

CHAPTER 1

Review of Measure Theory

1.1. Basic Definitions

Let E denote a set and $\mathcal{P}(E)$ denote the power set of E , that is, the set of all subsets of E . In what follows we will use calligraphic letters to denote a class of subsets of E , that is, a subset of $\mathcal{P}(E)$. Moreover, the reference set E will be called a space.

DEFINITION 1.1. A σ -algebra on E is a nonempty class $\mathcal{E} \subseteq \mathcal{P}(E)$ satisfying:

- (1) $E \in \mathcal{E}$.
- (2) If $B \in \mathcal{E}$ then $B^c := E \setminus B \in \mathcal{E}$.
- (3) If $B_i \in \mathcal{E}, i = 1, 2, 3, \dots$, then $\bigcup_{i=1}^{\infty} B_i \in \mathcal{E}$.

DEFINITION 1.2. We say that a σ -algebra $\mathcal{G}$ is a sub- σ -algebra of $\mathcal{E}$ if $\mathcal{G} \subseteq \mathcal{E}$, i.e., $A \in \mathcal{G} \Rightarrow A \in \mathcal{E}$.

DEFINITION 1.3. The pair $(E, \mathcal{E})$ is called a measurable space. The elements of $\mathcal{E}$ are the measurable sets of $(E, \mathcal{E})$.

EXAMPLE 1.4. For any set E , both the power set $\mathcal{P}(E)$ and the class $\{\emptyset, E\}$ are σ -algebras, the largest and the smallest that can be defined on E , respectively.

EXAMPLE 1.5. If E is a set such that $\#E := \{\text{number of elements of } E\} < \infty$, we usually take the σ -algebra $\mathcal{P}(E)$ which has $2^{\#E}$ elements.

It is easy to show that an arbitrary intersection of σ -algebras is a σ -algebra again. This fact motivates the following definition.

DEFINITION 1.6. Given $\mathcal{G}$ any class of subsets we define $\sigma(\mathcal{G})$ the σ -algebra generated by $\mathcal{G}$ as the smallest σ -algebra containing $\mathcal{G}$, which coincides with the intersection of all σ -algebras containing $\mathcal{G}$.

DEFINITION 1.7. Let $(E, \mathcal{E})$ a measurable space. A measure μ on $(E, \mathcal{E})$ is a mapping $\mu : \mathcal{E} \rightarrow [0, +\infty]$ satisfying $\mu(\emptyset) = 0$ and

$$\mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mu(B_i), \quad (\sigma\text{-additivity}),$$

whenever $\{B_i\}_{i \geq 1}$ is a sequence of elements of $\mathcal{E}$ that are disjoint ($B_i \cap B_j = \emptyset$, if $i \neq j$).

The quantity $\mu(E)$ is called the total mass of μ . If $\mu(E) < +\infty$, we say that the measure μ is finite. If there exists $\{B_i\}_{i \geq 1}$ disjoint such that $E = \bigcup_{i=1}^{\infty} B_i$ and $\mu(B_i) < +\infty, i \geq 1$, we say that the measure μ is σ -finite. Usually we will make an abuse of notation and we will say a measure μ on E instead of a measure on $(E, \mathcal{E})$, particularly when it is clear from the context which is the σ -algebra on E .

DEFINITION 1.8. Let E be an space, $\mathcal{E}$ be a σ -algebra of E and μ be a measure on $\mathcal{E}$. The triple $(E, \mathcal{E}, \mu)$ is called a measure space.

EXAMPLE 1.9. Any measurable space with the measure $\mu \equiv +\infty$ or $\mu \equiv 0$ are measure spaces.

DEFINITION 1.10. If $\#E < \infty$, then we can define a measure μ by assigning a positive number $\mu(x)$ to each element x of the set E , that is,

$$\mu(B) = \sum_{x \in B} \mu(x), \quad B \in \mathcal{P}(E),$$

and $(E, \mathcal{P}(E), \mu)$ is a measure space.

DEFINITION 1.11. Let $(E, \mathcal{E}, \mu)$ be a measure space and $N \subset \mathcal{P}(E)$. We say that N is negligible if there exists $A \in \mathcal{E}$ such that $N \subset A$ and $\mu(A) = 0$. We denote the set of all negligible sets by $\mathcal{N}_\mu$.

DEFINITION 1.12. A measure space $(E, \mathcal{E}, \mu)$ is complete if all negligible sets are measurable.

EXAMPLE 1.13. The following are some non trivial examples of measures and measurable spaces.

- (1) **Borel measures:** Let (E, d) be a separable metric space (we will usually consider $\mathbb{R}^d$ with the Euclidian distance). Consider the Borel σ -algebra of subsets of E , denoted by $\mathcal{B}(E)$. This σ -algebra is the smallest σ -algebra containing all the open sets in E . More generally, let $S \in \mathcal{B}(E)$, its Borel σ -algebra is defined as

$$\mathcal{B}(S) = \{A \cap S : A \in \mathcal{B}(E)\}.$$

Elements of $\mathcal{B}(S)$ are called Borel sets and any measure on $(S, \mathcal{B}(S))$ is called a Borel measure. The paradigmatic example of Borel measure on $\mathbb{R}^d$ is given by the Lebesgue measure. The Lebesgue measure, denoted by λ_d , is defined by

$$\lambda_d(A) = \prod_{i=1}^d (b_i - a_i),$$

where $A = (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d)$ and $-\infty < a_i < b_i < \infty$. Note that λ_d is not finite but it is σ -finite.

- (2) **Lebesgue-Stieltjes measures:** A Lebesgue-Stieltjes measure μ on $\mathbb{R}^d$ is a Borel measure μ such that $\mu(A) < \infty$ for any $A \in \mathcal{B}(\mathbb{R}^d)$ bounded. The Lebesgue measure is a particular case of Lebesgue-Stieltjes measure. Lebesgue-Stieltjes measures are generated by distribution functions, that is, right-continuous non-decreasing functions.
- (3) **Probability measures:** Here we use the notation $(E, \mathcal{E}) = (\Omega, \mathcal{F})$. The elements of Ω represents the outcomes of a random experiment and the elements of $\mathcal{F}$ are called events. Any measure P on $(\Omega, \mathcal{F})$ with $P(\Omega) = 1$ is called a probability measure. $P(B), B \in \mathcal{F}$ is the probability that the outcome belong to the event B .
- (4) **Counting measures:** Let E be a countable set and consider the σ -algebra given by its power set $\mathcal{P}(E)$. For any $A \in \mathcal{P}(A)$ denote the cardinality of the set A (which may be infinite) by $\#A$. The counting measure is defined as $\mu(A) = \#A, A \in \mathcal{P}(E)$.

1.2. Measurable Mappings and Functions

Let $T : E \rightarrow F$ be a mapping/transformation from a space E to another space F . The inverse image $T^{-1}B$ of a subset $B \subset F$ is defined to be

$$T^{-1}B = \{x \in E : T(x) \in B\}.$$

DEFINITION 1.14. Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be two measurable spaces. An mapping $T : E \rightarrow F$ is called $(\mathcal{E}, \mathcal{F})$ -measurable if $T^{-1}B \in \mathcal{E}$ whenever $B \in \mathcal{F}$. If the σ -algebras involved are clear from the context it is called measurable.

Note that if $\mathcal{E} = \mathcal{P}(E)$ then any mapping is measurable. The following theorem simplifies checking if a transformation is measurable.

THEOREM 1.15. Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be two measurable spaces and $T : E \rightarrow F$ a mapping. Let $\mathcal{G}$ be a class of subsets of F such that $\sigma(\mathcal{G}) = \mathcal{F}$. Then, T is $(\mathcal{E}, \mathcal{F})$ -measurable if and only if $T^{-1}B \in \mathcal{E}$ for all $B \in \mathcal{G}$.

DEFINITION 1.16. Let $(E, \mathcal{E})$ be a measurable space and E a space. A function $f : E \rightarrow \mathbb{R}$ is called Borel measurable if it is $(\mathcal{E}, \mathcal{B}(\mathbb{R}))$ -measurable.

DEFINITION 1.17. Let $(F, \mathcal{F})$ be a measurable space, E a space and $T : E \rightarrow F$ a transformation. We define $\sigma(T)$, the σ -algebra generated by T , as the σ -algebra generated by the family of sets $\{T^{-1}(B) : B \in \mathcal{F}\}$, which is a σ -algebra on E .

$\sigma(T)$ is the smallest σ -algebra on E that makes the transformation T measurable.

THEOREM 1.18. Let $(E, \mathcal{E})$, $(F, \mathcal{F})$ and $(G, \mathcal{G})$ be measurable spaces and $T : E \rightarrow F$ and $S : F \rightarrow G$, two measurable transformations. Then $S \circ T : E \rightarrow G$ is $(\mathcal{E}, \mathcal{G})$ -measurable.

The following lemma simplifies to check if a particular function is measurable.

THEOREM 1.19. Let $(E, \mathcal{E})$ be a measurable space and $f : E \rightarrow \mathbb{R}$ a function on E . Then, f is measurable if and only if $\{x \in E : f(x) \leq a\} \in \mathcal{E}$ for all $a \in \mathbb{R}$.

THEOREM 1.20. Let f, g measurable functions on $(E, \mathcal{E})$. Then,

- (1) $af + gf, a, b \in \mathbb{R}$ is a measurable function.
- (2) fg and f/g are measurable functions.
- (3) The functions $f^+ := \max(0, f(x))$ and $f^- := -\min(0, f(x))$ are positive measurable functions. Note that $f = f^+ - f^-$ and $|f| = f^+ + f^-$.
- (4) The sets $\{x \in E : f(x) = g(x)\}$, $\{x \in E : f(x) < g(x)\}$ and $\{x \in E : f(x) \leq g(x)\}$ are measurable.

THEOREM 1.21. Let $\{f_n\}_{n \geq 1}$ be a sequence of measurable functions. Then, the following functions are measurable

- (1) $\sup_{n \geq 1} f_n(x)$.
- (2) $\inf_{n \geq 1} f_n(x)$.
- (3) $\limsup_{n \rightarrow \infty} f_n(x) := \inf_{n \geq 1} \sup_{k \geq n} f_k(x)$.
- (4) $\liminf_{n \rightarrow \infty} f_n(x) := \sup_{n \geq 1} \inf_{k \geq n} f_k(x)$.

The following result shows how to induce a measure on a measurable space from a measure space through a measurable transformation.

DEFINITION 1.22. Let $(E, \mathcal{E}, \mu)$ be a measure space, $(F, \mathcal{F})$ be a measurable space and $T : E \rightarrow F$ be a measurable transformation. Then, the set function

$$\mu_T(B) := \mu \circ T^{-1}(B) = \mu(T^{-1}B), \quad B \in \mathcal{F},$$

is a measure in $\mathcal{F}$. The measure μ_T is called the image measure induced by T on $\mathcal{F}$ or the push-forward measure.

DEFINITION 1.23. Let $(E, \mathcal{E}, \mu)$ be a measure space and suppose that some property hold at all point of $B \in \mathcal{E}$ where $\mu(B^c) = 0$. Then, this property is said to hold almost everywhere (abbreviated a.e. or μ -a.e.). The set B^c is called the exceptional set for that property. If $\mu(E) = 1$, we say that the property holds almost surely (abbreviated a.s. or μ -a.s.).

1.3. Integration Theory

Assume that we have a measure space $(E, \mathcal{E}, \mu)$.

DEFINITION 1.24. Let $f(x) = \sum_{i=1}^n a_i \mathbf{1}_{B_i}(x)$ be a nonnegative simple function defined on E , where $\{B_i\}_{i=1, \dots, n} \subset \mathcal{E}$ are disjoint and $E = \bigcup_{i=1}^n B_i$. Then, the integral of f with respect to μ is defined to be

$$\int_E f d\mu := \int_E f(x) \mu(dx) := \sum_{i=1}^n a_i \mu(B_i).$$

DEFINITION 1.25. Let f a nonnegative measurable function. The integral of f with respect to μ is defined to be

$$\int_E f d\mu = \lim_{n \rightarrow \infty} \int_E f_n d\mu,$$

where $\{f_n\}_{n \geq 1}$ is the sequence of measurable simple functions defined by

$$f_n(x) = \begin{cases} \frac{i-1}{2^n} & \text{if } \frac{i-1}{2^n} \leq f(x) < \frac{i}{2^n}, \quad i = 1, \dots, n2^n \\ n & \text{if } f(x) \geq n \end{cases}.$$

Note that $f_n(x)$ is an increasing sequence that converges pointwise to $f(x)$.

