B

Basics in Probability

The results below may be found in classical books such as [30, 65, 66, 89]. We provide recalls on probability spaces, random variables, convergence of random variables, then on conditional expectation, conditional probability and stochastic kernels. We conclude with the Monte Carlo method.

B.1 Probability Space

We give the definition of a probability space, after having recalled the notions of measurable space and of measure.

B.1.1 Measurable Space

A *σ-field* on a set Ω is a collection $\mathcal{A}$ of subsets of Ω such that:

- $\emptyset \in \mathcal{A}$,
- if the sequence $\{B_n\}_{n\in N}$ is such that $B_n \in \mathcal{A}$, for $n \in N$ where N is countable, then $\bigcup_{n\in N} B_n \in \mathcal{A}$,
- if $B \in \mathcal{A}$, then the complementary set $B^{\mathrm{c}} = \Omega\backslash B \in \mathcal{A}$.

Given any collection $\mathcal{C}$ of subsets of Ω, the σ-field $\sigma(\mathcal{C})$ generated by $\mathcal{C}$ is defined to be the smallest σ-field in Ω such that $\mathcal{C} \subset \sigma(\mathcal{C})$. It is called the σ-field *generated* by $\mathcal{C}$.

A *measurable space* is a set Ω together with a σ-field $\mathcal{A}$ on Ω, and is denoted by $(\Omega, \mathcal{A})$. The elements of $\mathcal{A}$ are called *measurable sets*.

Let $(\Omega, \mathcal{A})$ and $(\mathbb{Y}, \mathcal{Y})$ be two measurable spaces. A *mapping* $\boldsymbol{Y} : \Omega \to \mathbb{Y}$ is said to be *measurable* if $\boldsymbol{Y}^{-1}(\mathcal{Y}) \subset \mathcal{A}$. The collection $\boldsymbol{Y}^{-1}(\mathcal{Y})$ of subsets of $\mathcal{A}$ is a σ-field, denoted by $\sigma(\boldsymbol{Y})$ and called the σ-field *generated* by $\boldsymbol{Y}$.

Definition B.1. *For any topological space $\mathbb{Y}$, the* Borel σ-field *of Ω is the σ-field $\mathcal{B}^{\mathrm{o}}_{\mathbb{Y}}$ generated by the open sets of $\mathbb{Y}$. Elements of $\mathcal{B}^{\mathrm{o}}_{\mathbb{Y}}$ are called* Borel sets*. When we use measurable mappings with values in a topological space (as* *metric or Hilbert spaces), measurability implicitly refers to the Borel σ-field. A continuous mapping between topological spaces is measurable.*

B.1.2 Measure

Let $(\Omega, \mathcal{A})$ be a measurable space. A *measure* on $(\Omega, \mathcal{A})$ is a function $\mu\colon \mathcal{A} \to \mathbb{R} \cup \{+\infty\}$ with values in the extended real numbers, such that:

1. $\mu(B) \geq 0$ for $B \in \mathcal{A}$, with equality if $B = \emptyset$;
2. if the sequence $\{B_n\}_{n\in N}$ is such that $B_n \in \mathcal{A}$, for $n \in N$ where N is countable, and the B_n are mutually disjoints, then $\mu(\bigcup_{n\in N} B_n) = \sum_{n\in N} \mu(B_n)$.

The second property is called σ-additivity, or countable additivity. The triple $(\Omega, \mathcal{A}, \mu)$ is called a *measure space*.

A measure μ is said to be *finite* if $\mu(\Omega) < +\infty$. A measure μ is said to be σ-*finite* (or a measure μ is a σ-finite measure) if there exists a countable sequence $\{B_n\}_{n\in N}$ in $\mathcal{A}$ such that $\bigcup_{n\in N} B_n = \Omega$ and $\mu(B_n) < +\infty$ for all $n \in N$.

Say that a subset $C \subset \Omega$ is μ-*negligible* in the measure space $(\Omega, \mathcal{A}, \mu)$ if there exists $B \in \mathcal{A}$ such that $C \subset B$ and $\mu(B) = 0$. The measure space $(\Omega, \mathcal{A}, \mu)$ is called μ-*complete* [26, p. 2] if every μ-negligible subset is in $\mathcal{A}$. Given a measure space $(\Omega, \mathcal{A}, \mu)$, we define a new σ-field $\mathcal{A}^\mu$ which consists of all the sets $B \subset \Omega$ for which there exists $B_+, B_- \in \mathcal{A}$ such that $B_- \subset B \subset B_+$ and $\mu(B_+ - B_-) = 0$. The extension of μ on $\mathcal{A}^\mu$ is unique and the measure space $(\Omega, \mathcal{A}, \mu)$ is complete if $\mathcal{A}^\mu = \mathcal{A}$. On a measurable space $(\Omega, \mathcal{A})$ without explicit reference to a measure, it is possible to define a σ-field called the σ-field of universally measurable sets over $(\Omega, \mathcal{A})$ [21, Definition 7.18]. It is the σ-field $\widehat{\mathcal{A}} := \bigcap_\mu \mathcal{A}^\mu$ obtained when the intersection is over the finite measures μ on $\mathcal{A}$. A σ-field $\mathcal{A}$ is said to be *complete* or *universally complete* if $\mathcal{A} = \widehat{\mathcal{A}}$. As an example, suppose that μ is a σ-finite measure on $\mathcal{A}$ and $\mathcal{A}$ is μ-complete. Since $\mathcal{A} \subset \widehat{\mathcal{A}} = \bigcap_{\mu'} \mathcal{A}^{\mu'} \subset \mathcal{A}^\mu = \mathcal{A}$ we have that $\mathcal{A} = \widehat{\mathcal{A}}$, so that the σ-field $\mathcal{A}$ is universally complete.

