1

STOCHASTIC PROCESSES

This appendix provides a very basic introduction to the language of probability theory and stochastic processes. We assume the reader is familiar with the general measure and integration theory, we start this chapter with the notions that are specific to Probability Theory.

1.1 Random variables, laws

A *probability space* is a triple $(\Omega, \mathcal{F}, P)$ where Ω is a set, $\mathcal{F}$ is a σ-field of subsets of Ω and P is a probability measure on $\mathcal{F}$. A subset A of Ω is *negligible* if there exists $B \in \mathcal{F}$ such that $A \subset B$ and $P(B) = 0$. Note that this notion is relative to $\mathcal{F}$ and P but we shall not mention them. The probability space $(\Omega, \mathcal{F}, P)$ is called *complete* if $\mathcal{F}$ contains all the negligible sets.

An incomplete probability space $(\Omega, \mathcal{F}, P)$ can easily be completed in the following way. Let $\mathcal{N}$ denote the set of all negligible sets. Let $\overline{\mathcal{F}}$ be the σ-algebra generated by $\mathcal{F}$ and $\mathcal{N}$. It is easy to see that it coincides with the set of subsets of Ω which are of the form $B \cup N$ for some $B \in \mathcal{F}$ and $N \in \mathcal{N}$. One extends P to a probability measure $\overline{P}$ on $\overline{\mathcal{F}}$ by putting $\overline{P}(B \cup N) = P(B)$. The probability space $(\Omega, \overline{\mathcal{F}}, \overline{P})$ is then complete. It is called the *completion* of $(\Omega, \mathcal{F}, P)$. From now on, all our probability spaces are implicitely assumed to be complete.

A *random variable* is a measurable map $X : \Omega \mapsto \mathbb{R}$ where $\mathbb{R}$ is equipped with its usual Borel σ-field $\mathcal{B}(\mathbb{R})$. The *law*, or *distribution*, of a random variable X is the (probability) measure $\mu_X = X \circ P$ on $\mathbb{R}$, the image of P under X, that is,

$$\mu_X(A) = P\left(X^{-1}(A)\right)$$

for all $A \in \mathcal{B}(\mathbb{R})$. Very often, the above quantity is denoted by $P(X \in A)$.

The *expectation* of an integrable random variable X is the integral

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, dP(\omega) = \int_{\mathbb{R}} x\, d\mu_X(x).$$

The *variance* of a square integrable random variable X is the quantity

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

We have defined random variables as being real-valued only, but we will often consider complex-valued functions of random variables. If X is a real-valued random variable and if $f = g + ih$ is a measurable function from $\mathbb{R}$ to $\mathbb{C}$ (where g and h are the real and the imaginary parts of f respectively), then one extends the expectation linearly by

$$\mathbb{E}[f(X)] = \mathbb{E}[g(X)] + i\,\mathbb{E}[h(X)],$$

when this is well-defined.

The *characteristic function* of a random variable X is the Fourier transform of its law, that is, the function $\widehat{\mu}_X$ on $\mathbb{R}$ defined by

$$\widehat{\mu}_X(t) = \mathbb{E}\left[e^{itX}\right],\ t \in \mathbb{R}.$$

The invertibility of the Fourier transform has the following consequence.

Proposition 1.1. *The characteristic function of a random variable completely determines its law.* □

Let n be an integer greater than 1. A *n-tuple* of random variables (one says *pair* , *triple* when $n = 2, 3$ respectively) is a family $(X_1, \ldots, X_n)$ of n random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$. The *law*, or *distribution*, of a n-uple $(X_1, \ldots, X_n)$ is the probability measure $\mu_{(X_1,\ldots,X_n)}$ on $\mathbb{R}^n$, the image of P under $(X_1, \ldots, X_n)$, characterized by

$$\mu_{(X_1,\ldots,X_n)}(A_1 \times \cdots \times A_n) = P\left(\bigcap_{i=1}^{n} X_i^{-1}(A_i)\right)$$

for all $A_1, \ldots, A_n$ in $\mathcal{B}(\mathbb{R})$.

We say that n random variables $X_1, \ldots, X_n$ are *independent* if the law $\mu_{(X_1,\ldots,X_n)}$ of their n-tuple is the product measure $\mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$.

Proposition 1.2. *The random variables $X_1, \ldots, X_n$ are independent if and only if*

$$\mathbb{E}\left[f_1(X_1) \ldots f_n(X_n)\right] = \mathbb{E}\left[f_1(X_1)\right] \ldots \mathbb{E}\left[f_n(X_n)\right] \tag{1.1}$$

for all bounded Borel functions $f_1, \ldots, f_n$ on $\mathbb{R}$.

Proof. If (1.1) holds true for all bounded Borel functions $f_1, \ldots, f_n$ on $\mathbb{R}$ then taking the f_i to be indicator functions of some Borel sets ($f_i = \mathbb{1}_{A_i}$) gives

$$\begin{aligned}
\mu_{(X_1,\ldots,X_n)}(A_1 \times \ldots \times A_n) &= P\left(X_1^{-1}(A) \cap \ldots \cap X_n^{-1}(A_n)\right) \\
&= \mathbb{E}\left[\mathbb{1}_{A_1}(X_1) \ldots \mathbb{1}_{A_n}(X_n)\right] \\
&= \mathbb{E}\left[\mathbb{1}_{A_1}(X_1)\right] \ldots \mathbb{E}\left[\mathbb{1}_{A_n}(X_n)\right] \\
&= P\left(X_1^{-1}(A_1)\right) \ldots P\left(X_n^{-1}(A_n)\right) \\
&= \mu_{X_1}(A_1) \ldots \mu_{X_n}(A_n).
\end{aligned}$$

This proves the "if" part.