DEFINITION 1.26. Let f be a measurable function. We say that the integral of f with respect to μ exists if $\int_E f^+ d\mu < \infty$ or $\int_E f^- d\mu < \infty$ and it is defined to be

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu.$$

We say that f is integrable if $\int_E f d\mu$ exists and it is finite and we say that $f \in L^1(E, \mathcal{E}, \mu)$.

Note that f is integrable if and only if $\int_E |f| d\mu < \infty$.

THEOREM 1.27. *Let f, g be measurable functions such that $f = g$, a.e.. Then, if the integral of f exists then the integral of g also exists and they coincide. In particular, if f is integrable then g is also integrable.*

From the previous theorem one deduces that f only need to be defined a.s. to consider its integral with respect to μ . That is, it may exist a set $N \in \mathcal{E}$ with $\mu(N) = 0$ such that f is not defined on N . Define $g = f\mathbf{1}_{N^c} + 0\mathbf{1}_N$. The function g is measurable and defined for all $x \in E$. We can define, in a consistent manner, the integral of f with respect to μ by

$$\int_E f d\mu := \int_E g d\mu.$$

THEOREM 1.28. *Let f, g be integrable functions. We have that*

(1)

$$\left| \int_E f d\mu \right| \leq \int_E |f| d\mu.$$

(2) $af + bg$ is integrable and

$$\int_E (af + bg) d\mu = a \int_E f d\mu + b \int_E g d\mu,$$

(3) If $f \geq g$, a.e. then

$$\int_E f d\mu \geq \int_E g d\mu.$$

(4) If $f = g$, a.e. then

$$\int_E f d\mu = \int_E g d\mu.$$

(5) If $B \in \mathcal{E}$, then $f\mathbf{1}_B$ is measurable and we may define

$$\int_B f d\mu = \int_E f\mathbf{1}_B d\mu.$$

Moreover, if $\int_B f d\mu = 0$ for all $B \in \mathcal{E}$ then $f = 0$, a.e.

DEFINITION 1.29. For $p \geq 1$, we define the space $L^p = L^p(E, \mathcal{E}, \mu)$ to be the set of equivalence classes of measurable functions $f : E \rightarrow \mathbb{R}$ such that $\int_E |f|^p d\mu < \infty$, where two measurable functions are equivalent if they are equal μ -a.e.

One can prove that L^p is a vector space. Moreover, in L^p we can define a norm $\|\cdot\|_{L^p}$ by

$$\|f\|_{L^p} := \left(\int_E |f|^p d\mu \right)^{1/p}.$$

THEOREM 1.30. *For $p \geq 1$ the normed vector space $(L^p, \|\cdot\|)$ is complete, i.e., every Cauchy sequence is convergent to an element of L^p . That means, if $\lim_{n,m \rightarrow \infty} \|f_n - f_m\|_{L^p} = 0$ then there exist $f \in L^p$ such that $\lim_{n \rightarrow \infty} \|f_n - f\|_{L^p} = 0$.*

THEOREM 1.31. $L^2(E, \mathcal{E}, \mu)$ is also a Hilbert space with the inner product $(f, g) := \int_E fg d\mu$.

THEOREM 1.32. *The following inequalities hold.*

(1) (**Minkowski**): If $f, g \in L^p, p \geq 1$, then $f + g \in L^p$ and

$$\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}.$$

- (2) (**Hölder**): Let $p, q \geq 1$ such that $pq = p + q$. If $f \in L^p$ and $g \in L^q$, then $fg \in L^1$ and

$$\|fg\|_{L^1} \leq \|f\|_{L^p} \|g\|_{L^q}.$$

If $p = q = 2$, this result is known as Cauchy's inequality.

- (3) (**Chebyshev-Markov**): Let f be a measurable function and let g be a measurable function that is nonnegative and nondecreasing on the range of f . Then, for all $t > 0$ we have

$$\mu(\{x \in E : f(x) \geq t\}) \leq \frac{1}{g(t)} \int_E g(f(x)) \mu(dx).$$

The following is a general change of variable formula.

THEOREM 1.33 (Image Measure). Let $(E, \mathcal{E}, \mu)$ be a measure space, $(F, \mathcal{F})$ a measurable space, $T : E \rightarrow F$ a measurable transformation and $f : F \rightarrow \mathbb{R}$ a measurable function. Then, the integral of f as a function on $(F, \mathcal{F}, \mu_T)$ exists if and only if the integral of $f \circ T$ as a function on $(E, \mathcal{E}, \mu)$ exists, and if this is the case

$$\int_E f \circ T d\mu = \int_F f d\mu_T.$$

In particular, f is integrable if and only if $f \circ T$ is integrable in their respective spaces.

The previous theorem is a powerful tool to compute integrals of functions defined on an arbitrary measure space $(E, \mathcal{E}, \mu)$. The idea is to find an equivalent integral on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. Assume that we have a measurable function g defined on a measure space $(E, \mathcal{E}, \mu)$ and we want to compute its integral with respect to μ . Applying the image measure theorem with $T = g$ and $f = Id$ (the identity function) we get that

$$\int_E g d\mu = \int_E Id \circ g d\mu = \int_{\mathbb{R}} Id d\mu_g = \int_{\mathbb{R}} x \mu_g(dx).$$

1.4. Modes of Convergence

The following are the most basic notions of convergence for sequences of measurable functions.

DEFINITION 1.34. Let $\{f_n\}_{n \geq 1}$ be a sequence of measurable functions and let f be a measurable function. We say that

- (1) $\{f_n\}_{n \geq 1}$ converges to f almost everywhere (a.e.) if there exists a measurable set N with $\mu(N) = 0$ such that

$$\lim_{n \rightarrow \infty} f_n(x) = f(x), \quad x \notin N.$$

- (2) $\{f_n\}_{n \geq 1}$ converges to f in L^p if

$$\lim_{n \rightarrow \infty} \int_E |f_n - f|^p d\mu = 0.$$

- (3) $\{f_n\}_{n \geq 1}$ converges to f in measure if

$$\lim_{n \rightarrow \infty} \mu(\{x \in E : |f_n(x) - f(x)| > \varepsilon\}) = 0, \quad \forall \varepsilon > 0.$$

The following results relate the operation of taking limits and integration.

REMARK 1.35. We have the following relationships between modes of convergence:

- (1) Convergence in L^p implies convergence in measure.
- (2) Convergence in measure (or in L^p) implies the existence of a subsequence converging a.e..
- (3) If $\mu(E) < \infty$, convergence a.e. implies convergence in measure.
- (4) If $\{f_n\}_{n \geq 1}$ is such that $|f_n(x)| \leq g(x)$, where $g \in L^p$, then convergence a.e. (or in measure) implies convergence in L^p .

THEOREM 1.36 (**Monotone convergence**). *Let $\{f_n\}_{n \geq 1}$ be sequence of non-negative measurable functions that is a.e. increasing. Then, the integrals of $f_n, n \geq 1$ and $\lim_{n \rightarrow \infty} f_n(x)$ exist and*

$$\int_E \lim_{n \rightarrow \infty} f_n(x) d\mu = \lim_{n \rightarrow \infty} \int_E f_n d\mu.$$

In particular, if $\lim_{n \rightarrow \infty} \int_E f_n d\mu < +\infty$ then $f_n, n \geq 1$ and $\lim_{n \rightarrow \infty} f_n(x)$ are integrable.

THEOREM 1.37 (**Fatou's lemma**). *Let $\{f_n\}_{n \geq 1}$ be a sequence of non-negative measurable functions. Then, the integrals of $f_n, n \geq 1$ and $\liminf_{n \rightarrow \infty} f_n(x)$ exist and*

$$\int_E \liminf_{n \rightarrow \infty} f_n(x) d\mu \leq \liminf_{n \rightarrow \infty} \int_E f_n d\mu.$$

In particular, if $\liminf_{n \rightarrow \infty} \int_E f_n d\mu < +\infty$ then $\liminf_{n \rightarrow \infty} f(x)$ is integrable.

THEOREM 1.38 (**Dominated convergence**). *Let $\{f_n\}_{n \geq 1}$ be a sequence of measurable functions such that $|f_n(x)| \leq g(x)$, a.e., where $g \in L^1, p \geq 1$. Then, $f := \lim_{n \rightarrow \infty} f_n(x)$ (where the convergence is either a.e. or in measure) is integrable and*

$$\int_E f d\mu = \lim_{n \rightarrow \infty} \int_E f_n d\mu.$$

1.5. The Radon-Nikodym Theorem

DEFINITION 1.39. Let $(E, \mathcal{E}, \mu)$ be a measure space. Let $f : E \rightarrow \mathbb{R}_+$ be a measurable function. We define on $\mathcal{E}$ the measure

$$\nu(B) = \int_B f d\mu := \int_E \mathbf{1}_B f d\mu, \quad B \in \mathcal{E}.$$

We say that f is the density of ν with respect to μ and it is written $f = \frac{d\nu}{d\mu}$. We also say that f is the Radon-Nikodym derivative of ν with respect to μ .

DEFINITION 1.40. Let μ and ν be two measures on a measurable space $(E, \mathcal{E})$. We say that ν is absolutely continuous with respect to μ if and only if

$$\forall A \in \mathcal{E}, \text{ such that } \mu(A) = 0 \Rightarrow \nu(A) = 0,$$

and we write $\nu \ll \mu$. If $\mu \ll \nu$ and $\nu \ll \mu$ we say that the two measures are equivalent and we write $\mu \sim \nu$.

THEOREM 1.41 (**Radon-Nikodym**). *Let μ and ν be two measures on a measurable space $(E, \mathcal{E})$ and μ being σ -finite. Then, the following statements are equivalent:*

(1) $\nu \ll \mu$.

(2) There exists a measurable function $f : E \rightarrow \mathbb{R}_+$ such that $f = \frac{d\nu}{d\mu}$.

PROPOSITION 1.42. Let $(E, \mathcal{E}, \mu)$ be a measure space. Let $\nu \ll \mu$ with density $f = \frac{d\nu}{d\mu}$. Let $g : E \rightarrow \mathbb{R}$ measurable. Then the integral of g with respect to ν exists if and only if the integral of gf exists and, if this is the case,

$$\int_E g d\nu = \int_E gf d\mu.$$

In particular, f is integrable with respect to ν if and only if gf is integrable with respect to μ .

1.6. Measures on Product Spaces

DEFINITION 1.43. Let $\{(E_i, \mathcal{E}_i)\}_{i \in I}$ a family of measurable spaces. Let $E := \prod_{i \in I} E_i$ the Cartesian product of the family $\{E_i\}_{i \in I}$. Let $\pi_i : E \rightarrow E_i, i \in I$, the projection of E into its factors. We call the product σ -algebra of $\{\mathcal{E}_i\}_{i \in I}$ the σ -algebra generated by $\{\pi_i\}_{i \in I}$ and we will denote it by $\mathcal{E} = \bigotimes_{i \in I} \mathcal{E}_i$. That is,

$$\bigotimes_{i \in I} \mathcal{E}_i = \sigma(\{\pi_i^{-1}(A_i) : A_i \in \mathcal{E}_i, i \in I\}).$$

A set of the form $\pi_i^{-1}(A_i)$ is called a cylinder of base A_i . $(E, \mathcal{E})$ is called the measurable product space of $\{(E_i, \mathcal{E}_i)\}_{i \in I}$.

Observe that if $(\Omega, \mathcal{F})$ is another measurable space, then $g : \Omega \rightarrow E$ is measurable iff $\forall i \in I, \pi_i \circ g : \Omega \rightarrow E_i$ is measurable.

When we have finite product of measurable spaces the product σ -algebra can be defined, equivalently, in terms of rectangles.

DEFINITION 1.44. Let $(E, \mathcal{E})$ the measurable product space of $\{(E_i, \mathcal{E}_i)\}_{i=1, \dots, n}$. A measurable rectangle is a subset of E of the form $A_1 \times \dots \times A_n$ with $A_i \in \mathcal{E}_i$.

We have that the σ -algebra generated by the measurable rectangles coincides with $\mathcal{E}$.