B.1.3 Probability Space

If $\mu(\Omega) = 1$, then $(\Omega, \mathcal{A}, \mu)$ is called a *probability space*, and the measure μ is called a *probability measure*, generally denoted by $\mathbb{P}$ and called a *probability*. Elements of $\mathcal{A}$ are called *events*

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A condition holds *almost surely* on Ω if it holds on $\Omega \backslash N$, where N is a subset of Ω of measure 0, and abbreviated a.s. or $\mathbb{P}$-a.s..

Definition B.2. *We say that a probability space $(\Omega, \mathcal{A}, \mu)$ is* non-atomic*, or alternatively call μ non-atomic, if $\mu(A) > 0$ implies the existence of $B \in \mathcal{A}$, $B \subset A$ with $0 < \mu(B) < \mu(A)$.*

B.1.4 Product Probability Space

Let $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$, $i = 1, 2$ be two probability spaces. The *product probability space* is defined as $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$, where the product σ-field $\mathcal{A}_1 \otimes \mathcal{A}_2$ has been introduced at Remark 3.25 as the one generated by the rectangles $\{G_1 \times G_2 \mid G_i \in \mathcal{A}_i\ ,\ \ i = 1, 2\}$, and where the *product probability* $\mathbb{P}_1 \otimes \mathbb{P}_2$ is characterized on the rectangles by

$$\mathbb{P}_1 \otimes \mathbb{P}_2(G_1 \times G_2) = \mathbb{P}_1(G_1) \times \mathbb{P}_2(G_2)\ . \tag{B.1}$$

It can be shown that $\mathbb{P}_1 \otimes \mathbb{P}_2$ can be extended into a probability on the product σ-field $\mathcal{A}_1 \otimes \mathcal{A}_2$. By associativity, one can define the product of a finite number of probability spaces. The case of an infinite product of probability spaces is discussed at §B.7.1.

B.2 Random Variables

After having recalled the definition of a random variable, we turn to integration, with L^p-spaces and mathematical expectation. Recalls on probability image, Radon-Nikodym derivative and uniform integrability are also provided.

Definition B.3. *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space and* $(\mathbb{Y}, \mathcal{Y})$ *be a measurable space. A measurable mapping* $\boldsymbol{Y} : \Omega \to \mathbb{Y}$ *is called a* random variable.

Remark B.4. When the measurable mapping $\boldsymbol{Y}$ takes values in $(\mathbb{R}, \mathcal{B}^{o}_{\mathbb{R}})$ it is called a real-valued random variable. When it takes values in $(\mathbb{R}^n, \mathcal{B}^{o}_{\mathbb{R}^n})$ it is called a random element. ♢

Two random variables $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are said to be equal *almost surely*, or $\mathbb{P}$-*almost surely*, or $\mathbb{P}$-a.s., when $\mathbb{P}(\{\boldsymbol{Y} = \boldsymbol{Z}\}) = 1$. Recall that, in Probability Theory, it is customary to omit the variable ω and to write

$$\{\boldsymbol{Y} = \boldsymbol{Z}\} := \big\{\omega \in \Omega \;\big|\; \boldsymbol{Y}(\omega) = \boldsymbol{Z}(\omega)\big\}\ . \tag{B.2}$$

B.2.1 L^p-Spaces

Let $0 < p \leq +\infty$. The L^p-*norm* of a random variable $\boldsymbol{Y}$ with values in a Banach space is defined for $p < +\infty$ as

$$\left\|\boldsymbol{Y}\right\|_p := \left(\int_\Omega \left\|\boldsymbol{Y}\right\|^p \, \mathrm{d}\mathbb{P}\right)^{\frac{1}{p}}\ , \tag{B.3}$$

when the integral exists, and for $p = +\infty$ as

$$\left\|\boldsymbol{Y}\right\|_\infty = \operatorname{ess\,sup} \left\|\boldsymbol{Y}\right\| := \inf\big\{y \in \mathbb{R} \;\big|\; \mathbb{P}\big\{\omega \;\big|\; \left\|\boldsymbol{Y}(\omega)\right\| > y\big\} = 0\big\}\ . \tag{B.4}$$

The L^∞-norm $\|\boldsymbol{Y}\|_\infty$ of $\boldsymbol{Y}$ is called the *essential supremum* of $\|\boldsymbol{Y}\|$.