Conversely, if $X_1, \dots, X_n$ are independent then, by the same computation as above, we get that (1.1) holds true for $f_1, \dots, f_n$ being indicator functions. By linearity this remains true for $f_1, \dots, f_n$ being simple functions (linear combinations of indicator functions). By Lebesgue's theorem one gets that (1.1) holds true for any bounded Borel functions $f_1, \dots, f_n$, approximating them by simple functions. □

The *characteristic function* of a n-tuple $(X_1, \dots, X_n)$ is the Fourier transform of its law, that is, the function $\widehat{\mu}_{(X_1,\dots,X_n)}$ on $\mathbb{R}^n$ defined by

$$\widehat{\mu}_{(X_1,\dots,X_n)}(t_1, \dots, t_n) = \mathbb{E}\left[e^{i(t_1X_1+\dots+t_nX_n)}\right].$$

In the same way as for Proposition 1.1 we get the following.

Proposition 1.3. *The characteristic function of a n-tuple completely determines its law.* □

For a n-tuple $(X_1, \dots, X_n)$, we call *marginal laws*, or simply *marginals*, the laws of the (strict) sub-families $(X_{i_1}, \dots, X_{i_p})$ of $(X_1, \dots, X_n)$.

Proposition 1.4. *The law of a n-tuple $(X_1, \dots, X_n)$ completely determines its marginals.*

Proof. Consider a sub-family $(X_{i_1}, \dots, X_{i_p})$ taken from $(X_1, \dots, X_n)$. For any family of Borel sets $A_{i_1}, \dots, A_{i_p}$, define the sets $B_1, \dots, B_n$ by $B_j = A_{i_k}$ if $j = i_k$ for some k and $B_j = \mathbb{R}$ otherwise. We then have

$$\mu_{(X_{i_1},\dots,X_{i_p})}(A_{i_1} \times \dots \times A_{i_p}) = \mu_{(X_1,\dots,X_n)}(B_1 \times \dots \times B_n),$$

that is, the claim. □

Note that the converse of Proposition 1.4 is not true in general: the marginal laws do not determine the law of the whole n-tuple. For example, the law of a pair (X, Y) is not in general determined by the individual laws of X and Y, unless they are independent.

1.2 Basic examples

The very first non-trivial probability laws are the *Bernoulli laws*. This term is characteristic of any probability law μ on $\mathbb{R}$ which is supported by no more than two points. That is, probability measures μ such that $\mu(\{a\}) = p$ and $\mu(\{b\}) = 1 - p$ for some $a, b \in \mathbb{R}$ and some $p \in [0, 1]$. This law is denoted by

$$\mu = p\,\delta_a + (1 - p)\,\delta_b.$$

A way to concretely realise such a random variable is the following.Take $\Omega = \{0, 1\}$, define the measure $P(\{1\}) = p$ and $P(\{0\}) = 1 - p$ on Ω, the random

variable $X(1) = a, X(0) = b$ follows the law μ above. When $a = 1$ and $b = 0$ above, the measure P is called the *Bernoulli law with parameter p*.

More interesting is the way one can construct a *Bernoulli sequence*, that is, a sequence of random variables (X_n) on some probability space $(\Omega, \mathcal{F}, P)$, such that the X_n are independent and each of them is a Bernoulli random variable (whose law may vary with n). Put $\Omega = \{0,1\}^{\mathbb{N}}$, the set of sequences with values in $\{0,1\}$. Let $\mathcal{F}$ be the cylinder σ-field of Ω, that is, the σ-field generated by all the subsets $A \subset \Omega$ which are of the form

$$A = \{\omega \in \Omega\,;\, \omega_{n_1} = i_1, \ldots, \omega_{n_k} = i_k\}$$

for some $n_1, \ldots, n_k \in \mathbb{N}$, some $i_1, \ldots, i_k \in \{0,1\}$ and some $k \in \mathbb{N}$. Let P_n be the Bernoulli law on $\{0,1\}$ with parameter p_n, for all $n \in \mathbb{N}$. Let $P = \otimes_{n\in\mathbb{N}} P_n$ be the measure on $\mathcal{F}$ defined by

$$P(A) = P_{n_1}(\{i_1\}) \ldots P_{n_k}(\{i_k\})$$

if A is as above. Let ν_n be the coordinate mappings on Ω, that is, $\nu_n(\omega) = \omega_n$ for all $\omega \in \Omega$. Then it is easy to check that the ν_n are independent random variables on $(\Omega, \mathcal{F}, P)$ whose respective laws are $p_n\,\delta_1 + (1 - p_n)\,\delta_0$. Putting $X = a_n\nu_n + b_n$ will furnish any desired Bernoulli sequence (X_n).

Another fundamental law in probability theory is the *exponential law* and its naturally associated *Poisson law*. The exponential law is the "memory-lack law". Consider a positive random variable T which is a waiting time, say, such that, knowing that a period of duration s has passed without the expected event occuring, does not change the law of the remaining time to wait. In other words, the probability that $T > s + h$, knowing that $T > s$ already, should be the same as the probability that $T > h$. The first probability is $P(T > s + h)/P(T > s)$ and thus our assumption can be translated into

$$P(T > s + h) = P(T > s)P(T > h)$$

for all $s, h \geq 0$. It is well-known that the only measurable locally bounded real functions u defined on $\mathbb{R}^+$ such that $u(s + h) = u(s)u(h)$ for all $s, h \in \mathbb{R}^+$ are the function 0 or the functions $u(t) = e^{-\lambda t}$. Thus the distribution of our waiting time, if non-trivial, is of the form

$$P(T > t) = \mathbb{1}_{\mathbb{R}^+}(t)\, e^{-\lambda t}$$

for some $\lambda > 0$. This is the *exponential law* with parameter λ.