DEFINITION 1.45. Let $\{(E_i, \mathcal{E}_i, \mu_i)\}_{i=1, \dots, n}$ be a family of measure spaces. The product measure on $(\prod_{i=1, \dots, n} E_i, \bigotimes_{i=1, \dots, n} \mathcal{E}_i)$, denoted by $\bigotimes_{i=1, \dots, n} \mu_i$, is the measure satisfying

$$\mu_1 \otimes \dots \otimes \mu_n (A_1 \times \dots \times A_n) = \mu_1 (A_1) \cdot \dots \cdot \mu_n (A_n),$$

where $A_i \in \mathcal{E}_i, i = 1, \dots, n$.

THEOREM 1.46 (Tonelli-Fubini). Let $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ be two σ -finite measure spaces and $f : E_1 \times E_2 \rightarrow \mathbb{R}$ be a $\mathcal{E}_1 \otimes \mathcal{E}_2$ -measurable function. If either

$$\int_{E_1} \left(\int_{E_2} |f(x, y)| \mu_2(dy) \right) \mu_1(dx) < \infty,$$

or

$$\int_{E_2} \left(\int_{E_1} |f(x, y)| \mu_1(dx) \right) \mu_2(dy) < \infty,$$

then, $f \in L^1(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2, \mu_1 \otimes \mu_2)$ and

$$\begin{aligned} \int_{E_1 \times E_2} f(x, y) \mu_1 \otimes \mu_2(dx, dy) &= \int_{E_1} \left(\int_{E_2} f(x, y) \mu_2(dy) \right) \mu_1(dx) \\ &= \int_{E_2} \left(\int_{E_1} f(x, y) \mu_1(dx) \right) \mu_2(dy). \end{aligned}$$

1.7. The Monotone Class Theorem and Dynkin's π - λ Theorem

In this section we present some fundamental technical results known as monotone class theorems. These allow to prove that certain families of sets are σ -algebras.

DEFINITION 1.47. A family of subsets $\mathcal{M}$ of another set E is said to be a monotone class of E if $A \in \mathcal{M}$ whenever $\{A_n\}_{n \geq 1}$ is a sequence of sets in $\mathcal{M}$ with either

- (1) $A_n \subseteq A_{n+1}$ and $\bigcup_{n \geq 1} A_n = A$, or
- (2) $A_{n+1} \subseteq A_n$ and $\bigcap_{n \geq 1} A_n = A$.

Given $\mathcal{K}$ a family of sets, there exists a smallest monotone class containing $\mathcal{K}$.

THEOREM 1.48 (Monotone Class Theorem). *Let E be a set, and $\mathcal{A}$ an algebra of subsets of E . Suppose $\mathcal{M}$ is a monotone class of subsets of E which contains $\mathcal{A}$. Then $\mathcal{M}$ contains the σ -algebra $\sigma(\mathcal{A})$. Furthermore, $\sigma(\mathcal{A})$ is the smallest monotone class containing $\mathcal{A}$.*

The usual way of applying the previous result is to consider a simple algebra $\mathcal{A}$, and to define $\mathcal{M}$ to be the collection of sets in $\sigma(\mathcal{A})$ where some property holds. Then, if we show that

- the algebra $\mathcal{A}$ lies in $\mathcal{M}$,
- limits of monotone sequences in $\mathcal{M}$ lie in $\mathcal{M}$,

then Theorem 1.48 allows us to conclude that $\mathcal{M}$ contains all of $\sigma(\mathcal{A})$.

Next we give a monotone class theorem for functions.

DEFINITION 1.49. A monotone vector space $\mathcal{H}$ on a space E is defined to be the collection of bounded, real-valued functions f on E satisfying the three conditions:

- (1) $\mathcal{H}$ is a vector space $\mathbb{R}$,
- (2) $\mathbf{1}_E \in \mathcal{H}$ (that is, constant functions are in $\mathcal{H}$), and
- (3) if $\{f_n\}_{n \geq 1} \subset \mathcal{H}$, and $0 \leq f_1 \leq f_2 \leq \dots \leq f_n \leq \dots$, and $\lim_{n \rightarrow \infty} f_n = f$, and f is bounded, then $f \in \mathcal{H}$.

DEFINITION 1.50. A collection $\mathcal{M}$ of real functions defined on a space E is said to be multiplicative if $f, g \in \mathcal{M}$ implies that $fg \in \mathcal{M}$.

For a collection of real-valued functions $\mathcal{M}$ defined on E , we let $\sigma(\mathcal{M})$ denote the space of functions defined on E which are measurable with respect to the σ -algebra on E generated by $\{f^{-1}(A) : A \in \mathcal{B}(\mathbb{R}), f \in \mathcal{M}\}$.

THEOREM 1.51 (Monotone Class Theorem for Functions). *Let $\mathcal{M}$ be a multiplicative class of bounded, real-valued functions defined on a space E . If $\mathcal{H}$ is a monotone vector space containing $\mathcal{M}$, then $\mathcal{H}$ contains all bounded, $\sigma(\mathcal{M})$ -measurable functions.*

The following results are sometimes also referred to as the monotone class theorem.

DEFINITION 1.52. A collection $\mathcal{A}$ of sets is called a λ -system (or d -system) on E if

- (1) $E \in \mathcal{A}$,
- (2) For any $A, B \in \mathcal{A}$ with $A \subseteq B$, $B \setminus A \in \mathcal{A}$,
- (3) If $\{A_n\}_{n \geq 1} \subseteq \mathcal{A}$ and $A_n \subseteq A_{n+1}$ for all n , then $\bigcup_{n \geq 1} A_n \in \mathcal{A}$.

A collection $\mathcal{A}$ of sets is called a π -system if $A \cap B \in \mathcal{A}$ whenever $A, B \in \mathcal{A}$.

LEMMA 1.53. A collection $\mathcal{K}$ of subsets of E is a σ -algebra if and only if it is both a π -system and a λ -system.

THEOREM 1.54 (Dynkin's π - λ -Systems Lemma). If $\mathcal{A}$ is a π -system, then any λ -system containing $\mathcal{A}$ contains $\sigma(\mathcal{A})$.

1.8. Exercises

- (1) Is the family $\mathcal{E}$ consisting of all finite subsets of E and their complements always a σ -algebra?
- (2) Let E and $\tilde{E}$ be arbitrary sets and $X : \tilde{E} \rightarrow E$ be any mapping. Show that if $\mathcal{E}$ is a σ -algebra on E then $\tilde{\mathcal{E}} = \{X^{-1}(A) : A \in \mathcal{E}\}$ is a σ -algebra on $\tilde{E}$.
- (3) Let $\mathcal{E}$ be a σ -algebra on E and let $A \subset E$. Show that $\mathcal{E}_A := \{A \cap B : B \in \mathcal{E}\}$ is a σ -algebra on A .
- (4) Show that if $\{\mathcal{E}_i\}_{i \in I}$ is any collection of σ -algebra defined on the same set E , then their intersection $\bigcap_{i \in I} \mathcal{E}_i$ is also a σ -algebra on E .
- (5) Let $(E, \mathcal{E})$ be a measurable space. We define the *upper limit* or *limit superior* of a sequence of sets $\{A_n\}_{n \geq 1} \subset E$ by

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k,$$

and the *lower limit* or *limit inferior* by

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

Show that if $\{A_n\}_{n \geq 1} \subset \mathcal{E}$, then

$$\limsup_{n \rightarrow \infty} A_n \in \mathcal{E}, \quad \liminf_{n \rightarrow \infty} A_n \in \mathcal{E}.$$

Show that

$$\begin{aligned} \limsup_{n \rightarrow \infty} A_n &= \{A_n \text{ occurs for infinitely many } n\}, \\ \liminf_{n \rightarrow \infty} A_n &= \{A_n \text{ occurs for all but finitely many } n\}. \end{aligned}$$

and that $\liminf_{n \rightarrow \infty} A_n \subset \limsup_{n \rightarrow \infty} A_n$.

- (6) Let $\Omega = \mathbb{N}$, and let $\mathcal{F} = \mathcal{P}(\Omega)$. Set $P(\{i\}) = \alpha_i, i = 1, 2, \dots$. Extend P to a probability measure defined on $\mathcal{F}$. What conditions have to be imposed on the numbers α_i ? Can they all be chosen the same?

- (7) Let $(E, \mathcal{E}, \mu)$ be a measure space. Show that μ is finitely subadditive, that is, for any $\{A_n\}_{n=1, \dots, N} \subset \mathcal{E}$ where $N \in \mathbb{N}$, one has that

$$\mu\left(\bigcup_{n=1}^N A_n\right) \leq \sum_{n=1}^N \mu(A_n).$$

- (8) Let $f : E \rightarrow \mathbb{R}$ be a measurable function such that $f \geq 0, \mu$ -a.e.. Prove that if $\int_E f d\mu = 0$ then $f = 0, \mu$ -a.e..
 (9) Let $f \in L^1(E, \mathcal{E}, \mu)$. Prove that if $\int_A f d\mu = 0, \forall A \in \mathcal{E}$ then $f = 0, \mu$ -a.e..
 (10) Consider the measure space $([0, 1], \mathcal{B}(\mathbb{R}), \lambda)$. Let $f(x) = 0, 0 \leq x \leq 1$ and

$$f_n(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1/n \\ 0 & \text{if } 1/n < x \leq 1 \end{cases}.$$

Does $\{f_n\}_{n \geq 1}$ converge to f

- (a) for all x ?
 - (b) a.e.?
 - (c) in measure?
 - (d) in L^p ?
- (11) Let P and Q be two probability measures defined on the same measurable space $(\Omega, \mathcal{F})$. Show that $P \sim Q$ iff $Q \ll P$ and $P(\frac{dQ}{dP} = 0) = 0$. What is the relationship between $\frac{dQ}{dP}$ and $\frac{dP}{dQ}$?
- (12) Let $(E, \mathcal{E}, \mu)$ be a measure space. Let $I \subset \mathbb{R}$ be an open interval and $t_0 \in I$. Let $\{f_t\}_{t \in I}$ be a family of measurable functions from E to $\mathbb{R}$ satisfying:
- (a) $f_t(x) : I \rightarrow \mathbb{R}$ is differentiable on I, μ -a.e.
 - (b) There exists a measurable function g such that $|f_t| + \left|\frac{d}{dt}f_t\right| \leq g, \mu$ -a.e., $\forall t \in I$.

Show that the function

$$\begin{aligned} F : I &\rightarrow \mathbb{R} \\ t &\mapsto \int_E f_t d\mu \end{aligned},$$

is well defined, differentiable at t_0 with $F'(t_0) = \int_E \frac{d}{dt}f_t \Big|_{t=t_0} d\mu$.

CHAPTER 2

Review of Probability Theory

2.1. Basic definitions

DEFINITION 2.1. A probability space $(\Omega, \mathcal{F}, P)$ is a measure space such that $P(\Omega) = 1$.

DEFINITION 2.2. A random variable with values in $\mathbb{R}^d$ (or d -dimensional random vector) $X : \Omega \rightarrow \mathbb{R}^d$, is a $(\mathcal{F}, \mathcal{B}(\mathbb{R}^d))$ -measurable mapping.

DEFINITION 2.3. The (joint) law of a random vector $X = (X_1, \dots, X_d)^T$, denoted by $\mathcal{L}(X)$, is the image measure (or push-forward measure) P_X on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, that is,

$$P_X(B) = P \circ X^{-1}(B) = P(X^{-1}B), \quad B \in \mathcal{B}(\mathbb{R}^d).$$

In this context, the law of X_i is called the marginal law of X_i . Note that P_X is a Lebesgue-Stieltjes measure on $\mathcal{B}(\mathbb{R}^d)$.

DEFINITION 2.4. The (joint) distribution function $F_X(x)$ of a random vector X is the following distribution function associated to the Lebesgue-Stieltjes measure P_X :

$$\begin{aligned} F_X(x_1, \dots, x_d) &= P_X((-\infty, x_1] \times \dots \times (-\infty, x_d]) \\ &= P(\{X_1 \leq x_1, \dots, X_d \leq x_d\}), \quad (x_1, \dots, x_d) \in \mathbb{R}^d. \end{aligned}$$

We note that the joint law of X determines the marginal laws of X_i , but not viceversa. Moreover,

$$F_{X_i}(x_i) = \lim_{x_j \rightarrow \infty} F_X(x_1, \dots, x_d).$$

DEFINITION 2.5. We say that a random vector X has an absolutely continuous law (with respect to the Lebesgue measure) if $P_X(B) = 0$ for any $B \in \mathcal{B}(\mathbb{R}^d)$ such that $\lambda_d(B) = 0$. We write $P_X \ll \lambda_d$.