The set of random variables with finite L^p-norm forms a vector space V with the usual pointwise addition and scalar multiplication of functions. Two random variables are said to be equivalent when their difference has zero L^p-norm: the L^p*-space* on Ω is the quotient space of V by this equivalence relation. Thus, random variables in L^p-space are defined up to equivalence almost surely. We use the notation $L^p(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{Y})$ to specify the domain and images spaces, and the notation $L^p(\Omega, \mathcal{A}, \mathbb{P})$ for real-valued random variables. For $1 \leq p \leq +\infty$ the space $L^p(\Omega, \mathcal{A}, \mathbb{P})$ is a Banach space.

B.2.2 Mathematical Expectation

A real-valued random variable $\boldsymbol{Y}$ is said to be *integrable* when $\boldsymbol{Y} \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ or, equivalently, when $\int_\Omega |\boldsymbol{Y}| \, \mathrm{d}\mathbb{P} < +\infty$. The mathematical *expectation* of $\boldsymbol{Y}$ is

$$\mathbb{E}(\boldsymbol{Y}) := \int_\Omega \boldsymbol{Y} \, \mathrm{d}\mathbb{P} \, . \tag{B.5}$$

With this notation, $\boldsymbol{Y} \in L^1(\Omega, \mathcal{A}, \mathbb{P}) \iff \mathbb{E}(|\boldsymbol{Y}|) < +\infty$.

When the dependence w.r.t. the probability $\mathbb{P}$ has to be stressed, one uses the notation $\mathbb{E}_\mathbb{P}(\boldsymbol{Y})$:

$$\mathbb{E}_\mathbb{P}(\boldsymbol{Y}) := \int_\Omega \boldsymbol{Y} \, \mathrm{d}\mathbb{P} \, . \tag{B.6}$$

The space $L^2(\Omega, \mathcal{A}, \mathbb{P})$ is a Hilbert space, equipped with the scalar product

$$\langle \boldsymbol{Y} \, , \boldsymbol{Z} \rangle := \int_\Omega \boldsymbol{Y}(\omega) \boldsymbol{Z}(\omega) \, \mathrm{d}\mathbb{P}(\omega) \, . \tag{B.7}$$

Random variables in $L^2(\Omega, \mathcal{A}, \mathbb{P})$ are said to be *square integrable*.

B.2.3 Probability Image

Let $\boldsymbol{Y}$ be a random variable. The image measure

$$\mathbb{P}_{\boldsymbol{Y}} := \mathbb{P} \circ \boldsymbol{Y}^{-1} \tag{B.8}$$

is a probability on $(\mathbb{Y}, \mathcal{Y})$, called the *probability law of* $\boldsymbol{Y}$ or *probability distribution of* $\boldsymbol{Y}$. It is also denoted by

$$\boldsymbol{Y}_\star(\mathbb{P}) := \mathbb{P} \circ \boldsymbol{Y}^{-1} \, . \tag{B.9}$$

For any measurable bounded real-valued function φ, the mapping $\varphi(\boldsymbol{Y}) : \Omega \to \mathbb{R}$ is a random variable, and one has that

$$\mathbb{E}_\mathbb{P}(\varphi(\boldsymbol{Y})) = \mathbb{E}_{\mathbb{P}_{\boldsymbol{Y}}}(\varphi) \, . \tag{B.10}$$

B.2.4 Radon-Nikodym Derivative

Let $\mathbb{P}$ and $\mathbb{Q}$ be two probabilities on $(\Omega, \mathcal{A})$. The probability $\mathbb{Q}$ is said to *have a density* w.r.t. $\mathbb{P}$ if there exists a nonnegative ($\boldsymbol{R} \geq 0$, $\mathbb{P}$-a.s.) integrable random variable $\boldsymbol{R} \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ such that

$$\mathbb{E}_{\mathbb{Q}}(\boldsymbol{Z}) = \mathbb{E}_{\mathbb{P}}(\boldsymbol{R}\boldsymbol{Z}) \ , \ \ \forall \boldsymbol{Z} \in L^1(\Omega, \mathcal{A}, \mathbb{Q}) \ . \tag{B.11}$$

The random variable $\boldsymbol{R}$ is uniquely defined $\mathbb{P}$-a.s., is called a *density* , and is denoted by $\boldsymbol{R} = \mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$.

The probabilities $\mathbb{P}$ and $\mathbb{Q}$ are said to be *equivalent* if $\mathbb{Q}$ is absolutely continuous w.r.t. $\mathbb{P}$ and $\mathbb{P}$ is absolutely continuous w.r.t. $\mathbb{Q}$. This is denoted by $\mathbb{P} \sim \mathbb{Q}$. In that case, $\boldsymbol{R} = \mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$ is uniquely defined $\mathbb{P}$-a.s. and $\mathbb{Q}$-a.s., and we have that $\boldsymbol{R} > 0$ and $\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q} = 1/\boldsymbol{R}$:

$$\mathbb{P} \sim \mathbb{Q} \text{ and } \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \boldsymbol{R} \in L^1(\Omega, \mathcal{A}, \mathbb{P}) \ , \ \ \boldsymbol{R} > 0 \ , \ \ \mathbb{P}\text{-a.s. or } \mathbb{Q}\text{-a.s.} \ . \tag{B.12}$$