Now, imagine a sequence of independent random variables $X_1, X_2, \ldots$ each of which follows the exponential law with parameter λ. We think of this sequence as successive intervals of time for the occurences of some events (clients arriving, phone calls etc.). Define the random variable

$$N = \inf\{n\,;\, X_1 + X_2 + \ldots + X_n \geq 1\}\,,$$

that is, the number of arrivals during the time interval $[0, 1]$.

Proposition 1.5. *For every $k \in \mathbb{N}$ we have*

$$P(N = k) = \frac{\lambda^k}{k!}\, e^{-\lambda}\,.$$

Proof. The probability $P(N = k)$ is equal to

$$P(X_1 + \ldots + X_k \leq 1 < X_1 + \ldots X_{k+1})\,.$$

By the independence of the X_i's, this is equal to

$$\int_{x_1+\ldots+x_k\leq 1<x_1+\ldots x_{k+1}} \lambda^{k+1}\, e^{-\lambda x_1} \ldots e^{-\lambda x_{k+1}}\, dx_1 \ldots dx_{k+1}\,.$$

With the change of variables $s_1 = x_1$, $s_2 = x_1 + x_2$, $\ldots$, $s_{k+1} = x_1 + \ldots + x_{k+1}$ this gives

$$\lambda^k \int_{s_k\leq 1<s_{k+1}} \lambda e^{-\lambda s_{k+1}}\, ds_1 \ldots ds_{k+1} = \frac{\lambda^k}{k!} e^{-\lambda}. \quad \square$$

The law for N above is called the *Poisson law* with parameter λ.

A similar computation shows that the random variable $N(t)$ representing the number of arrivals during the interval $[0, t]$ follows a Poisson law with parameter λt.

The last fundamental law we shall describe here is the *Normal law* , or *Gaussian law* . A random variable is said to follow the normal law $\mathcal{N}(m, \sigma)$ if its probability law is of the form

$$\frac{1}{\sigma\sqrt{2\pi}}\, \exp\left(\frac{(x-m)^2}{2\sigma^2}\right) dx$$

for some $m \in \mathbb{R}$ and $\sigma > 0$. The expectation of such a random variable is m and its variance is σ.

Such distributions are central in probability theory for they have many interesting properties and they are the laws to which one converges by the Central Limit Theorem. We do not wish (and do not need) to develop all these properties here; the only point we want to develop here is the multidimensional extension of the notion of Gaussian random variable.

For a n-tuple $(X_1, \ldots, X_n)$ of random variables (or a n-dimensional random variable), we let $\mathbb{E}[X]$ be the n-dimensional vector $(\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n])$ and we let $\mathrm{Var}(X)$ be the so-called *covariance matrix* :

$$\left(\mathbb{E}\left[(X_i - \mathbb{E}[X_i])\,(X_j - \mathbb{E}[X_j])\right]\right)_{i,j} = \left(\mathbb{E}\left[X_i\, X_j\right] - \mathbb{E}[X_i]\,\mathbb{E}[X_j]\right]\right)_{i,j}\,.$$

The matrix $\mathrm{Var}(X)$ is always (real) positive semi-definite; it is positive definite if the law of $(X_1, \ldots, X_n)$ is non-degenerate.

Let Q be a $n \times n$ real, positive definite matrix. Define the associated quadratic form

$$q(x) = \langle x\,,\, Qx \rangle$$

on $\mathbb{R}^n$. A family $(X_1, \ldots, X_n)$ of random variables is said to be a *Gaussian family* if its n-tuple law is of the form

$$\gamma^{-1} \exp\left(-\frac{1}{2}\, q(x-a)\right) \tag{1.2}$$

for some Q as above, some $a \in \mathbb{R}^n$ and some $\gamma \in \mathbb{R}$. Straightforward computations show that:

– The constant γ needs to be equal to $(2\pi)^{n/2}\,\sqrt{\det Q^{-1}}$ for (1.2) to define the density of a probability measure;

– The expectation of $(X_1, \ldots, X_n)$ is a and its covariance matrix is Q^{-1}.

One of the remarkable point with Gaussian families is that independence can be read easily, as the following theorem proves it.

Theorem 1.6. *A Gaussian family $(X_1, \ldots, X_n)$ is made up of independent random variables if and only if its covariance matrix is diagonal.*

Proof. One direction is easy: if the random variables X_i are independent then the covariances $\mathbb{E}\big[X_i\, X_j\big] - \mathbb{E}[X_i]\,\mathbb{E}[X_j]$ vanish for all $i \neq j$. Conversely, if V is diagonal then the density

function (1.2) factorizes as a product of Gaussian densities in each of the variables x_i. This shows independence. □

Consider any invertible linear transformation A of $\mathbb{R}^n$. Put $B = A^{-1}$. Consider the random variables $(Y_1, \ldots, Y_n) = A(X_1, \ldots, X_n)$. The following is also a straighforward change of variable argument.

Proposition 1.7. *The family $(Y_1, \ldots, Y_n) = A(X_1, \ldots, X_n)$ is again a Gaussian family, but with associated matrix $Q' = B^* Q B$ and expectation Aa.* □

More important is the following.

Theorem 1.8. *All the sub-families of a Gaussian family is again a Gaussian family.*

Proof. The coefficient $q_{n,n}$ of the matrix Q associated to $(X_1, \ldots, X_n)$ by (1.2) is not null, for otherwise the integral of the density would diverge in the variable x_n. This implies that the linear transformation $(y_1, \ldots, y_n) = A(x_1, \ldots, x_n)$, given by $y_i = x_i$ for $i = 1, \ldots, n-1$ and $y_n = q_{1,n} x_1 + \ldots + q_{n,n} x_n$, is invertible. Furthermore, it is possible to write the quadratic form q in the form

$$q(x) = \frac{1}{q_{n,n}} y_n^2 + q'(y)$$

where q' is a quadratic form depending only on the variables $y_1, \ldots, y_{n-1}$ of $y = (y_1, \ldots, y_n)$. Now, applying Proposition 1.7, the density of $(Y_1, \ldots, Y_n)$ factorises and the laws of $(Y_1, \ldots, Y_{n-1}) = (X_1, \ldots, X_{n-1})$ and Y_n are again seen to be normal. □

We can now prove the following fundamental characterization of Gaussian families.