The following two definitions make sense thanks to the image measure theorem and the Radon-Nikodym theorem.

DEFINITION 2.6. Let $P_X \ll \lambda_d$. Then, we define the density f_X of X to be the Radon-Nikodym derivative $f_X = \frac{dP_X}{d\lambda_d}$ and

$$P_X(B) = \int_B dP_X = \int_B f_X d\lambda_d, \quad B \in \mathcal{B}(\mathbb{R}^d).$$

We have that $f_X(x_1, \dots, x_d) = \frac{\partial^d}{\partial x_1 \dots \partial x_d} F_X(x_1, \dots, x_d)$.

DEFINITION 2.7. Let $g : \mathbb{R}^d \rightarrow \mathbb{R}$ be a Borel measurable function. Then the expectation of $g(X)$ is defined to be

$$\mathbb{E}[g(X)] = \int_{\Omega} g \circ X dP = \int_{\mathbb{R}^d} g dP_X.$$

If $P_X \ll \lambda_d$, then

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^d} g f_X d\lambda_d.$$

DEFINITION 2.8. Let $X = (X_1, \dots, X_d)^T$. If $\mathbb{E}[|X_i|] < \infty, i = 1, \dots, d$ then $\mu = (\mu_1, \dots, \mu_d)^T = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])^T$, the mean vector of X , is well defined. If $X_i \in L^2, i = 1, \dots, d$ then

$$\begin{aligned} \Sigma = \text{Cov}(X) &= (\text{Cov}(X_i, X_j))_{i,j=1,\dots,d} \\ &= (\mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)])_{i,j=1,\dots,d} = \mathbb{E}[(X - \mu)(X - \mu)^T], \end{aligned}$$

the covariance matrix of X , is well defined. We will also use the notation $\text{Var}[X_i] := \text{Cov}(X_i, X_i), i = 1, \dots, d$.

REMARK 2.9. The covariance matrix of a random vector X is a symmetric nonnegative definite matrix, i.e. $\Sigma_{ij} = \Sigma_{ji}, i, j = 1, \dots, d$ and $y^T \Sigma y \geq 0$ for all $y \neq 0, y \in \mathbb{R}^d$. If $\det \Sigma > 0$ then the probability distribution of X is a truly d -dimensional distribution in the sense that cannot be concentrated on a lower dimensional subspace of $\mathbb{R}^d$ and we say that the distribution is non-degenerate or regular. If $\det \Sigma = 0$ then the probability distribution of X is concentrated on a lower dimensional subspace of $\mathbb{R}^d$ and we say that the distribution is degenerate or singular.

PROPOSITION 2.10. Let X be a d -dimensional random vector with mean vector μ and covariance matrix Σ . Further, let B be an $m \times d$ matrix, let b be a constant m -dimensional vector, and set $Y = BX + b$. Then, Y is a m -dimensional random vector with mean vector and covariance matrix given by

$$\mathbb{E}[Y] = B\mu + b, \quad \text{Cov}(Y) = B\Sigma B^T.$$

THEOREM 2.11. Let $S \in \mathcal{B}(\mathbb{R}^d)$ be partitioned into disjoint measurable subsets $S_0, S_1, \dots, S_m$ such that $S = \cup_{i=0}^m S_i$ and such that $\lambda_d(S_0) = 0$. Assume that for each $i = 1, \dots, m$, the mapping $g : S_i \rightarrow \mathbb{R}^d$ is injective (one to one) and continuously differentiable with non-vanishing Jacobian. Let $Y = g(X)$, where $X : \Omega \rightarrow S$ is a random vector with density f_X with respect to λ_d . Then, Y has density (with respect to λ_d) given by

$$f_Y(y) = \sum_{i=1}^m f_X(g_i^{-1}(y)) \left| \det J_{g_i^{-1}}(y) \right| \mathbf{1}_{g(S_i)}(y),$$

where g_i^{-1} denotes the inverse map $g_i^{-1} : g(S_i) \rightarrow S_i$ and $J_{g_i^{-1}}$ its corresponding Jacobian matrix.

We have different notions of convergence for sequences of random variables that are analogous to the ones for sequences of measurable functions. In particular, convergence almost everywhere is denoted convergence almost surely (a.s.), convergence in L^p is also denoted convergence in L^p and convergence in measure is denoted convergence in probability.

DEFINITION 2.12. Let $\{\mu_n\}_{n \geq 1}$ and μ probability measures defined on $(S, \mathcal{B}(S))$, where S is a complete, separable metric space. We say that $\{\mu_n\}_{n \geq 1}$ converges weakly to μ if

$$\int_{\Omega} f d\mu_n \rightarrow \int_{\Omega} f d\mu,$$

for all $f : \Omega \rightarrow \mathbb{R}$ continuous and bounded.

We say that sequence of random vectors $\{X_n\}_{n \geq 1}$ converges weakly (or in distribution, or in law) to the random vector X if P_{X_n} converges weakly to P_X , as probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

This notion of convergence for random variables is the weakest of all, that is, convergence in probability implies weak convergence. If the limit is a degenerate random variable (a constant), then convergence in distribution implies convergence in probability. It is important to note that the weak convergence of random variables only involves their laws. In particular, the random variables can be defined on different probability spaces.

Besides Minkowski, Hölder, Cauchy-Schwarz, Chebyshev-Markov inequalities from general measure theory we have the following important inequality for random variables.

THEOREM 2.13 (**Jensen's inequality**). *Let X be an integrable random variable and φ a convex function then*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

2.2. Independence

DEFINITION 2.14. Let $(\Omega, \mathcal{F}, P)$ be a probability space. A family of events $\{A_i\}_{i \in I} \subseteq \mathcal{F}$ are independent iff $\forall J \subset I, J$ finite, one has that

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

DEFINITION 2.15. Let $\{\mathcal{E}_i\}_{i \in I}, \mathcal{E}_i \subset \mathcal{F}$, be a family of collections of events. We say that $\{\mathcal{E}_i\}_{i \in I}$ are independent iff every family $\{A_i\}_{i \in I}$, with $A_i \in \mathcal{E}_i$, is independent.

DEFINITION 2.16. Let $\{X_i : \Omega \rightarrow \mathbb{R}\}_{i \in I}$ be a family of random variables. We say that $\{X_i\}_{i \in I}$ is independent iff the family of σ -algebras $\{\sigma(X_i)\}_{i \in I}$ is independent.

Taking into account the previous definitions we have that $\{X_i\}_{i \in I}$ is independent iff for all finite subset $\{i_1, \dots, i_n\} \subset I$, and for all $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$, one has that

$$\begin{aligned} P(\{X_{i_1} \in B_1, \dots, X_{i_n} \in B_n\}) &:= P(\{X_{i_1} \in B_1\} \cap \dots \cap \{X_{i_n} \in B_n\}) \\ &= P(X_{i_1} \in B_1) \cdots P(X_{i_n} \in B_n). \end{aligned}$$

The next proposition characterizes the independence of a finite number of r.v.

PROPOSITION 2.17. *Let $\{X_i\}_{i=1, \dots, d}$ be random variables with laws $P_{X_i}, i = 1, \dots, d$, respectively, and let $X = (X_1, \dots, X_d)$ be the random vector with law P_X . Then, the following statements are equivalent*

- (1) $\{X_i\}_{i=1, \dots, d}$ are independent;
- (2) $P_X = P_{X_1} \otimes \dots \otimes P_{X_d}$;

$$(3) F_X(x_1, \dots, x_d) = F_{X_1}(x_1) \cdots F_{X_d}(x_d);$$

Moreover, if we assume that $P_X \ll \lambda^d$ then $\{X_i\}_{i=1, \dots, d}$ are independent iff

$$f_X(x_1, \dots, x_d) = f_{X_1}(x_1) \cdots f_{X_d}(x_d).$$

The next proposition only gives necessary conditions for the independence.

PROPOSITION 2.18.

- (1) Let $(E_i, \mathcal{E}_i)_{i \in I}$ and $(G_i, \mathcal{G}_i)_{i \in I}$ be two families of measurable spaces. Let $\{X_i\}_{i \in I}$ be a family of random elements $X_i : \Omega \rightarrow E_i$, and $\{g_i\}_{i \in I}$ a family of measurable transformations $g_i : E_i \rightarrow G_i$. Then, $\{g_i \circ X_i\}_{i \in I}$ is a family of independent random elements $g_i \circ X_i : \Omega \rightarrow G_i$.
- (2) Let $\{X_i\}_{i=1, \dots, d}$ be independent random variables with finite expectation, i.e., $\mathbb{E}[|X_i|] < \infty, i = 1, \dots, d$, then $X_1 \cdots X_d$ has finite expectation and

$$\mathbb{E}[X_1 \cdots X_d] = \mathbb{E}[X_1] \cdots \mathbb{E}[X_d].$$

- (3) Let $\{X_i\}_{i=1, \dots, d}$ be independent random variables with $\mathbb{E}[|X_i|^2] < \infty, i = 1, \dots, d$, then

$$\text{Var}[X_1 + \cdots + X_d] = \text{Var}[X_1] + \cdots + \text{Var}[X_d].$$

2.3. Characteristic Functions

DEFINITION 2.19. Let μ be a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Its characteristic function $\varphi_\mu : \mathbb{R}^d \rightarrow \mathbb{C}$ is defined to be

$$\varphi_\mu(u) := \int_{\mathbb{R}^d} e^{i\langle u, x \rangle} \mu(dx), \quad u \in \mathbb{R}^d,$$

where $\langle u, x \rangle := u^T x = \sum_{i=1}^d u_i x_i$. If X is a random vector we define its characteristic function φ_X as the characteristic function of its law P_X , i.e.,

$$\varphi_X(u) = \int_{\mathbb{R}^d} e^{i\langle u, x \rangle} P_X(dx) = \mathbb{E}[e^{i\langle u, X \rangle}], \quad u \in \mathbb{R}^d.$$

THEOREM 2.20. X and Y have the same characteristic function if and only if they have the same law, i.e., $P_X = P_Y$.

LEMMA 2.21. The following properties of φ_X are elementary:

- (1) $|\varphi_X(u)| \leq 1$;
- (2) $\varphi_X(0) = 1$;
- (3) φ_X is uniformly continuous;
- (4) φ_X is Hermitian, i.e., $\varphi_X(-u) = \overline{\varphi_X(u)}$, where the bar denotes complex conjugation;
- (5) X is symmetric (with respect to the origin) if and only if φ_X is real-valued.
- (6) If $X = (X_1, \dots, X_d)^T$ and $\mathbb{E}[|X_j|^n] < \infty$ for some $1 \leq j \leq d$ and $n \in \mathbb{N}$, then

$$\mathbb{E}[X_j^n] = i^{-n} \frac{\partial^n}{\partial u_j^n} \varphi_X(u) \Big|_{u=0}.$$

If $M_X(u) := \varphi_X(-iu)$ exists in a neighbourhood of $u = 0$, then M_X is called the moment generating function of X . In this case, all the moments of X exist and can be obtained by partial differentiation of M_X .