B.2.5 Uniform Integrability

Consider $\{\boldsymbol{Y}_i\}_{i \in I}$ a collection of random variables with values in a Banach space. The collection $\{\boldsymbol{Y}_i\}_{i \in I}$ is said to be *uniformly continuous* if

$$\forall \epsilon > 0 \ , \ \ \exists \alpha > 0 \ \text{ such that } \ \mathbb{P}(A) \leq \alpha \Rightarrow \sup_{i \in I} \int_A \left\| \boldsymbol{Y}_i \right\| \, \mathrm{d}\mathbb{P} \leq \epsilon \ ,$$

and *uniformly integrable* if

$$\forall \epsilon > 0 \ , \ \ \exists \alpha > 0 \ \text{ such that } \ \sup_{i \in I} \int_{\left\| \boldsymbol{Y}_i \right\| > \alpha} \left\| \boldsymbol{Y}_i \right\| \, \mathrm{d}\mathbb{P} \leq \epsilon \ . \tag{B.13}$$

B.3 Convergence of Random Variables

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We present different notions of convergence of random variables.

B.3.1 Almost Sure Convergence

Let $\boldsymbol{Y}$ be a random variable and $\{\boldsymbol{Y}_n\}_{n \in \mathbb{N}}$ be a sequence of random variables with values in the same topological space. We say that $\{\boldsymbol{Y}_n\}_{n \in \mathbb{N}}$ converges *almost surely* towards $\boldsymbol{Y}$, denoted by $\boldsymbol{Y}_n \xrightarrow{\text{a.s.}} \boldsymbol{Y}$, if

$$\mathbb{P}\big(\boldsymbol{Y}_n \underset{n \to +\infty}{\longrightarrow} \boldsymbol{Y}\big) = 1 \ .$$

This is denoted by $\boldsymbol{Y}_n \xrightarrow{\text{a.s.}} \boldsymbol{Y}$.

B.3.2 Convergence in L^p Norm

Let $1 \leq p \leq +\infty$. Let $\boldsymbol{Y}$ be a random variable and $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ be a sequence of random variables with values in the same Banach space. The sequence $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ *converges in L^p norm* towards $\boldsymbol{Y}$ if

$$\left\|\boldsymbol{Y}_n - \boldsymbol{Y}\right\|_p \underset{n\to+\infty}{\longrightarrow} 0 .$$

This is denoted by $\boldsymbol{Y}_n \xrightarrow{L^p} \boldsymbol{Y}$. The L^2 convergence is called *mean square convergence.*

B.3.3 Convergence in Probability

Let $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ be a sequence of random variables, taking values in a metric space $(\mathbb{Y}, d)$. The sequence $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ *converges in probability* towards a random variable $\boldsymbol{Y}$ if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(d(\boldsymbol{Y}_n, \boldsymbol{Y}) \geq \varepsilon) = 0 .$$

This is denoted by $\boldsymbol{Y}_n \xrightarrow{\mathbb{P}} \boldsymbol{Y}$.

B.3.4 Convergence in Law

Let $\boldsymbol{Y}$ be a random variable and $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ be a sequence of random variables with values in the same topological space. The sequence $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ *converges in law* or *converges in distribution* towards a random variable $\boldsymbol{Y}$ if the sequence $\{\mathbb{P}_{\boldsymbol{Y}_n}\}_{n\in\mathbb{N}}$ of image probabilities (see §B.8) narrowly converges towards $\mathbb{P}_{\boldsymbol{Y}}$, that is, if

$$\lim_{n\to+\infty} \mathbb{E}(\varphi(\boldsymbol{Y}_n)) = \mathbb{E}(\varphi(\boldsymbol{Y})) , \tag{B.14}$$

for all bounded continuous function φ. This is denoted by $\boldsymbol{Y}_n \xrightarrow{\mathcal{D}} \boldsymbol{Y}$.

B.3.5 Relations Between Convergences

We have the following properties.

- Convergence almost surely implies convergence in probability.
- If a sequence converges in probability, there exists a sub-sequence which converges almost surely.
- If a sequence of random variables converges in L^2 norm, the sequence converges in probability.
- If the sequence $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ converges in probability to $\boldsymbol{Y}$ then $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ converges in law to $\boldsymbol{Y}$.
- When $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ converges in law to a constant random variable $\boldsymbol{Y}$, then $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ converges in probability to the constant $\boldsymbol{Y}$.

B.4 Conditional Expectation

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. In what follows, $\mathcal{G}$ denotes a *subfield* of $\mathcal{A}$, that is, $\mathcal{G} \subset \mathcal{A}$ and $\mathcal{G}$ is a σ-field. In this section, when not specified, a random variable is a real-valued random variable.