Theorem 1.9. *A family $(X_1, \ldots, X_n)$ of random variables is Gaussian if and only if any linear combination $Z = \alpha_1 X_1 + \ldots \alpha_n X_n$ is a Gaussian random variable.*

Proof. The "only if" part is an immediate corollary of Theorem 1.8. Let us prove the "if" part. If any linear combination $Z = a_1 X_1 + \ldots a_n X_n$ is Gaussian then the characteristic function of $(X_1, \ldots, X_n)$ must be of the form

$$\mathbb{E}\left[\exp\left(i\sum_i a_i X_i\right)\right] = \exp\left[im(a) - \frac{1}{2}\sigma(a)^2\right]$$

where $a = (a_1, \ldots, a_n)$, where $m(a) = \sum_i a_i \mathbb{E}[X_i]$ and $\sigma(a)^2 = \mathbb{E}\left[(\sum_i a_i X_i)^2\right] - m(a)^2$.

Put $m_i = \mathbb{E}[X_i]$ and $m = (m_1, \ldots, m_n)$. Let $V_{i,j} = \mathbb{E}\left[(X_i - m_i)(X_j - m_j)\right]$ be the coefficients of the covariance matrix V of $(X_1, \ldots, X_n)$. Assume that the matrix V is non-degenerate (otherwise this would reduce the number of pertinent variables). Putting all this together, we easily get that the distribution of $(X_1, \ldots, X_n)$ is given by the density

$$(2\pi)^{n/2}\sqrt{\det V^{-1}}\,\exp\left[-\frac{1}{2}(x-m)V^{-1}(x-m)^t\right].$$

This proves the "if" part and the theorem. □

1.3 Stochastic processes

The main goal of stochastic calculus is to study "stochastic processes", that is, families $(X_t)_{t\in I}$ of random variables X_t indexed by a subset I of $\mathbb{R}$ (in the sequel we shall only consider the cases where I is $\mathbb{R}^+$ or a sub-interval of $\mathbb{R}^+$). We shall give a rigorous framework for the study of stochastic processes and their laws.

Let I be a subset of $\mathbb{R}$. A *process* (or *stochastic process*) *indexed by* I is a family $(X_t)_{t\in I}$ of random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$. When I is $\mathbb{R}^+$ one simply says a *process* and $(X_t)_{t\in I}$ may be denoted $(X_t)_{t\geq 0}$ or just $X_.$.

For each finite subset $T = \{t_1, \ldots, t_n\}$ of I one denotes by $\mu_{X,T}$ the law of the n-tuple $(X_{t_1}, \ldots, X_{t_n})$. If $(X_t)_{t\in I}$ and $(Y_t)_{t\in I}$ are two processes indexed

by the same set I and if $\mu_{X,T} = \mu_{Y,T}$ for all finite subsets T of I, we say that $(X_t)_{t\in I}$ and $(Y_t)_{t\in I}$ are *versions* of each other.

Let I be a subset of $\mathbb{R}$. Suppose we are given for any finite subset T of I a probability measure μ^T on $\mathbb{R}^{|T|}$ (where $|T|$ denotes the cardinal of T). It is natural to wonder if one can construct a process $(X_t)_{t\in I}$ such that $\mu_{X,T} = \mu^T$ for all T. The answer is given by the well-known Kolmogorov Consistency Theorem that we state without proof.

Let I be a set. Let $(\Omega, \mathcal{F})$ be a measurable space. Suppose that for all finite subsets T of I we are given a probability measure μ^T on $(\prod_{t\in T} \Omega, \bigotimes_{t\in T} \mathcal{F})$. The family $(\mu^T)_T$ is called *consistent* if, for all finite subsets T_1, T_2 of I such that $T_1 \subset T_2$, the restriction of μ^{T_2} to the σ-algebra $\bigotimes_{t\in T_1} \mathcal{F}$ coincides with μ^{T_1}.

Theorem 1.10 (Kolmogorov's consistency theorem). *Let $(\Omega, \mathcal{F})$ be a Polish space together with its Borel σ-algebra. Let I be a subset of $\mathbb{R}$. If for all finite subsets T of I there exists a probability measure μ^T on $(\prod_{t\in T} \Omega, \bigotimes_{t\in T} \mathcal{F})$ such that the family $(\mu^T)_T$ is consistent, then there exists a unique probability measure μ on $\left(\prod_{t\in I} \Omega, \bigotimes_{t\in I} \mathcal{F}\right)$ which extends all the probability measures μ^T.* □

Consider a process $(X_t)_{t\in I}$ on $(\Omega, \mathcal{F}, P)$. Let $\mathbb{R}^I = \prod_{t\in I} \mathbb{R}$ and let μ_X be the unique probability measure on $\left(\mathbb{R}^I, \mathcal{B}(\mathbb{R}^I)\right)$ which extends the consistent family of probability measures $\{\mu_{X,T};\ T \subset I,\ T \text{ finite}\}$. This measure μ_X on $\left(\mathbb{R}^I, \mathcal{B}(\mathbb{R}^I)\right)$ is called the *law* of the process $(X_t)_{t\in I}$.