PROPOSITION 2.22. *Let X be a d -dimensional random vector with mean vector μ and covariance matrix Σ . Further, let B be an $m \times d$ matrix, let b be a constant m -dimensional vector, and set $Y = BX + b$. Then,*

$$\varphi_Y(u) = e^{i\langle u, b \rangle} \varphi_X(B^T u), \quad u \in \mathbb{R}^d.$$

PROPOSITION 2.23. *Let $X = (X_1, \dots, X_d)$ be a random vector. Then $\{X_i\}_{i=1, \dots, d}$ are independent if and only if*

$$\varphi_X(u_1, \dots, u_d) = \varphi_{X_1}(u_1) \cdots \varphi_{X_d}(u_d).$$

DEFINITION 2.24. Let $\psi : \mathbb{R}^d \rightarrow \mathbb{C}$. We say that

- (1) ψ is positive definite if for all $n \in \mathbb{N}, c_1, \dots, c_n \in \mathbb{C}$, and $u_1, \dots, u_n \in \mathbb{R}^d$ we have

$$\sum_{i,j=1}^n c_j \overline{c_i} \psi(u_j - u_i) \geq 0.$$

- (2) ψ is conditionally positive definite if for all $n \in \mathbb{N}, c_1, \dots, c_n \in \mathbb{C}$ such that $\sum_{i=1}^n c_i = 0$, and $u_1, \dots, u_n \in \mathbb{R}^d$ we have

$$\sum_{i,j=1}^n c_j \overline{c_i} \psi(u_j - u_i) \geq 0.$$

THEOREM 2.25 (Bochner's theorem). *A mapping $\varphi : \mathbb{R}^d \rightarrow \mathbb{C}$ is the characteristic function of a probability measure if and only if $\varphi(0) = 1$, φ is positive definite, and φ is continuous at zero.*

THEOREM 2.26 (Schoenberg Correspondence). *A mapping $\psi : \mathbb{R}^d \rightarrow \mathbb{C}$ is Hermitian and conditionally positive definite if and only if $e^{t\psi}$ is positive definite for each $t > 0$.*

THEOREM 2.27 (Glivenko). *Let $\{\mu_n\}_{n \geq 1}$ be a sequence of probability measures and μ be a probability measure, all defined on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then, $\varphi_{\mu_n}(u) \rightarrow \varphi_\mu(u), u \in \mathbb{R}^d$ if and only if μ_n converges weakly to μ . The convergence of the characteristic functions is uniformly on compacts.*

THEOREM 2.28 (Lévy's continuity theorem). *If $(\varphi_n)_{n \geq 1}$ is a sequence of characteristic functions and there exists a function $\psi : \mathbb{R}^d \rightarrow \mathbb{C}$ such that for all $u \in \mathbb{R}^d, \varphi_n(u) \rightarrow \psi(u)$ and ψ is continuous at 0 then ψ is the characteristic function of a probability measure.*

2.4. The Multivariate Normal Distribution

DEFINITION 2.29. A d -dimensional random vector X is Gaussian (or multivariate normal) iff every linear combination $a^T X = \sum_{i=1}^d a_i X_i$ has a (one dimensional) Gaussian distribution (possibly degenerate). The notation $X \sim \mathcal{N}(\mu, \Sigma)$ is used to denote that X has a multivariate normal distribution with mean vector μ and covariance matrix Σ .

We say that X is a standard multivariate normal random vector if $X \sim \mathcal{N}(0, I_d)$ where I_d is the d -dimensional identity matrix.

THEOREM 2.30. *Suppose that $X \sim \mathcal{N}(\mu, \Sigma)$ and set $Y = BX + b$. Then $Y \sim \mathcal{N}(B\mu + b, B\Sigma B^T)$.*

The previous theorem is useful to obtain a random vector with $X \sim \mathcal{N}(\mu, \Sigma)$ from an affine transform of a random vector $Z \sim \mathcal{N}(0, I_d)$. One has that $X = \Sigma^{1/2}Z + \mu$, where $\Sigma^{1/2}$ is a square root of the covariance matrix Σ , that is, $\Sigma^{1/2}$ is a matrix B such that $BB^T = \Sigma$.

THEOREM 2.31. *Let $X \sim \mathcal{N}(\mu, \Sigma)$ and $\det \Sigma > 0$. Then $P_X \ll \lambda_d$ and*

$$f_X(x) = \frac{dP_X}{d\lambda_d}(x) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right), \quad x \in \mathbb{R}^d.$$

THEOREM 2.32. *Let $X \sim \mathcal{N}(\mu, \Sigma)$. Then $\{X_i\}_{i=1, \dots, d}$ are independent iff $\Sigma_{i,j} = \text{Cov}(X_i, X_j) = 0$. If, in addition, $\det \Sigma > 0$ then*

$$f_X(x) = \prod_{i=1}^d \frac{1}{\sqrt{2\pi \Sigma_{i,i}}} \exp\left(-\frac{(x - \mu_i)^2}{\Sigma_{i,i}}\right).$$

2.5. Borel-Cantelli Lemma

Here we give one of the most important/useful results on sequences of measurable sets.

THEOREM 2.33. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{A_n\}_{n \geq 1}$ be a sequence of sets in $\mathcal{F}$. Then*

(1) *If $\sum_{n \geq 1} P(A_n) < \infty$, then it follows that*

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = 0.$$

(2) *If $\sum_{n \geq 1} P(A_n) = \infty$ and the events $\{A_n\}_{n \geq 1}$ are independent, then*

$$P\left(\left(\limsup_{n \rightarrow \infty} A_n\right)^c\right) = P\left(\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right)^c\right) = 0.$$

2.6. Conditional Expectation

The concept of conditional expectation is crucial in many areas of modern Probability Theory. Important types of stochastic processes, like martingales and Markov processes, are defined using conditional expectation.

DEFINITION 2.34. Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G}$ be a sub- σ -algebra of $\mathcal{F}$ and $X \in L^1(\Omega, \mathcal{F}, P)$. The conditional expectation of X given the σ -algebra $\mathcal{G}$, denoted by $\mathbb{E}[X|\mathcal{G}]$, is the (P -a.s unique) $\mathcal{G}$ -measurable random variable satisfying

$$\mathbb{E}[\mathbf{1}_B X] = \int_B X dP = \int_B \mathbb{E}[X|\mathcal{G}] dP = \mathbb{E}[\mathbf{1}_B \mathbb{E}[X|\mathcal{G}]], \quad \forall B \in \mathcal{G}.$$

One can show, using the Radon-Nikodym theorem, that the conditional expectation is well defined.

LEMMA 2.35. Assume that $X, Y \in L^1$.

- (1) *Conservation of the expectation:* $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.
- (2) *Linearity:* $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$, for all $a, b \in \mathbb{R}$.
- (3) *Conditional Jensen's inequality:* Let $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function, $\varphi(X) \in L^1$. Then,

$$\varphi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\varphi(X)|\mathcal{G}].$$

- (4) *Tower property:* Let $\mathcal{H} \subseteq \mathcal{G}$ be a sub- σ -algebra of $\mathcal{G}$. Then,

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

- (5) *“What is measurable goes out”:* Assume that $XZ \in L^1$ and that Z is $\mathcal{G}$ -measurable. Then,

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

In particular, if X is $\mathcal{G}$ -measurable then $\mathbb{E}[X|\mathcal{G}] = X$.

- (6) *“Role of independence”:* If $\mathcal{H}$ is a σ -algebra independent of $\sigma(\mathcal{G}, \sigma(X))$, then

$$\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X|\mathcal{G}].$$

In particular, if X is independent of $\mathcal{H}$ then $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$.

- (7) *Substitution property:* Let $h : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ be a measurable function. If X is $\mathcal{G}$ -measurable, $h(X, Y) \in L^1$, and Y is independent of $\mathcal{G}$ then

$$\mathbb{E}[h(X, Y)|\mathcal{G}] = \mathbb{E}[h(a, Y)]|_{a=X}.$$

PROPOSITION 2.36 (Doob-Dynkin lemma). If $Y : \Omega \rightarrow \mathbb{R}^d$, then $X : \Omega \rightarrow \mathbb{R}$ is $\sigma(Y)$ -measurable if and only if there exists a measurable function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $X = \varphi(Y)$.

2.7. Exercises

- (1) Let X be a discrete random variable, that is, X only can take a countable number of values. Describe how is the σ -algebra $\sigma(X)$. Describe how are the $\sigma(X)$ measurable functions.
- (2) Let X be a random variable and $Y = X^2$. Show that Y is also a random variable. Is X measurable with respect to $\sigma(Y)$?
- (3) Let Q be the distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ given by

$$Q((a, b]) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz, \quad a < b \in \mathbb{R}.$$

Construct a probability space and a random variable Z defined on it such that the law of Z is given by Q . We say that the random variable Z is a standard Normal (or Gaussian) random variable. Compute its mean and variance. Let $X = \sigma Z + \mu$, where $\mu \in \mathbb{R}$ and $\sigma > 0$. Check that X is a random variable, find its distribution function and compute $\mathbb{E}[X]$ and $\text{Var}[X]$. We say that X is normal random variable with mean μ and variance σ^2 and we denote it by $X \sim \mathcal{N}(\mu, \sigma^2)$.

- (4) Let $X \sim \mathcal{N}(\mu, \sigma^2)$ defined on some probability space $(\Omega, \mathcal{F}, P)$. Compute $\psi(\theta) = \mathbb{E}[e^{\theta X}]$ for $\theta \in \mathbb{R}$. Define $L(X; \theta) := e^{\theta X} / \psi(\theta)$ and show that $Q_\theta(A) = \mathbb{E}[L(X; \theta) \mathbf{1}_A]$, $A \in \mathcal{F}$ defines a probability measure on $(\Omega, \mathcal{F})$. Show that $Q_\theta \ll P$. Find the law of X under Q_θ , that is, the law of X as a random variable defined on $(\Omega, \mathcal{F}, Q_\theta)$.
- (5) Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Define $Y = \exp(X)$. Show that Y is a random variable. The law of Y is called lognormal with mean μ and variance σ^2 , show that $P_Y \ll \lambda$ and find its density function $\frac{dP_Y}{d\lambda}$. Compute $\mathbb{E}[Y^n]$, $n \geq 1$ and $\text{Var}[Y]$.
- (6) Let X and Y be two independent and identically distributed (i.i.d.) random variables with law $\mathcal{N}(\mu, \sigma^2)$. Find the density of $(U, V) = (X + Y, X - Y)$. Under which conditions on μ and σ^2 are U and V independent?
- (7) Let X and Y be two i.i.d. random variables with law $\mathcal{N}(0, \sigma^2)$ independent random variables. Find the density of $(U, V) = (\sqrt{X^2 + Y^2}, X/Y)$, where V is defined as zero if $Y = 0$. Are U and V independent?
- (8) Let $Y \sim \mathcal{N}(0, 1)$ and for $a > 0$ define $Z = Y \mathbf{1}_{\{|Y| \leq a\}} - Y \mathbf{1}_{\{|Y| > a\}}$. Show that Z is Gaussian r.v. but (Y, Z) is not multivariate Gaussian.
- (9) Let X be a random variable and $f : \mathbb{R} \rightarrow \mathbb{R}_+$ an increasing, Borel measurable function. Let $a \in \mathbb{R}$ such that $f(a) > 0$. Prove that

$$P(X \geq a) \leq \frac{\mathbb{E}[f(X)]}{f(a)}.$$

- (10) Let $Y \in L^2(\Omega, \mathcal{F}, P)$. Show that if $\mathbb{E}[Y|X] = X$ and $\mathbb{E}[Y^2|X] = X^2$ then $X = Y$, P -a.s.
- (11) In the setup of Exercise 4, let $\mathcal{G}$ be a sub- σ -algebra of $\mathcal{F}$. Show that for any $Y \in L^1(\Omega, \mathcal{F}, Q_\theta)$ one has that

$$\mathbb{E}_{Q_\theta}[Y|\mathcal{G}] = \frac{\mathbb{E}[Y L(X; \theta) | \mathcal{G}]}{\mathbb{E}[L(X; \theta) | \mathcal{G}]}, \quad Q_\theta\text{-a.s.}$$

- (12) Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $K > 0$. Compute

$$\mathbb{E}[\max(0, e^X - K)].$$

You can express the solution in terms of the cumulative distribution function of a standard normal random variable

$$\Phi(x) := \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz.$$

CHAPTER 3

Review of Stochastic Processes

3.1. Basic Definitions

DEFINITION 3.1. Let $(E, \mathcal{E})$ be a measurable space and I a set. An E -valued stochastic process X indexed by I is a family of E -valued random variables $X = \{X_t\}_{t \in I}$ defined on the same probability space $(\Omega, \mathcal{F}, P)$.

We can think of a stochastic process as a function

$$\begin{aligned} X : I \times \Omega &\longrightarrow E \\ (t, \omega) &\mapsto X_t(\omega) \end{aligned} .$$

For every $\omega \in \Omega$ fixed, the process X defines a function

$$\begin{aligned} X.(\omega) : I &\longrightarrow E \\ t &\mapsto X_t(\omega) \end{aligned} ,$$

which is called a *trajectory* or a *sample path* of the process. Hence, we can look at X as a mapping

$$\begin{aligned} X : \Omega &\longrightarrow E^I \\ \omega &\mapsto X.(\omega) \end{aligned} ,$$

where E^I is the cartesian product of I copies of E , which is the set of all functions from I to E . That is, we can see X as a mapping from Ω to a space of functions.

REMARK 3.2. We will usually take $E = \mathbb{R}$ or $E = \mathbb{R}^d$. Moreover, we will usually take $I = \mathbb{R}_+$, $I = \mathbb{N}$ or $I = [0, T]$, where $T \in \mathbb{R}_+$.