Let $L^2(\Omega, \mathcal{G}, \mathbb{P})$ be the closed vector subspace of square integrable $\mathcal{G}$-measurable functions in the Hilbert space $L^2(\Omega, \mathcal{A}, \mathbb{P})$. We can thus define the orthogonal projection on $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Definition B.5. *If $\boldsymbol{Y}$ is a square integrable random variable, we define the* conditional expectation of $\boldsymbol{Y}$ knowing (the σ-field) $\mathcal{G}$, *and we denote by* $\mathbb{E}\big(\boldsymbol{Y} \mid \mathcal{G}\big)$, *the orthogonal projection of $\boldsymbol{Y}$ on $L^2(\Omega, \mathcal{G}, \mathbb{P})$:*

$$\boldsymbol{Z} = \mathbb{E}\big(\boldsymbol{Y} \mid \mathcal{G}\big) \iff \boldsymbol{Z} = \underset{\boldsymbol{T} \in L^2(\Omega, \mathcal{G}, \mathbb{P})}{\arg\min} \ \mathbb{E}\Big(\big\|\boldsymbol{Y} - \boldsymbol{T}\big\|^2\Big) . \tag{B.15}$$

Thus, the conditional expectation solves an optimization problem under measurability constraints (among $\mathcal{G}$-measurable random variables). This was discussed in §3.5 in the case of a finite probability space.

The conditional expectation may be extended to L^1 random variables. If $\boldsymbol{Y}$ is an integrable random variable, $\mathbb{E}\big(\boldsymbol{Y} \mid \mathcal{G}\big)$ is the unique $\boldsymbol{Z} \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ such that

$$\mathbb{E}\big(\boldsymbol{Y}\boldsymbol{T}\big) = \mathbb{E}\big(\boldsymbol{Z}\boldsymbol{T}\big) \ , \ \ \forall \boldsymbol{T} \in L^\infty(\Omega, \mathcal{G}, \mathbb{P}) . \tag{B.16}$$

In fact, the above result holds true under the weaker assumptions that $\boldsymbol{Y}$ is bounded either below or above by an integrable random variable.

The conditional expectation may be extended componentwise to L^1 random variables with values in $\mathbb{R}^d$.

Elementary Properties

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two integrable random variables, λ a real number. Then

$$\mathbb{E}\big(\lambda \boldsymbol{X} + \boldsymbol{Y} \mid \mathcal{G}\big) = \lambda \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{G}\big) + \mathbb{E}\big(\boldsymbol{Y} \mid \mathcal{G}\big) \ , \tag{B.17}$$

$$\boldsymbol{X} \geq 0 \Rightarrow \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{G}\big) \geq 0 \, , \tag{B.18}$$

$$\mathbb{E}\big(\mathbb{E}\big(\boldsymbol{X} \mid \mathcal{G}\big)\big) = \mathbb{E}\big(\boldsymbol{X}\big) \ , \tag{B.19}$$

$$\boldsymbol{Y} \in L^\infty(\Omega, \mathcal{G}, \mathbb{P}) \Rightarrow \mathbb{E}\big(\boldsymbol{Y}\boldsymbol{X} \mid \mathcal{G}\big) = \boldsymbol{Y}\mathbb{E}\big(\boldsymbol{X} \mid \mathcal{G}\big) \ . \tag{B.20}$$

If $\mathcal{G}^\flat$, $\mathcal{G}^\sharp$ are two subfields of $\mathcal{A}$, we have that

$$\mathcal{G}^\flat \subset \mathcal{G}^\sharp \Rightarrow \mathbb{E}\Big(\mathbb{E}\big(\boldsymbol{X} \mid \mathcal{G}^\sharp\big) \ \Big| \ \mathcal{G}^\flat\Big) = \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{G}^\flat\big) . \tag{B.21}$$

Dependence Upon the Probability Law

Proposition B.6. *Let $\mathbb{P}$ and $\mathbb{Q}$ be two equivalent probabilities on $(\Omega, \mathcal{A})$, with positive Radon-Nikodym derivative $\boldsymbol{R}$ (see §B.2.4). We have that*

$$\mathbb{E}_{\mathbb{Q}}\left(\boldsymbol{X} \mid \mathcal{G}\right) = \frac{\mathbb{E}_{\mathbb{P}}\left(\boldsymbol{R}\boldsymbol{X} \mid \mathcal{G}\right)}{\mathbb{E}_{\mathbb{P}}\left(\boldsymbol{R} \mid \mathcal{G}\right)} \;, \quad \mathbb{P}\text{-a.s. or } \mathbb{Q}\text{-a.s.} \;, \tag{B.22}$$

for all bounded random variable $\boldsymbol{X}$.

Proof. Let $\boldsymbol{X} \in L^\infty(\Omega, \mathcal{A}, \mathbb{P})$. We first note that the right hand side of Equation (B.22) is well defined because $\boldsymbol{X}$ is bounded and $\boldsymbol{R} \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. For any $G \in \mathcal{G}$, we have that