Proposition 1.11. *Two processes $(X_t)_{t\in I}$ and $(Y_t)_{t\in I}$ are versions of each other if and only if they have the same law.*

Proof. Straightforward. □

Let $(X_t)_{t\in I}$ be a process on $(\Omega, \mathcal{F}, P)$. For all $t \in I$, define the function Y_t on $\mathbb{R}^I$ by $Y_t(x) = x(t)$, that is Y_t is the t-th coordinate mapping. This defines a process $(Y_t)_{t\in I}$ on $\left(\mathbb{R}^I, \mathcal{B}(\mathbb{R}^I), \mu_X\right)$ and we have

$$\begin{aligned}\mu_{Y,T}(A_{t_1} \times \cdots \times A_{t_n}) &= \mu_X\left(Y_{t_1}^{-1}(A_{t_1}) \cap \cdots \cap Y_{t_n}^{-1}(A_{t_n})\right)\\ &= \mu_X(\{x \in \mathbb{R}^I\,;\, x(t_1) \in A_{t_1}, \ldots, x(t_n) \in A_{t_n}\})\\ &= \mu_{X,T}(A_{t_1} \times \cdots \times A_{t_n}).\end{aligned}$$

Thus $(Y_t)_{t\in I}$ is a version of $(X_t)_{t\in I}$. This version is called the *canonical version* of $(X_t)_{t\in I}$. The probability space $(\mathbb{R}^I, \mathcal{B}(\mathbb{R}^I), \mu_X)$ is called the *canonical space* of $(X_t)_{t\in I}$.

From now on, unless otherwise precised, all the processes are indexed by $\mathbb{R}^+$.

Two processes $X_\cdot$ and $Y_\cdot$ defined on the same probability space are *modifications* of each other if, for all $t \in \mathbb{R}^+$, X_t equals Y_t almost surely. They are *indistinguishable* if, for almost all $\omega \in \Omega$, $X_t(\omega)$ equals $Y_t(\omega)$ for all t.

A subset A of $\mathbb{R}^+ \times \Omega$ is *evanescent* if there exists a negligible subset B of Ω such that $A \subset \mathbb{R}^+ \times B$.

From all these definitions one easily checks the following properties.

Proposition 1.12.

a) If $X_.$ and $Y_.$ are indistinguishable then they are modifications of each other.

b) If $X_.$ and $Y_.$ are modifications of each other then they are versions of each other.

c) $X_.$ and $Y_.$ are indistinguishable if and only if the set of (ω, t) such that $X_t(\omega) \neq Y_t(\omega)$ is evanescent. □

Let $X_.$ be a process. For all $\omega \in \Omega$, the mapping $t \mapsto X_t(\omega)$ defines a function on $\mathbb{R}^+$. These functions are called the *paths* of the process $X_.$. One says that the process $X_.$ has *continuous paths* (or simply, is *continuous*) if for almost all $\omega \in \Omega$ the path $X_.(\omega)$ is continuous on $\mathbb{R}^+$. In the same way one defines *right-continuous processes* , *left-continuous processes* (we take the convention that any process is left-continuous at 0). A process is said to be *càdlàg* [1] if its paths are right-continuous and admit left limits at all points.

Proposition 1.13. *Let $X_.$ and $Y_.$ be two right-continuous (resp. left-continuous) processes. If they are modifications of each other then they are indistinguishable.*

Proof. There exists a negligible set $\mathcal{N}$ such that for all $\omega \notin \mathcal{N}$ the paths $X_.(\omega)$ and $Y_.(\omega)$ are right-continuous and $X_t(\omega) = Y_t(\omega)$ for all t rational. Thus passing to the limit we have $X_t(\omega) = Y_t(\omega)$ for all t. □

A process $X_.$ is *measurable* if it is measurable as a mapping from $\mathbb{R}^+ \times \Omega$ to $\mathbb{R}$, where $\mathbb{R}^+ \times \Omega$ is equipped with the σ-algebra $\mathcal{B}(\mathbb{R}^+) \otimes \mathcal{F}$.

We end up this section with a notion which is fundamental, in particular for the study of martingales (cf Chapter **??**). Let $U = \{X_i\,; i \in I\}$ be any family of integrable random variables on $(\Omega, \mathcal{F}, P)$, one can think of U as a subset of $L^1(\Omega, \mathcal{F}, P)$. The family U is *uniformly integrable* if

$$\sup_{X \in U} \mathbb{E}\left[|X|\, \mathbb{1}_{|X| \geq a}\right] \xrightarrow[a \to +\infty]{} 0. \tag{1.3}$$

Proposition 1.14. *Let U be a subset of $L^1(\Omega, \mathcal{F}, P)$. The following assertions are equivalent.*

i) U is uniformly integrable.

ii) One has

[1] The name càdlàg comes from the french "continu à droite avec des limites à gauche" which means "right-continuous with left-limits" (the english version of càdlàg should then be "rcll").

a) $\sup_{X\in U}\mathbb{E}\left[|X|\right]<\infty$
and

b) for all $\varepsilon>0$*, there exists a* $\delta>0$ *such that* $A\in\mathcal{F}$ *and* $P(A)\leq\delta$ *imply* $\mathbb{E}\left[|X|\,\mathbb{1}_A\right]\leq\varepsilon$*, for all* $X\in U$*.*

Proof. If U is uniformly integrable and $A\in\mathcal{F}$ then

$$\begin{aligned}\sup_{X\in U}\mathbb{E}\left[|X|\,\mathbb{1}_A\right] &= \sup_{X\in U}\mathbb{E}\left[|X|\,\mathbb{1}_{|X|\geq a}\mathbb{1}_A\right]+\sup_{X\in U}\mathbb{E}\left[|X|\,\mathbb{1}_{|X|<a}\mathbb{1}_A\right]\\ &\leq \sup_{X\in U}\mathbb{E}\left[|X|\,\mathbb{1}_{|X|\geq a}\mathbb{1}_A\right]+a\sup_{X\in U}\mathbb{E}\left[\mathbb{1}_{|X|<a}\mathbb{1}_A\right].\end{aligned}$$

Let c be such that the first term above is smaller than $\varepsilon/2$ for all $a\geq c$. We then have $\sup_{X\in U}\mathbb{E}\left[|X|\,\mathbb{1}_A\right]\leq\varepsilon/2+cP(A)$. Taking $A=\Omega$ gives a*)*. Taking $\delta=\varepsilon/2c$ gives b). We have proved that *i)* implies *ii)*.