3.2. Law of a Stochastic Process

If we think of a stochastic process as a mapping $X : \Omega \rightarrow \mathbb{R}^{\mathbb{R}_+}$ and we put on $\mathbb{R}^{\mathbb{R}_+}$ a σ -algebra such that X is measurable, we will have an $\mathbb{R}^{\mathbb{R}_+}$ -valued random variable and we can talk about the law of the process X . We expect the law of the process X to be related to the laws of the random variables X_t . For instance, what is the value of $P(X_t \in B)$ should make sense as a question on the r.v. X_t as well as a question on the law of X . Moreover, the answer must coincide.

We will consider in $\mathbb{R}^{\mathbb{R}_+}$ the product σ -algebra of $\mathbb{R}_+$ copies of $\mathcal{B}(\mathbb{R})$, denoted by $\mathcal{B}(\mathbb{R})^{\mathbb{R}_+}$. Recall that this σ -algebra is the smallest σ -algebra that makes the projections $\{\pi_t\}_{t \in \mathbb{R}_+}$, that is,

$$\begin{aligned} \pi_t : \mathbb{R}^{\mathbb{R}_+} &\longrightarrow \mathbb{R} \\ f &\mapsto f(t) \end{aligned} ,$$

Borel measurable functions for all $t \in \mathbb{R}_+$.

PROPOSITION 3.3. If $X = \{X_t\}_{t \in \mathbb{R}_+}$ is a stochastic process and we consider in $\mathbb{R}^{\mathbb{R}_+}$ the product σ -algebra $\mathcal{B}(\mathbb{R})^{\mathbb{R}_+}$, then the application $X : \Omega \rightarrow \mathbb{R}^{\mathbb{R}_+}$ is measurable.

DEFINITION 3.4. The law of real stochastic process $X = \{X_t\}_{t \in \mathbb{R}_+}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ is the law of the random variable $X : \Omega \rightarrow \mathbb{R}^{\mathbb{R}_+}$, where the σ -algebra on $\mathbb{R}^{\mathbb{R}_+}$ is $\mathcal{B}(\mathbb{R})^{\mathbb{R}_+}$.

Let μ be a probability measure on $(\mathbb{R}^{\mathbb{R}_+}, \mathcal{B}(\mathbb{R})^{\mathbb{R}_+})$. We can consider, for each $J \subset \mathbb{R}_+$ finite, the image measure of μ by the projection π_J , which we will denote by μ_J . That is, μ_J is the probability measure on $(\mathbb{R}^{\#J}, \mathcal{B}(\mathbb{R})^{\#J})$ defined by $\mu_J(B) := \mu(\pi_J^{-1}(B))$, $B \in \mathcal{B}(\mathbb{R})^{\#J}$.

DEFINITION 3.5. The family of probabilities $\{\mu_J : J \subset \mathbb{R}_+, \#J < +\infty\}$ is called the family of finite dimensional distributions of μ .

This family satisfies that if $K \subset J$ then

$$(3.2.1) \quad \mu_K(B) = \mu_J(\pi_{J,K}^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R})^{\#K},$$

where $\pi_{J,K}$ is the natural projection of $\mathbb{R}^{\#J}$ on $\mathbb{R}^{\#K}$.

THEOREM 3.6 (Kolmogorov's Extension). Let $\{\mu_J : J \subset \mathbb{R}_+, \#J < +\infty\}$ be a family of probabilities satisfying the consistency property (3.2.1). Then there exists a unique probability measure μ on $(\mathbb{R}^{\mathbb{R}_+}, \mathcal{B}(\mathbb{R})^{\mathbb{R}_+})$ such that the family of its finite dimensional distributions coincides with the initially given family. The probability measure μ is called the projective limit probability measure of the family.

COROLLARY 3.7. The law of a stochastic process is completely determined by the associated family of finite dimensional distributions, that is, by

$$\mu_{X;t_1, \dots, t_d}(B) := P((X_{t_1}, \dots, X_{t_d}) \in B), \quad B \in \mathcal{B}(\mathbb{R}^d), \quad t_1, \dots, t_d \in \mathbb{R}, \quad d \in \mathbb{N}.$$

The previous theorem shows that provided we have a reasonable (consistent) family of finite dimensional distributions we can construct a real stochastic process (actually a probability space containing it) such that its law has as family of finite dimensional distributions this family. Note that we can always realize the process on $(\Omega, \mathcal{F}, P) = (\mathbb{R}^{\mathbb{R}_+}, \mathcal{B}(\mathbb{R})^{\mathbb{R}_+}, \mu)$ where μ is the projective limit probability measure and the process is $X(f) = f$, where $f \in \mathbb{R}^{\mathbb{R}_+}$.

The following set of definitions only depend on the law of a process X .

DEFINITION 3.8. Let X be a stochastic process. Then

- (1) X is called Gaussian if its family of finite dimensional distributions are multivariate Gaussian distributions.
- (2) X has independent increments if $X_t - X_s$ is independent of $X_r - X_u$, for all $u \leq r \leq s \leq t$.
- (3) X is stationary if for all $h > 0$ and for all $t_1, \dots, t_n \in \mathbb{R}_+, n \in \mathbb{N}$, we have that

$$\mathcal{L}(X_{t_1}, \dots, X_{t_n}) = \mathcal{L}(X_{t_1+h}, \dots, X_{t_n+h}).$$

- (4) X is stationary in wide sense if $\mathbb{E}[X_t^2] < \infty$ and for all $h > 0$

$$E[X_t] = E[X_{t+h}],$$

$$\text{Cov}(X_t, X_s) = \text{Cov}(X_{t+h}, X_{s+h}), \quad t, s \in \mathbb{R}_+.$$

(5) X has stationary increments if for all $s \leq t \in \mathbb{R}_+$ we have that

$$\mathcal{L}(X_t - X_s) = \mathcal{L}(X_{t-s}).$$

Note that, as the multivariate Gaussian distributions are characterized by the moments of first and second order, the law of a Gaussian process X is characterized by $E[X_t]$ and $\text{Cov}(X_t, X_s), s, t \in \mathbb{R}_+$.

3.3. Equivalence of Processes and Path Properties

Suppose that we have two processes X and Y defined on the same probability space and with the same law. As application from Ω to $\mathbb{R}^{\mathbb{R}_+}$, X and Y may be very different. For instance, Let X_t be $\text{Unif}[0, 1]$, independent and we define $Y_t = 1 - X_t$. Then, X and Y have the same law, but $P(X_t = Y_t) = 0$ for all $t \in \mathbb{R}_+$. We need stronger concepts of equality for processes than equality in law.

DEFINITION 3.9. Let X and Y be two processes defined on the same probability space $(\Omega, \mathcal{F}, P)$. We say that Y is a *modification* or a *version* of X if

$$\forall t \in \mathbb{R}_+, \quad P(X_t = Y_t) = 1.$$

DEFINITION 3.10. Let X and Y be two processes defined on the same probability space $(\Omega, \mathcal{F}, P)$. We say that Y is indistinguishable of X if the set

$$\{\omega : X_t(\omega) \neq Y_t(\omega), \quad \forall t \in \mathbb{R}_+\},$$

is a negligible set.

If X and Y are indistinguishable then one is a version of the other. If X is a version of Y then X and Y have the same law.

THEOREM 3.11 (Kolmogorov's continuity criterion). *Let X be a stochastic process defined in some probability space $(\Omega, \mathcal{F}, P)$. Assume that for every $T > 0$, there exist $\alpha > 0, r > 0$ and $K > 0$ such that*

$$\mathbb{E}[|X_t - X_s|^r] \leq K |t - s|^{1+\alpha}, \quad 0 \leq s < t \leq T.$$

Then, there exists a modification $\tilde{X}$ of X with all the paths being continuous.

Regarding path regularity properties the most important classes of processes are the following.

DEFINITION 3.12. Let $X = \{X_t\}_{t \in \mathbb{R}_+}$ be a real valued stochastic process.

- (1) We say that X is continuous if its sample paths are continuous.
- (2) We say that X is càdlàg or RCLL if its sample paths are continuous from the right and have limits from the left. That is $X_t = X_{t+} := \lim_{s \downarrow t} X_s$ and $X_{t-} := \lim_{s \uparrow t} X_s$ exists and is finite.
- (3) We say that X is càglàd or LCRL if its sample paths are continuous from the left and have limits from the right. That is $X_t = X_{t-} := \lim_{s \uparrow t} X_s$ and $X_{t+} := \lim_{s \downarrow t} X_s$ exists and is finite.

The properties in the previous definition may hold only P -a.s and then we will say that a process P -a.s continuous or P -a.s. RCLL, etc... Finally, if we have a process Y that is version of a process X and both are P -a.s. RCLL (or P -a.s. LCRL) then X and Y are indistinguishable. In particular, if X and Y are P -a.s. continuous and Y is a version of X then X and Y are indistinguishable.

3.4. Measurability Properties of Stochastic Processes

Note that from the measurability of each X_t is not possible to deduce the measurability of X as a function from $\mathbb{R}_+ \times \Omega$ to $\mathbb{R}$.

DEFINITION 3.13. We say that a process X is (jointly) measurable when it is $(\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}, \mathcal{B}(\mathbb{R}))$ -measurable, that is, it is measurable as a function $X : \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}$.

THEOREM 3.14. Let X be a measurable stochastic processes. If

$$\int_{\mathbb{R}_+ \times \Omega} |X(t, \omega)| \lambda \otimes P(dt, d\omega) < \infty,$$

where λ is the Lebesgue measure, then

$$\int_{\mathbb{R}_+ \times \Omega} X(t, \omega) \lambda \otimes P(dt, d\omega) = \int_{\mathbb{R}_+} \mathbb{E}[X_t] dt = \mathbb{E} \left[\int_{\mathbb{R}_+} X_t dt \right] < \infty.$$

DEFINITION 3.15. Let $(\Omega, \mathcal{F}, P)$ be a probability space. A filtration on $(\Omega, \mathcal{F}, P)$ is a family $\mathbb{F} = \{\mathcal{F}_t\}_{t \in \mathbb{R}_+}$ of sub- σ -algebras of $\mathcal{F}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ if $s \leq t$. We say that $(\Omega, \mathcal{F}, \mathbb{F}, P)$ is a filtered probability space.

DEFINITION 3.16. A filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t \in \mathbb{R}_+}$ is called right continuous if $\mathcal{F}_t = \mathcal{F}_{t+}$ for every $t \in \mathbb{R}_+$, where $\mathcal{F}_{t+} = \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}$. The σ -algebra $\mathcal{F}_{t+}$ represents what we know if we look ahead an infinitesimal amount of time.

DEFINITION 3.17. A filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$ is said to satisfy the “usual conditions” if $(\Omega, \mathcal{F}, P)$ is complete and

- (1) $\mathcal{F}_0$ contains all negligible sets in $\mathcal{F}$.
- (2) $\mathbb{F}$ is right continuous.

DEFINITION 3.18. A stochastic process X is $\mathbb{F}$ -adapted if for all $t \in \mathbb{R}_+$ the r.v. X_t is $\mathcal{F}_t$ -measurable.

Note that we can always find a filtration with respect to which a process X is adapted. We can consider the *natural* filtration generated by X , denoted by $\mathbb{F}^X$ and defined by

$$\mathbb{F}^X = \{\mathcal{F}_t^X := \sigma(X_s, s \leq t)\}_{t \in \mathbb{R}_+}.$$

The natural filtration $\mathbb{F}^X$ represents the information revealed by the process X as time evolves. For instance, the process $Y = \{Y_t := X_t^2\}_{t \in \mathbb{R}_+}$ is $\mathbb{F}^X$ -adapted while the process $Z = \{Z_t = X_1 + X_t^2\}_{t \in \mathbb{R}_+}$ is not $\mathbb{F}^X$ -adapted.

Given a process X , the filtration $\mathbb{F}^X$ does not need to satisfy the usual conditions. However, we can always consider (and we do it) the *minimal augmented* (also know as *the usual augmentation of the*) filtration generated by X , which is the smallest filtration satisfying the usual conditions and with respect to which the process X is adapted. This

is done by enlarging $\mathbb{F}^X$ with $\mathcal{N}_P$, the P -negligible sets in $\mathcal{F}$, and then taking its limit from the right, that is, considering the filtration

$$\left\{ \bigcap_{\varepsilon > 0} \sigma(\mathcal{F}_{t+\varepsilon}^X \cup \mathcal{N}_P) \right\}_{t \in \mathbb{R}_+}.$$

We will use the same notation to denote the natural filtration generated by X and the minimal augmented filtration generated by X , that is, $\mathbb{F}^X$. Moreover, all the processes that we will be dealing with in this course are examples of the so called strong Markov processes: Brownian motion, solutions of stochastic differential equations, Lévy processes... For these processes, it suffices to augment their natural filtration with the P -negligible sets of $\mathcal{F}$ to automatically get a right continuous filtration. Many results that we will see in this course need the filtration under consideration to satisfy the usual conditions to be true, but not all. That the filtration $\mathbb{F}$ contains all the P -negligible sets ensure, for instance, that if a process X is $\mathbb{F}$ -adapted and Y is a version of X , then Y is also $\mathbb{F}$ -adapted.