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}\left(\boldsymbol{X}\mathbf{1}_G\right) &= \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{R}\boldsymbol{X}\mathbf{1}_G\right) \quad \text{by (B.11)} \\
&= \mathbb{E}_{\mathbb{P}}\left(\mathbb{E}_{\mathbb{P}}\left(\boldsymbol{R}\boldsymbol{X} \mid \mathcal{G}\right)\mathbf{1}_G\right) \quad \text{by (B.16)} \\
&= \mathbb{E}_{\mathbb{P}}\left(\mathbb{E}_{\mathbb{P}}\left(\boldsymbol{R} \mid \mathcal{G}\right)\frac{\mathbb{E}_{\mathbb{P}}\left(\boldsymbol{R}\boldsymbol{X} \mid \mathcal{G}\right)}{\mathbb{E}_{\mathbb{P}}\left(\boldsymbol{R} \mid \mathcal{G}\right)}\mathbf{1}_G\right) \quad \text{since } \boldsymbol{R} > 0 \\
&= \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{R}\frac{\mathbb{E}_{\mathbb{P}}\left(\boldsymbol{R}\boldsymbol{X} \mid \mathcal{G}\right)}{\mathbb{E}_{\mathbb{P}}\left(\boldsymbol{R} \mid \mathcal{G}\right)}\mathbf{1}_G\right) \quad \text{by (B.16)} \\
&= \mathbb{E}_{\mathbb{Q}}\left(\frac{\mathbb{E}_{\mathbb{P}}\left(\boldsymbol{R}\boldsymbol{X} \mid \mathcal{G}\right)}{\mathbb{E}_{\mathbb{P}}\left(\boldsymbol{R} \mid \mathcal{G}\right)}\mathbf{1}_G\right) \quad \text{by (B.11)} \;.
\end{aligned}$$

The proof is complete by Equation (B.16). □

Remark B.7. If $\mathrm{d}\mathbb{Q}/\,\mathrm{d}\mathbb{P}$ is $\mathcal{G}$-measurable, then the conditional expectation operators $\mathbb{E}_{\mathbb{Q}}\left(\cdot \mid \mathcal{G}\right)$ and $\mathbb{E}_{\mathbb{P}}\left(\cdot \mid \mathcal{G}\right)$ coincide ($\mathbb{Q}$-a.s. or $\mathbb{P}$-a.s.). In other words, the conditional expectation operator $\mathbb{E}_{\mathbb{P}}\left(\cdot \mid \mathcal{G}\right)$ depends upon $\mathcal{G}$ and the equivalence class of $\mathbb{P}$ for the relation $\mathbb{P} \sim_{\mathcal{G}} \mathbb{Q} \iff \mathbb{P} \sim \mathbb{Q}$ and $\mathrm{d}\mathbb{Q}/\,\mathrm{d}\mathbb{P}$ is $\mathcal{G}$-measurable. ♢

Proposition B.8. *Let $\Phi : (\mathbb{X}, \mathcal{X}) \to (\Omega, \mathcal{A})$ be measurable, and let $\mathbb{Q}$ be a probability on $(\mathbb{X}, \mathcal{X})$. Let $\Phi_\star(\mathbb{Q})$ the probability on $(\Omega, \mathcal{A})$, image of $\mathbb{Q}$ by Φ (see §B.8). For any random variable $\boldsymbol{Y}$ on Ω such that $\mathbb{E}_{\Phi_\star(\mathbb{Q})}\left(\left|\boldsymbol{Y}\right|\right) < +\infty$, we have that*

$$\mathbb{E}_{\Phi_\star(\mathbb{Q})}\left(\boldsymbol{Y} \mid \mathcal{G}\right) \circ \Phi = \mathbb{E}_{\mathbb{Q}}\left(\boldsymbol{Y} \circ \Phi \mid \Phi^{-1}(\mathcal{G})\right) \;, \quad \mathbb{P}\text{-a.s.} \;. \tag{B.23}$$

Proof. For any $G \in \mathcal{G}$, we have that

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}\left(\mathbf{1}_{\Phi^{-1}(G)} \times (\boldsymbol{Y} \circ \Phi)\right) &= \mathbb{E}_{\mathbb{Q}}\left((\mathbf{1}_G \circ \Phi) \times (\boldsymbol{Y} \circ \Phi)\right) \\
&= \mathbb{E}_{\Phi_\star(\mathbb{Q})}\left(\mathbf{1}_G \times \boldsymbol{Y}\right) \quad \text{by (B.10)} \\
&= \mathbb{E}_{\Phi_\star(\mathbb{Q})}\left(\mathbf{1}_G \times \mathbb{E}_{\Phi_\star(\mathbb{Q})}\left(\boldsymbol{Y} \mid \mathcal{G}\right)\right) \quad \text{by (B.16)} \\
&= \mathbb{E}_{\mathbb{Q}}\left(\mathbf{1}_G \circ \Phi \times \mathbb{E}_{\Phi_\star(\mathbb{Q})}\left(\boldsymbol{Y} \mid \mathcal{G}\right) \circ \Phi\right) \quad \text{by (B.10)} \\
&= \mathbb{E}_{\mathbb{Q}}\left(\mathbf{1}_{\Phi^{-1}(G)} \times \mathbb{E}_{\Phi_\star(\mathbb{Q})}\left(\boldsymbol{Y} \mid \mathcal{G}\right) \circ \Phi\right) \;.
\end{aligned}$$

We conclude that (B.23) holds true since $\Phi^{-1}(\mathcal{G})$ is equal to $\{\Phi^{-1}(G) \mid G \in \mathcal{G}\}$ and since the function $\mathbb{E}_{\Phi_\star(\mathbb{Q})}\left(\boldsymbol{Y} \mid \mathcal{G}\right) \circ \Phi$ is $\Phi^{-1}(\mathcal{G})$-measurable and is in $L^1(\mathbb{X}, \mathcal{X}, \mathbb{Q})$. □

Conditional Expectation w.r.t. an Atomic σ-Field

A subfield $\mathcal{G}$ of $\mathcal{A}$ is called *atomic* (see Definition 3.26) if it is generated by a countable partition $\{\Omega_n\}_{n\in N}$: $\mathcal{G} = \sigma(\Omega_n, n \in N)$, where N is countable (in bijection with a subset of $\mathbb{N}$).