Conversely, assume *ii)* is satisfied. Let $\varepsilon>0$ and let $\delta>0$ be as in b). Let $c=\sup_{X\in U}\mathbb{E}[|X|]/\delta<\infty$. Let A be the event $(|X|\geq c)$. We have $P(A)\leq\mathbb{E}[|X|]/c\leq\delta$. Thus $\mathbb{E}\left[|X|\,\mathbb{1}_A\right]\leq\varepsilon$ for all $X\in U$. □

One of the main application of the notion of uniform integrability is the following.

Theorem 1.15. *Let* (X_n) *be a sequence of random variables belonging to* $L^1(\Omega,\mathcal{F},P)$*. Suppose that* (X_n) *converges almost surely to a random variable* $X_\infty\in L^1(\Omega,\,\mathcal{F},\,P)$*. Then* (X_n) *converges to* X_∞ *in* $L^1(\Omega,\mathcal{F},P)$ *if and only if* (X_n) *is uniformly integrable.*

Proof. If (X_n) converges to X_∞ in $L^1(\Omega,\mathcal{F},P)$ we then have

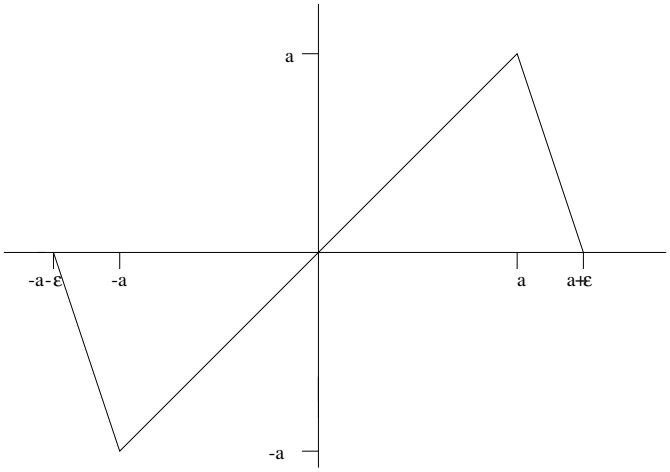
$$\sup_n\mathbb{E}[|X_n|]\leq\sup_n\mathbb{E}[|X_n-X_\infty|]+\mathbb{E}[|X_\infty|]<+\infty$$

and

$$\mathbb{E}[|X_n|\,\mathbb{1}_A]\leq\mathbb{E}[|X_\infty|\,\mathbb{1}_A]+\mathbb{E}[|X_n-X_\infty|\,\mathbb{1}_A]\,.$$

The second term of the right hand side is dominated by ε for n large enough (independently of A); as any finite sequence of random variables is always uniformly integrable, the conclusion follows.

Conversely, if (X_n) is uniformly integrable, let $\varepsilon>0$ and a_ε be such that if $a\geq a_\varepsilon$ then $\sup_n\mathbb{E}[|X_n|\,\mathbb{1}_{|X_n|\geq a}]\leq\varepsilon/3$. Let ϕ_a be the following function.



By Fatou's lemma we have $\mathbb{E}[|X_\infty|\mathbb{1}_{|X_\infty|\geq a}] \leq \varepsilon/3$ and

$$\begin{aligned}\mathbb{E}\left[|X_n - X_\infty|\right] &\leq \mathbb{E}\left[|X_n - \phi_a(X_n)|\right] + \mathbb{E}\left[|\phi_a(X_n) - \phi_a(X_\infty)|\right] \\ &\qquad + \mathbb{E}\left[|\phi_a(X_\infty) - X_\infty|\right] \\ &\leq \mathbb{E}\left[|X_n|\mathbb{1}_{|X_n|\geq a}\right] + \mathbb{E}\left[|\phi_a(X_n) - \phi_a(X_\infty)|\right] \\ &\qquad + \mathbb{E}\left[|X_\infty|\mathbb{1}_{|X_\infty|\geq a}\right] .\end{aligned}$$

As $\mathbb{E}[|\phi_a(X_n) - \phi_a(X_\infty)|]$ tends to 0 by Lebesgue's theorem, this gives the result. □

1.4 Brownian motion

In this section and in the next one, we shall consider two fundamental examples of stochastic processes: *Brownian motion* and the Poisson process. They are fundamental for they are cornerstones of classical stochastic calculus theory, but they also appear to be key processes in the probabilistic interpretations of quantum stochastic calculus.

Let us first study Brownian motion. We wish to construct a process $(W_t)_{t\in\mathbb{R}^+}$ with values in $\mathbb{R}$ such that

1) for all $s < t$, the random variable $W_t - W_s$ is independent of the random variables W_u, $u \leq s$),

2) for all $t \geq 0$, the random variable W_t admits the Gaussian $\mathcal{N}(0, t)$ probability law.

Let us construct such a process. The following is an easy extension of the construction we described for constructing Bernoulli random walks.

Theorem 1.16. *For any given probability measure μ on $\mathbb{R}$, there exists a probability space $(\Omega, \mathcal{F}, P)$ and a sequence (X_n) of random variables on Ω*

such that the random variables X_n are independent and each have the law μ. □

We can now prove the following.

