T))| \leq \frac{C_T}{\sqrt{N}},$$

where $C_T$ is a constant (growing exponentially with $T$).

**Remark.** For ODE's, the convergence of the corresponding Euler scheme is better, namely

$$|h(X_T^{(N)}) - h(X_T)| \leq \frac{C_T}{N}.$$

The explanation for this difference in the order of convergence is that the approximation (3) is correct up to a term of order $O(t-s)$, whereas for ODE's, the corresponding approximation gives $X_t \approx X_s + f(X_s)\,(t-s)$ up to a term of order $O((t-s)^2)$, therefore much smaller that $O(t-s)$ when $t-s$ is small.

The final approximation formula is

$$\mathbb{E}(h(X_T)) \approx \frac{1}{M}\sum_{j=1}^{M} h(X_T^{(j,N)}),$$

where $X_T^{(j,N)}$ represents the final value at time $T$ of the $j^{th}$ run of the numerical scheme.

**Milstein's scheme.**

A better approximation than (3) turns out to be the following:

$$X_t \approx X_s + f(X_s)\,(t-s) + g(X_s)\,(B_t - B_s) + g(X_s)\,g'(X_s)\,\left(\frac{(B_t-B_s)^2 - (t-s)}{2}\right).$$

This approximation is correct up to a term of order $O((t-s)^{3/2})$. From there, we define the following numerical scheme:

$$\begin{cases} X_0^{(N)} = x_0 \\ \\ X_{(n+1)\Delta}^{(N)} = X_{n\Delta}^{(N)} + f(X_{n\Delta}^{(N)})\,\Delta + g(X_{n\Delta}^{(N)})\,\sqrt{\Delta}\,\xi_{n+1} \\ \\ \qquad\qquad + g(X_{n\Delta}^{(N)})\,g'(X_{n\Delta}^{(N)})\,\Delta\,\left(\frac{\xi_{n+1}^2 - 1}{2}\right), \quad 0 \leq n \leq N-1, \end{cases}$$

where $(\xi_n,\, n \geq 1)$ is again a sequence of i.i.d.$\sim \mathcal{N}(0,1)$ random variables. Under additional regularity assumptions on the function $g$, it can be shown that in this case,

$$|\mathbb{E}(h(X_T^{(N)})) - \mathbb{E}(h(X_T))| \leq \frac{C_T}{N},$$

which has the same order as that obtained for ODE's with the classical Euler scheme (but here, a higher order expansion is needed in order to obtain the result).