Proposition B.9. *If $\mathcal{G}$ is an atomic σ-field, generated by the countable partition $\{\Omega_n\}_{n\in N}$, and if $\boldsymbol{Y}$ is an integrable random variable, then*

$$\mathbb{E}\left(\boldsymbol{Y} \mid \mathcal{G}\right) = \sum_{n, \mathbb{P}(\Omega_n)>0} \frac{\mathbb{E}\left(\mathbf{1}_{\Omega_n}\boldsymbol{Y}\right)}{\mathbb{P}(\Omega_n)} \mathbf{1}_{\Omega_n} . \tag{B.24}$$

Conditional Expectation Knowing a Random Variable

Definition B.10. *If $\boldsymbol{Y}$ is an integrable random variable and $\boldsymbol{Z}$ is a random variable, we define* the conditional expectation of $\boldsymbol{Y}$ knowing (the random variable) $\boldsymbol{Z}$ *as the random variable*

$$\mathbb{E}\left(\boldsymbol{Y} \mid \boldsymbol{Z}\right) := \mathbb{E}\left(\boldsymbol{Y} \mid \sigma(\boldsymbol{Z})\right) , \tag{B.25}$$

where $\sigma(\boldsymbol{Z})$ is the σ-field generated by the random variable $\boldsymbol{Z}$.

The following proposition results from Proposition 3.46.

Proposition B.11. *Let $\boldsymbol{Y}$ be an integrable random variable. For any random variable $\boldsymbol{Z}$, there exists a unique measurable function Ψ (unique $\mathbb{P}_{\boldsymbol{Z}}$-a.s.) such that $\mathbb{E}\left(\boldsymbol{Y} \mid \boldsymbol{Z}\right) = \Psi(\boldsymbol{Z})$. Therefore, we can define*

$$\mathbb{E}\left(\boldsymbol{Y} \mid \boldsymbol{Z} = z\right) := \Psi(z) , \quad \forall z \text{ such that } \mathbb{P}(Z = z) > 0 . \tag{B.26}$$

Notice that Ψ depends functionally upon the random variable $(\boldsymbol{Y}, \boldsymbol{Z})$, hence, in particular, upon $\boldsymbol{Z}$. To insist upon this dependence, we rephrase the above result as follows. For any random variable $\boldsymbol{Z}$, there exists a unique measurable function $\Psi_{[\boldsymbol{Y},\boldsymbol{Z}]}$ (unique $\mathbb{P}_{\boldsymbol{Z}}$-a.s.) and $\Omega' \subset \Omega$ such that $\mathbb{P}(\Omega') = 1$ and

$$\mathbb{E}\left(\boldsymbol{Y} \mid \boldsymbol{Z}\right)(\omega) = \Psi_{[\boldsymbol{Y},\boldsymbol{Z}]}\big(\boldsymbol{Z}(\omega)\big) , \quad \forall \omega \in \Omega' . \tag{B.27}$$

If $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are discrete random variables, by (B.24), we have that

$$\Psi_{[\boldsymbol{Y},\boldsymbol{Z}]}(z) = \sum_{y \in \boldsymbol{Y}(\Omega), z' \in \boldsymbol{Z}(\Omega)} y \frac{\mathbb{P}(\boldsymbol{Y} = y, \boldsymbol{Z} = z')}{\mathbb{P}(\boldsymbol{Z} = z')} \mathbf{1}_{\{z'\}}(z) , \tag{B.28}$$

which indeed depends functionally upon $\boldsymbol{Y}$ and $\boldsymbol{Z}$, by the terms $\mathbb{P}(\boldsymbol{Y} = y, \boldsymbol{Z} = z')$ and $\mathbb{P}(\boldsymbol{Z} = z')$.

B.5 Conditional Probability

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. In what follows, $\mathcal{G}$ denotes a subfield of $\mathcal{A}$.