Stopping times are one of the most useful concepts in stochastic analysis as they are the basic tool for localization procedures.

DEFINITION 3.19. A random variable $\tau : \Omega \rightarrow [0, +\infty]$ is a stopping time with respect to the filtration $\mathbb{F}$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \in \mathbb{R}_+$.

DEFINITION 3.20. Let τ be a $\mathbb{F}$ -stopping time. Then, the σ -algebra of events occurring up to time τ can be defined by

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : \text{for each } t > 0, A \cap \{\tau \leq t\} \in \mathcal{F}_t\}.$$

One can check that if $\tau = t$, P -a.s. then $\mathcal{F}_\tau = \mathcal{F}_t$ and that if $\tau_1 \leq \tau_2$, P -a.s. then $\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}$. Therefore, one can use stopping times to randomize the flow of information, that is, to randomize a filtration.

The right continuity of the filtration ensure that some random times, for instance the first time that a $\mathbb{F}$ -adapted process with right continuous paths X hits a open set $A \in \mathcal{B}(\mathbb{R})$, mathematically defined by $\tau_A(\omega) = \inf\{t \in \mathbb{R}_+ : X_t(\omega) \in A\}$, are actually stopping times.

DEFINITION 3.21. A stochastic process X is progressively measurable or progressive with respect to $\mathbb{F}$ if, for all $t \in \mathbb{R}_+$, the restriction

$$\begin{aligned} X|_{[0,t] \times \Omega} : [0,t] \times \Omega &\longrightarrow \mathbb{R} \\ (s, \omega) &\longmapsto X_s(\omega) \end{aligned},$$

is $(\mathcal{B}([0,t]) \otimes \mathcal{F}_t, \mathcal{B}(\mathbb{R}))$ -measurable.

The notion of progressive measurability is useful to ensure that some processes are adapted. For instance, if X is progressive then the process $Y_t = \int_0^t X_s ds$ is adapted. Moreover, if X is progressive and τ is a stopping time with respect to the same filtration, then the random variable $X_\tau := X_{\tau(\omega)}(\omega)$ is $\mathcal{F}_\tau$ -measurable and the process stopped at X , defined by $X^\tau := X_{\min(t,\tau)}$ is also progressive.

DEFINITION 3.22. A stochastic process X is predictable or previsible with respect to $\mathbb{F}$ if, as a mapping from $\mathbb{R}_+ \times \Omega$ to $\mathbb{R}$, X is $(\mathcal{P}, \mathcal{B}(\mathbb{R}))$ -measurable, where $\mathcal{P}$ is the predictable σ -algebra, that is, the σ -algebra generated by all left-continuous, $\mathbb{F}$ -adapted processes.

The notion of predictability will be essential to construct the stochastic integral with respect to processes with jumps.

PROPOSITION 3.23. *Let X be a stochastic process. Then,*

- (1) *If X is predictable then it is progressive.*
- (2) *If X is progressive then it is measurable and adapted.*
- (3) *If X is adapted and has right continuous paths (or left continuous paths) then it is progressive.*
- (4) *There are progressive processes that are not predictable.*
- (5) *There are adapted processes that are not measurable.*
- (6) *There are measurable processes that are not adapted.*
- (7) *There are adapted and measurable processes that are not progressive, but one can always find a modification that it is progressive.*

3.5. Examples of Stochastic Processes

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

3.5.1. Brownian Motion and Related Processes.

DEFINITION 3.24. A **$\mathbb{F}$ -Brownian motion** with variance parameter σ^2 is a stochastic process W adapted to $\mathbb{F}$ satisfying

- (1) W has continuous paths P -a.s.,
- (2) $W_0 = 0$, P -a.s.,
- (3) For all $0 \leq s < t$, the random variable $W_t - W_s$ is independent of $\mathcal{F}_s$.
- (4) For all $0 \leq s < t$, the law of $W_t - W_s$ is a $\mathcal{N}(0, \sigma^2(t - s))$.

REMARK 3.25.

- (1) If $\sigma^2 = 1$, we say that W is a standard $\mathbb{F}$ -Brownian motion.
- (2) We can define a Brownian motion without any reference to a filtration by just requiring that W has independent increments.
- (3) To have a $\mathbb{F}$ -Brownian motion is a stronger statement than to only have a Brownian motion. If $W_t - W_s$ is independent of $\mathcal{F}_s$ then it is independent from $W_u - W_v$, $0 \leq v \leq u \leq s$ and W is a process with independent increments.
- (4) If we only have a Brownian motion then we also have a $\mathbb{F}^W = \{\mathcal{F}_t^W\}_{t \in \mathbb{R}_+}$ -Brownian motion but W does not need to be a $\mathbb{F}$ -Brownian motion where $\mathbb{F}$ is an enlargement of $\mathbb{F}$, i.e., $\mathcal{F}_t^W \subseteq \mathcal{F}_t$, $t \in \mathbb{R}_+$.
- (5) One can show that a Brownian motion exists using Daniell-Kolmogorov's extension theorem and Kolmogorov's continuity criterion.

DEFINITION 3.26. A d -dimensional Brownian motion $W = \{(W_t^1, \dots, W_t^d)\}_{t \in \mathbb{R}_+}$ is a stochastic process with values in $\mathbb{R}^d$ such that each component W^i , $i = 1, \dots, d$ is a real Brownian motion independent of each other.

For instance the first component W^1 of a d -dimensional Brownian motion is a $\mathbb{F}^{W^1}$ -Brownian motion but it is also a $\mathbb{F}^W$ -Brownian motion. In general, if we have a $\mathbb{F}$ -Brownian motion and we enlarge the filtration $\mathbb{F}$ with some events independent from $\mathbb{F}$ we still have a Brownian motion with respect to the enlarged filtration.

DEFINITION 3.27. A process Y is a **Brownian motion with drift** μ and volatility σ if it can be written as

$$Y_t = \mu t + \sigma W_t, \quad t \in \mathbb{R}_+,$$

where W is a standard Brownian motion.

This process has almost all the properties of Brownian motion. It starts at zero, has independent increments and the increments have Gaussian laws. However, the Gaussian law of $Y_t - Y_s, s < t$ is not centered but has mean $\mu(t - s)$. Brownian motion with drift is also a Gaussian process.

This process was used in the first attempt to model stock prices behavior by Bachelier. Many people consider Bachelier's PhD dissertation entitled *Théorie de la Espéculation* (1900) the beginning of Mathematical Finance. Brownian motion with drift, as a model for stock prices, allows for negative prices with positive probability, which is major pitfall.

DEFINITION 3.28. A process S is a **geometric Brownian motion** (or **exponential Brownian motion**) with drift μ and volatility σ if it can be written as

$$S_t = \exp(\mu t + \sigma W_t), \quad t \in \mathbb{R}_+,$$

where W is a standard Brownian motion.

The law of a geometric Brownian motion is not Gaussian. Actually, the random variable S_t has lognormal distribution with mean μt and variance $\sigma^2 t$. It does not have independent and stationary increments like Brownian motion or Brownian motion with drift. On the other hand, its relative increments

$$\frac{S_{t_n} - S_{t_{n-1}}}{S_{t_{n-1}}}, \frac{S_{t_{n-1}} - S_{t_{n-2}}}{S_{t_{n-2}}}, \dots, \frac{S_{t_1} - S_{t_0}}{S_{t_0}}, \quad 0 \leq t_0 < t_1 < \dots < t_n,$$

are independent and stationary. Equivalently,

$$\frac{S_{t_n}}{S_{t_{n-1}}}, \frac{S_{t_{n-1}}}{S_{t_{n-2}}}, \dots, \frac{S_{t_1}}{S_{t_0}}, \quad 0 \leq t_0 < t_1 < \dots < t_n,$$

and

$$\log \left(\frac{S_{t_n}}{S_{t_{n-1}}} \right), \log \left(\frac{S_{t_{n-1}}}{S_{t_{n-2}}} \right), \dots, \log \left(\frac{S_{t_1}}{S_{t_0}} \right), \quad 0 \leq t_0 < t_1 < \dots < t_n,$$

are also independent and stationary. Moreover, the law of $\frac{S_t}{S_s}, s < t$ is a lognormal distribution with parameters $\mu(t - s)$ and $\sigma^2(t - s)$ or, equivalently, the law of $\log \left(\frac{S_t}{S_s} \right), s < t$ is $\mathcal{N}(\mu(t - s), \sigma^2(t - s))$.

This process is the standard model for stock prices. Note that in this case the prices are always positive. The model was first used by Samuelson (1964) and later on by Black-Scholes (1973) in their theory of option pricing.

3.5.2. Poisson Process and Related Processes.

DEFINITION 3.29. A **$\mathbb{F}$ -Poisson process** with intensity/parameter $\lambda > 0$ is a stochastic process N adapted to $\mathbb{F}$ satisfying:

- (1) N has R.C.L.L paths P -a.s.
- (2) $N_0 = 0$, P -a.s.,
- (3) For all $0 \leq s < t$, the random variable $N_t - N_s$ is independent of $\mathcal{F}_s$.
- (4) For all $0 \leq s < t$, the law of $N_t - N_s$ is a $\text{Pois}(\lambda(t-s))$, i.e.,

$$P(N_t - N_s = n) = \frac{(\lambda(t-s))^n}{n!} e^{-\lambda(t-s)}.$$

REMARK 3.30.

- (1) If $\lambda = 1$, we say that N is a standard $\mathbb{F}$ -Poisson process.
- (2) We can define a Poisson process without any reference to a filtration by just requiring that N has independent increments.
- (3) As for the Brownian motion, N it is always a Poisson process with respect to its natural filtration but it may not be a Poisson process with respect to an enlarged filtration.
- (4) The existence of a Poisson process is again guaranteed by the Daniell-Kolmogorov's extension theorem.

DEFINITION 3.31. A process $\tilde{N}$ is a **compensated Poisson process** with intensity λ if it can be written as

$$\tilde{N}_t = N_t - \lambda t, \quad t \in \mathbb{R}_+,$$

where N is a **Poisson process** with intensity λ .

Note that $\mathbb{E}[N_t] = \lambda t$ and, therefore, $\mathbb{E}[\tilde{N}_t] = 0$.

DEFINITION 3.32. Given a sequence $\{X_k\}_{k \geq 1}$ of independent, identically distributed random variables and N an independent Poisson process, the process

$$Y_t = \sum_{k=1}^{N_t} X_k, \quad t \in \mathbb{R}_+,$$

is called a **compound Poisson process**.

3.6. Levy Processes

The Brownian motion and the Poisson process are particular cases of a more general type of process.

DEFINITION 3.33. A stochastic process $L = \{L_t\}_{t \in \mathbb{R}_+}$ is a Lévy process starting at zero if it satisfies:

- (1) $L_0 = 0$, P -a.s.,
- (2) L has independent increments,
- (3) L has stationary increments, i.e., for all $0 \leq s < t$, the law of $L_t - L_s$ coincides with the law of L_{t-s} .
- (4) L is stochastically continuous, i.e., $\lim_{s \rightarrow t} P(|L_t - L_s| > \varepsilon) = 0, \forall \varepsilon > 0, t \in \mathbb{R}_+$.

REMARK 3.34.

- (1) Condition (4) in Definition 3.33 serves to exclude processes with jumps at fixed (nonrandom) times.
- (2) Given a Lévy process we can choose a unique modification whose paths are right continuous and with left limits. This property is satisfied by the Brownian motion and the Poisson process.
- (3) Obviously, conditions (2) and (3) strongly restrict the possible laws of the process L and its family of finite dimensional distributions. Actually, a Lévy process L is determined by the law of L_1 but this law cannot be arbitrary, it must be infinitely divisible.
- (4) Given a filtration $\mathbb{F}$ one can also define a $\mathbb{F}$ -Lévy process analogously to the $\mathbb{F}$ -Brownian motion, that is, by requiring that for all $0 \leq s < t$, the random variable $L_t - L_s$ is independent of $\mathbb{F}$.