Theorem 1.17. *Let $\mathcal{H}$ be a separable Hilbert space. There exists a probability space $(\Omega, \mathcal{F}, P)$ and a family $X(h)$, $h \in \mathcal{H}$, of random variables on Ω such that*

i) the mapping $h \longmapsto X(h)$ is linear,

ii) each random variable $X(h)$ follows the $\mathcal{N}(0, \|h\|)$ law.

Proof. Let (e_n) be an orthonormal basis of $\mathcal{H}$. By Theorem 1.16, there exists a probability space $(\Omega, \mathcal{F}, P)$ and a sequence (Z_n) of independent random variables on Ω, with individual law $\mathcal{N}(0, 1)$.

For any $h \in \mathcal{H}$, put $X(h) = \sum_n \langle e_n\,,\, h\rangle\, Z_n$. This series is convergent in $L^2(\Omega, \mathcal{F}, P)$ and defines a random variable $X(h)$ on Ω. The family $X(h)$, $h \in \mathcal{H}$, satisfies the assumptions of the theorem, as can easily be checked. □

Note that, by construction we have

$$\langle X(h')\,,\, X(h)\rangle_{L^2(\Omega,\mathcal{F},P)} = \langle h'\,,\, h\rangle\,.$$

Also note that the set $\{X(h),\, h \in \mathcal{H}\}$ is a *Gaussian subspace* of $L^2(\Omega, \mathcal{F}, P)$, that is, any linear combination of the $X(h)$ is a Gaussian random variable. By Theorem 1.9, this means that every finite family $(X(h_1), \ldots, X(h_n))$ is Gaussian.

Now the construction of a Brownian motion is easy. Take $\mathcal{H} = L^2(\mathbb{R}^+)$ and construct a family $X(h)$, $h \in \mathcal{H}$, as in Theorem 1.17. We claim that the process $W_t = X(\mathbb{1}_{[0,t]})$, $t \in \mathbb{R}^+$, is a Brownian motion. Indeed, the random variable W_t follows the law $\mathcal{N}(0, t)$ by ii) above and thus satisfies condition 2) in the definition of Brownian motion; furthermore, we have, for all $u \le s \le t$

$$\langle W_t - W_s\,,\, W_u\rangle_{L^2(\Omega,\mathcal{F},P)} = \langle \mathbb{1}_{[s,t]}\,,\, \mathbb{1}_{[0,u]}\rangle_{\mathcal{H}} = 0\,.$$

But the pair $(W_t - W_s, W_u)$ is Gaussian, hence by Theorem 1.6 they are independent random variables. This gives the condition 1) in the definition of a Brownian motion.

We shall now review some of the main properties of Brownian motion paths. First of all, we shall establish the continuity of the Brownian paths. Actually, we shall establish a stronger property: the Hölder continuity of order α for every $\alpha < 1/2$. This property is based on a general criterion due to Kolmogorov that we state here without proof.

Theorem 1.18. [Kolmogorov criterion] *Let $X.$ be a process such that there exists strictly positive constants γ, C and ε such that*

$$\mathbb{E}[|X_t - X_s|^\gamma] \le C|t - s|^{1+\varepsilon}\,.$$

Then there exists a modification of $X.$ whose paths are Hölder continuous of order α for every $\alpha \in [0, \varepsilon/\gamma]$.

Once this is admited, the fact that the increments of the Brownian motion are Gaussian gives

$$\mathbb{E}\left[(W_t - W_s)^{2p}\right] = C_p|t-s|^p$$

for every $p > 0$. This immediately yields the following.

Theorem 1.19. *Up to modification, the paths of Brownian motion are locally Hölder continuous of order α for every $\alpha < 1/2$. In particular they are continuous.* □

Another very important property of the Brownian motion is that it admits a non-trivial quadratic variation.

Theorem 1.20. *For any $t \in \mathbb{R}^+$, the quantity*

$$\sum_{i;t_i \le t} \left(W_{t_{i+1}} - W_{t_i}\right)^2$$

converges to t, in the L^2 sense when the diameter δ of the partition $\{t_i\,;\, i \in \mathbb{N}\}$ tends to 0.

Proof. Let us compute the L^2-norm of the difference (for simplicity, we assume that the partition ends at t):

$$\begin{aligned}\left\| \sum_{i;t_i \le t} (W_{t_{i+1}} - W_{t_i})^2 - t \right\|^2 &= \mathbb{E}\left[\left(\sum_{i;t_i \le t} (W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i) \right)^2 \right] \\ &= \sum_{i;t_i \le t} \mathbb{E}\left[\left((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i) \right)^2 \right],\end{aligned}$$

where we used the independence of the increments and the fact that

$$\mathbb{E}[(W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i)] = 0\,.$$

Furthermore, for a Gaussian random variable Y with mean 0, we have $\mathbb{E}[Y^4] = 3\mathbb{E}[Y^2]^2$. This gives

$$\left\| \sum_{i;t_i \le t} (W_{t_{i+1}} - W_{t_i})^2 - t \right\|^2 = 2 \sum_{i;t_i \le t} (t_{i+1} - t_i)^2$$

which converges to 0 with the diameter of the partition. □

It is actually possible to prove the same result for almost sure convergence, but we do not give a proof of that fact here.

This quadratic variation property has many important consequences that we shall develop in part II of this book, but one very important application of it is the following.

Theorem 1.21. *The paths of the Brownian motion are almost surely of infinite variation on any interval.*

The paths of the Brownian motion are almost surely nowhere locally Hölder continuous of order α for $\alpha > 1/2$.

In particular, the paths of the Brownian motion are almost surely nowhere differentiable.