3.7. Martingales

The following type of processes will be essential in the theory of stochastic integration and option pricing.

DEFINITION 3.35. A stochastic process X is a martingale if $\mathbb{E}[|X_t|] < \infty, t \in \mathbb{R}_+$ and

$$\mathbb{E}[X_t | \sigma(X_u : 0 \leq u \leq s)] = X_s, \quad 0 \leq s \leq t.$$

We can also define a martingale with respect to a filtration $\mathbb{F}$ and this is definition that we will use the most.

DEFINITION 3.36. Let X be a stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$. The process X is a $\mathbb{F}$ -martingale if $\mathbb{E}[|X_t|] < \infty, t \in \mathbb{R}_+$, X is $\mathbb{F}$ -adapted and

$$\mathbb{E}[X_t | \mathcal{F}_s] = X_s, \quad 0 \leq s \leq t.$$

The martingale property essentially says that the best prediction of the value (or the expected value) of the process at some future time t given the information up to time s is precisely the value of the process at time s .

EXAMPLE 3.37. A $\mathbb{F}$ -Brownian motion is a $\mathbb{F}$ -martingale: by definition W is $\mathbb{F}$ -adapted, as $W_t \sim \mathcal{N}(0, t)$ has finite moments of all orders and

$$\mathbb{E}[W_t - W_s | \mathcal{F}_s] = \mathbb{E}[W_t - W_s] = 0,$$

where we have used that $W_t - W_s$ is independent from $\mathcal{F}_s$, that if Z is independent from $\mathcal{G}$ we have that $\mathbb{E}[Z | \mathcal{G}] = \mathbb{E}[Z]$ and that $\mathbb{E}[W_t] = 0$, for all $t \in \mathbb{R}_+$. From the previous equality and using that W_s is $\mathcal{F}_s$ -measurable and, hence, $\mathbb{E}[W_s | \mathcal{F}_s] = W_s$, we can conclude.

The following result will be needed in the development of the stochastic integral.

THEOREM 3.38. If $\{M_t\}_{t \geq 0}$ is a martingale with almost surely continuous paths then we have that:

(1) For all $T \geq 0, \lambda > 0$ and $p \geq 1$

$$P \left(\sup_{0 \leq t \leq T} |M_t| > \lambda \right) \leq \frac{1}{\lambda^p} \mathbb{E} [|M_T|^p].$$

(2) For all $T \geq 0, p > 1$

$$\mathbb{E} [|M_T|^p] \leq \mathbb{E} \left[\sup_{0 \leq t \leq T} |M_t|^p \right] \leq \left(\frac{p}{p-1} \right)^p \mathbb{E} [|M_T|^p].$$

3.8. Markov Processes

DEFINITION 3.39. A real-valued stochastic process $X = \{X_t\}_{t \in \mathbb{R}_+}$ is a Markov process if

$$P(X_t \in B | \sigma(X_{s_1}, X_{s_2}, \dots, X_{s_n})) = P(X_t \in B | \sigma(X_{s_n})),$$

for all $0 \leq s_1 < s_2 < \dots < s_n \leq t$ and $B \in \mathcal{B}(\mathbb{R})$.

A Markov process is the stochastic analogous of what happens in a system governed by a differential equation with unique solution: A value given at some time determines what happens after that time, we do not need more information from the past. Essentially, a Markov process is independent of the past given the present.

We can also consider the Markov property with respect to a filtration $\mathbb{F}$.

DEFINITION 3.40. Let X be a stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$. The process X is a Markov process with respect to $\mathbb{F}$ if X is $\mathbb{F}$ -adapted and

$$P(X_t \in B | \mathcal{F}_s) = P(X_t \in B | \sigma(X_s)),$$

for all $0 \leq s < t$ and $B \in \mathcal{B}(\mathbb{R})$.

The following proposition provides an alternative (and more useful for our goals) characterization of the Markov property.

PROPOSITION 3.41. *Under the assumptions of the previous definition the following two statements are equivalent:*

(1) For all $0 \leq s < t$ and $B \in \mathcal{B}(\mathbb{R})$,

$$P(X_t \in B | \mathcal{F}_s) = P(X_t \in B | \sigma(X_s)).$$

(2) For all $g : \mathbb{R} \rightarrow \mathbb{R}$ Borel measurable such that $\mathbb{E} [|g(X_t)|] < \infty, t \in \mathbb{R}_+$ and for all $0 \leq s \leq t$

$$\mathbb{E} [g(X_t) | \mathcal{F}_s] = \mathbb{E} [g(X_t) | X_s].$$

Note that if a process X is Markov with respect to a filtration $\mathbb{F}$ then it is Markov with respect to its natural filtration.

Let X be a Markov process and for each $0 \leq s \leq t < \infty, A \in \mathcal{B}(\mathbb{R}), x \in \mathbb{R}$, define

$$p_{s,t}(x, A) = P(X_t \in A | X_s = x).$$

By the properties of conditional probability, we can choose $p_{s,t}(x, A)$ to be a transition probability measure from $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, that is, $p_{s,t} : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ is a mapping satisfying

- (1) $\forall x \in \mathbb{R}, p_{s,t}(x, \cdot) : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- (2) $\forall A \in \mathcal{B}(\mathbb{R}), p_{s,t}(\cdot, A) : \mathbb{R} \rightarrow [0, 1]$ is $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}_+))$ -measurable.

They give the probabilities of “transitions” of the process from the point x at time s to the set A at time t .

THEOREM 3.42 (Chapman-Kolmogorov). *If X is a Markov processes then for each $0 \leq r \leq s \leq t, x \in \mathbb{R}, A \in \mathcal{B}(\mathbb{R})$,*

$$p_{r,t}(x, A) = \int_{\mathbb{R}} p_{s,t}(y, A) p_{r,s}(x, dy).$$

DEFINITION 3.43. Let $\{p_{s,t}; 0 \leq s \leq t < \infty\}$ be a family of mappings from $\mathbb{R} \times \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$. We say that they are a family of Markov transition probabilities if, for each $0 \leq s \leq t$:

- (1) the maps $x \mapsto p_{s,t}(x, A)$ are measurable for each $A \in \mathcal{B}(\mathbb{R})$;
- (2) $p_{s,t}(x, \cdot)$ is a probability measure on $\mathcal{B}(\mathbb{R})$ for each $x \in \mathbb{R}$.
- (3) they satisfy the Chapman-Kolmogorov equations.

THEOREM 3.44. *If $\{p_{s,t}; 0 \leq s \leq t < \infty\}$ is a family of Markov transition probabilities and μ is a fixed probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, there exists a probability space $(\Omega, \mathcal{F}, P_\mu)$ and a Markov process $X = \{X_t\}_{t \in \mathbb{R}_+}$ defined on that space such that:*

- (1) $P(X_t \in A | X_s = x) = p_{s,t}(x, A)$ for each $0 \leq s \leq t, x \in \mathbb{R}$ and $A \in \mathcal{B}(\mathbb{R})$.
- (2) X_0 has law μ .

DEFINITION 3.45. A Markov process is said to be (time-) homogeneous if

$$p_{s,t}(x, A) = p_{0,t-s}(x, A),$$

for each $0 \leq s \leq t, x \in \mathbb{R}$ and $A \in \mathcal{B}(\mathbb{R})$. For homogeneous Markov processes we will always write the transition probabilities $p_{0,t}$ as p_t .

EXAMPLE 3.46. A $\mathbb{F}$ -Brownian motion is a $\mathbb{F}$ -Markov process with homogeneous transition probabilities given by

$$p_s(x, A) = \int_A \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{(y-x)^2}{2s}\right) dy,$$

and initial distribution the Dirac's delta distribution δ_0 .

3.9. Exercises

- (1) Recall that $X = \{X_t\}_{t \in \mathbb{R}_+}$ is a Gaussian process if for any $\{t_i\}_{i=1, \dots, n} \subset \mathbb{R}_+, n \in \mathbb{N}$ we have that the vector $(X_{t_1}, \dots, X_{t_n})$ is multivariate Gaussian. Prove that the following alternative definition of Brownian motion is equivalent to the one given in class. A real valued process X is a Brownian motion if X has P -a.s. continuous paths and is a centered Gaussian process with covariance function $K(s, t) = \mathbb{E}[X_s X_t] = \min(s, t)$.
- (2) Let $W = \{W_t\}_{t \in \mathbb{R}_+}$ be a Brownian motion and $a > 0$. Let $\mathbb{F}$ be the minimal augmented filtration generated by W . Which of the following processes are Brownian motions? Which are $\mathbb{F}$ -Brownian motions?
 - (a) $X_t = -W_t, t \in \mathbb{R}_+$.
 - (b) $X_t = W_{a+t} - W_a, t \in \mathbb{R}_+, .$

- (c) $X_t = W_{at^2}, t \in \mathbb{R}_+$.
- (3) Let $W = \{W_t\}_{t \in [0, T]}$ be a Brownian motion. Prove that the following processes are $\mathbb{F}^W$ -martingales:
- (a) $X_t = \exp(\theta W_t - \frac{\theta^2}{2}t), t \in [0, T]$.
- (b) $Z_t = W_t^2 - t, t \in [0, T]$.
- (4) Let $(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_t\}_{t \in \mathbb{R}_+}, P)$ and let $M = \{M_t\}_{t \in \mathbb{R}_+}$ be a $\mathbb{F}$ -martingale such that $\mathbb{E}[|M_t|^2] < \infty$. Prove that

$$\mathbb{E}[(M_t - M_s)^2 | \mathcal{F}_s] = \mathbb{E}[M_t^2 | \mathcal{F}_s] - \mathbb{E}[M_s^2 | \mathcal{F}_s], \quad s \leq t.$$

- (5) Let $(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_t\}_{t \in \mathbb{R}_+}, P)$ be a filtered probability space. We say that an $\mathbb{F}$ -adapted and integrable process $X = \{X_t\}_{t \in \mathbb{R}_+}$ is a submartingale if $\mathbb{E}[X_t | \mathcal{F}_s] \geq X_s, 0 \leq s \leq t$ (and we say that X is a supermartingale if $-X$ is a submartingale). Let $M = \{M_t\}_{t \in \mathbb{R}_+}$ be a $\mathbb{F}$ -martingale. Show that if φ is convex and $Y_t = \varphi(M_t)$ is integrable then Y is a submartingale.
- (6) Prove that a process $X = \{X_t\}_{t \in \mathbb{R}_+}$ with independent increments and with constant mean is a martingale.
- (7) Prove that an integrable process $X = \{X_t\}_{t \in \mathbb{R}_+}$ with independent increments is a Markov process.
- (8) Let $(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_t\}_{t \in \mathbb{R}_+}, P)$ be a filtered probability space and let $X \in L^p(\Omega, \mathcal{F}, P)$ for some $p \geq 1$. Show that $M_t = \mathbb{E}[X | \mathcal{F}_t]$ is a $\mathbb{F}$ -martingale and $M_t \in L^p$, for all $t \in \mathbb{R}_+$.
- (9) Let $M = \{M_n\}_{n \geq 0}$ be a discrete time martingale with respect to some filtration $\{\mathcal{F}_n\}_{n \geq 0}$ and let $H = \{H_n\}_{n \geq 0}$ be *predictable* process with respect to the same filtration, i.e., $H_0 = 0$ and H_n is $\mathcal{F}_{n-1}$ -measurable $\forall n \geq 1$. Show that if H is bounded ($|H_n| < C_n, P$ -a.s. for some constants $C_n, n \geq 1$), then the process $G = \{G_n\}_{n \geq 0}$ defined by

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

is a martingale.

- (10) Let $(\Omega, \mathcal{F}_T, \{\mathcal{F}_t\}_{t \in [0, T]}, P)$ be a filtered probability space. Assume that $Q \ll P$ with $Z_T := \frac{dQ}{dP}$. Prove that $Q|_{\mathcal{F}_t} \ll P|_{\mathcal{F}_t}$ for all $t \in [0, T]$ and, if we define $Z_t := \frac{dQ|_{\mathcal{F}_t}}{dP|_{\mathcal{F}_t}}$, then Z_t is a $\mathbb{F}$ -martingale under P . Prove that a $\mathcal{F}_t$ -adapted process Y is a martingale under Q if and only if ZY is a martingale under P .