Proof. Note that almost surely we have

$$\sum_{i;t_i\leq t}(W_{t_{i+1}}(\omega)-W_{t_i}(\omega))^2 \leq$$
$$\leq \left(\sup_{i;t_i\leq t}|W_{t_{i+1}}(\omega)-W_{t_i}(\omega)|\right)\left(\sum_{i;t_i\leq t}|W_{t_{i+1}}(\omega)-W_{t_i}(\omega)|\right).$$

The first term in the right hand side converges to 0 by the continuity of Brownian motion. The second term is dominated by the total variation of the Brownian path. As the left hand side converges to a finite quantity, this forces the total variation to be infinite.

The case of the non-Hölder property is treated by following the same idea: for all $\alpha > 1/2$ we have

$$\sum_{i;t_i\leq t}(W_{t_{i+1}}(\omega)-W_{t_i}(\omega))^2 \leq$$
$$\leq t\left(\sup_{i;t_i\leq t}|t_{i+1}-t_i|^{2\alpha-1}\right)\left(\sup_{i;t_i\leq t}\frac{|W_{t_{i+1}}(\omega)-W_{t_i}(\omega)|^2}{|t_{i+1}-t_i|^{2\alpha}}\right).$$

If the Brownian paths were Hölder of order α the last term above would be dominated independently of the partition. The rest of the right hand side converges to 0. This contradicts the fact that the left hand side converges to t. This proves the non Hölderian character of Brownian motion for $\alpha > 1/2$.

Non-differentiability is immediate. □

We have not yet said if the Brownian paths are Hölder-continuous of order $1/2$ or not. It so happens that they are not, but this result needs further developments; we just mention it as a remark.

1.5 The Poisson process

We now construct the *Poisson process*. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let (T_n) be a strictly increasing sequence of positive random variables. The T_n's are thought as arrival times. A process $(X_t)_{t\in\mathbb{R}^+}$ such that, for all $t \in \mathbb{R}^+$

$$X_t = \sum_n \mathbb{1}_{T_n \leq t}$$

is called a *counting process associated to* (T_n). It is valued in $\mathbb{N} \cup \{+\infty\}$. If $\sup_n T_n = \infty$ a.s. one says that $(X_t)_{t\in\mathbb{R}^+}$ is a *non-exploding counting process*.

A *Poisson process* is a non-exploding counting process $(N_t)_{t\in\mathbb{R}^+}$ whose increments are independent and stationary. That is,

1) $N_t - N_s$ is independent of all the random variables N_u, $u \leq s$

2) $N_t - N_s$ has the same law as $N_{t+h} - N_{s+h}$ for all $t \geq s \geq 0$ and $h \geq 0$.

Theorem 1.22. *Poisson processes exist and they are all of the following form: there exists a $\lambda \in \mathbb{R}^+$ such that*

$$P(N_t = n) = \frac{(\lambda t)^n}{n!}\, e^{-\lambda t} \tag{1.4}$$

for all $n \in \mathbb{N}$. In other words, the associated sequence (T_n) consists of sums of independent times each of which follows an exponential distribution with parameter λ.

Proof. One direction is easy. If (T_n) is the sequence given by $T_n = \sum_{m\leq n} S_m$ where the S_m's are independent, identically distributed random variables following the exponential law with parameter $\lambda \geq 0$, it is then easy to check that the associated counting process

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

is a Poisson process and follows the Poisson law (1.4).

The converse is more difficult, we only sketch the proof. From the hypothesis we have

$$P(N_t = 0) = P(N_s = 0)P(N_t - N_s = 0) = P(N_s = 0)P(N_{t-s} = 0)$$

and thus

$$P(N_t = 0) = e^{-\lambda t}$$

for some $\lambda \geq 0$ and for all $t \in \mathbb{R}^+$.

We now claim that $P(N_t \geq 2) = o(t)$. Indeed, divide $[0, 1]$ into n intervals of the same length. Let S_n be the number of subintervals which contain at least two times of the sequence $(T_m)_{m\in\mathbb{N}}$. Clearly S_n has a binomial distribution $B(n, P(N_{1/n} \geq 2))$. Therefore $\mathbb{E}[S_n] = nP(N_{1/n} \geq 2)$. For a fixed ω, for n sufficiently large there is no interval with more than one stopping time. Thus $\lim_{n\to+\infty} S_n(\omega) = 0$ a.s. and $\lim_{n\to+\infty} \mathbb{E}(S_n) = 0$ by dominated convergence (here is the difficult argument that we skip!). This gives the announced estimate.

We now have

$$P(N_t = 1) = 1 - P(N_t = 0) - P(N_t \geq 2)$$

and thus

$$\lim_{t\to 0} \frac{1}{t} P(N_t = 1) = \lim_{t\to 0} \frac{1 - e^{-\lambda t} + o(t)}{t} = \lambda \, .$$

Finally, for $\beta \in [0,1]$ put $f(t) = \mathbb{E}[\beta^{N_t}]$. Clearly $f(t+s) = f(t) + f(s)$ and f is of the form $f(t) = e^{tg(\beta)}$. But

$$\begin{aligned} f(t) &= \sum_{n=0}^{\infty} \beta^n P(N_t = n) \\ &= P(N_t = 0) + \beta P(N_t = 1) + \sum_{n=2}^{\infty} \beta^n P(N_t = n) \end{aligned}$$

and $g(\beta) = f'(0)$. This gives

$$\begin{aligned} g(\alpha) &= \lim_{t\to 0} \frac{P(N_t = 0) - 1}{t} + \frac{\alpha P(N_t = 1)}{t} + \frac{1}{t} o(t) \\ &= -\lambda + \lambda\alpha, \end{aligned}$$

so

$$f(t) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n \beta^n}{n!}$$

and the required result follows. □

The parameter λ is called the *intensity* of $(N_t)_{t\in\mathbb{R}^+}$. In particular we have

$$\begin{aligned} \mathbb{E}[N_t] &= \lambda t \\ \mathrm{Var}[N_t] &= \lambda t \, . \end{aligned}$$