B.5.1 Conditional Probability w.r.t. an Event

Let $B \in \mathcal{A}$ such that $\mathbb{P}(B) > 0$, and $F \in \mathcal{A}$. We define the *conditional probability of (the event) F knowing (the event)B* as the real number

$$\mathbb{P}(F \mid B) := \frac{\mathbb{P}(F \cap B)}{\mathbb{P}(B)} . \tag{B.29}$$

Let $B \in \mathcal{A}$ such that $\mathbb{P}(B) > 0$. The *conditional probability* $\mathbb{P}_{|B} : \mathcal{A} \to [0,1]$ defined by

$$\mathbb{P}_{|B}(F) := \mathbb{P}(F \mid B) , \ \forall F \in \mathcal{A} \tag{B.30}$$

is a probability over $(\Omega, \mathcal{A})$. The probability $\mathbb{P}_{|B}$ has the density $\mathbf{1}_B / \mathbb{P}(B)$ w.r.t. $\mathbb{P}$:

$$\frac{\mathrm{d}\mathbb{P}_{|B}}{\mathrm{d}\mathbb{P}} = \frac{\mathbf{1}_B}{\mathbb{P}(B)} . \tag{B.31}$$

B.5.2 Conditional Expectation w.r.t. an Event

Let $B \in \mathcal{A}$ with $\mathbb{P}(B) > 0$. As $(\Omega, \mathcal{A}, \mathbb{P}_{|B})$ is a probability space, we can define the expectation under this probability, denoted $\mathbb{E}(\cdot \mid B)$. Let $\boldsymbol{Y}$ be an integrable random variable. Then *the conditional expectation of $\boldsymbol{Y}$ knowing the event B* is the real number

$$\mathbb{E}\big(\boldsymbol{Y} \mid B\big) := \mathbb{E}_{\mathbb{P}_{|B}}(\boldsymbol{Y}) . \tag{B.32}$$

For any integrable random variable $\boldsymbol{Y}$, we have that

$$\mathbb{E}\big(\boldsymbol{Y} \mid B\big) = \frac{\mathbb{E}\big(\boldsymbol{Y} \mathbf{1}_B\big)}{\mathbb{P}(B)} . \tag{B.33}$$

Notice that, with the above notation, Equation (B.24) can be written as:

$$\mathbb{E}\big(\boldsymbol{Y} \mid \sigma(\{\Omega_n\}_{n \in N})\big) = \sum_{n \in N, \mathbb{P}(\Omega_n) > 0} \mathbb{E}\big(\boldsymbol{Y} \mid \Omega_n\big) \mathbf{1}_{\Omega_n} . \tag{B.34}$$

B.5.3 Conditional Probability w.r.t. a $\boldsymbol{\sigma}$-field

Let $B \in \mathcal{A}$. The *conditional probability of (the event) B knowing (the subfield) $\mathcal{G}$* is the *random variable*:

$$\mathbb{P}(B \mid \mathcal{G}) := \mathbb{E}\left(\mathbf{1}_B \mid \mathcal{G}\right) . \tag{B.35}$$

If $\mathcal{G}$ is an atomic σ-field, generated by the countable partition $\{\Omega_n\}_{n \in N}$, we obtain using (B.24):

$$\mathbb{P}(B \mid \sigma(\{\Omega_n\}_{n \in N})) = \sum_{n \in N, \mathbb{P}(\Omega_n) > 0} \mathbb{P}(B \mid \Omega_n) \mathbf{1}_{\Omega_n} . \tag{B.36}$$

If $\boldsymbol{Z}$ is a random variable, we define *the conditional probability of the event B knowing (the random variable) $\boldsymbol{Z}$* as the random variable

$$\mathbb{P}(B \mid \boldsymbol{Z}) := \mathbb{E}\left(\mathbf{1}_B \mid \sigma(\boldsymbol{Z})\right) , \tag{B.37}$$

where $\sigma(\boldsymbol{Z})$ is the σ-field generated by the random variable $\boldsymbol{Z}$.

B.6 Stochastic Kernels

The main objective of this section is to formulate Proposition B.22, which states the following, widely used property. When computing a conditional expectation with respect to a σ-field $\mathcal{G}$, all the $\mathcal{G}$-measurable variables can be "frozen" during the conditional expectation evaluation. For this purpose, we review results on Borel spaces and on regular conditional laws.

B.6.1 Borel Spaces

Here we follow [21, chap. 7].

Definition B.12. *Let $\mathbb{X}$ be a topological space. We denote by $\mathcal{B}^{o}_{\mathbb{X}}$ the σ-field generated by the open subsets of $\mathbb{X}$. The elements of $\mathcal{B}^{o}_{\mathbb{X}}$ are called the* Borel subsets *of $\mathbb{X}$. A mapping φ between topological spaces $\mathbb{X}$ and $\mathbb{X}'$ is said to be* Borel-measurable *if $\varphi^{-1}(\mathcal{B}^{o}_{\mathbb{X}'}) \subset \mathcal{B}^{o}_{\mathbb{X}}$.*

Definition B.13. *A topological space $\mathbb{X}$ is a* Borel space *if there exists a separable[1] complete metric space $\mathbb{X}'$ as well as a Borel subset $B \in \mathcal{B}^{o}_{\mathbb{X}'}$ such that $\mathbb{X}$ is homeomorphic[2] to B. A* Borel isomorphism *φ between Borel spaces $\mathbb{X}$ and $\mathbb{X}'$ is a one-to-one Borel-measurable mapping such that φ^{-1} is Borel-measurable on $\varphi(\mathbb{X})$.*

The spaces $\mathbb{R}^n$, as well as their Borel subsets, are Borel spaces. Every Borel space is metrizable and separable, and any complete separable metric space is a Borel space. Every uncountable Borel space is Borel isomorphic to $[0,1]$, metrizable and separable. If $\mathbb{X}$ is a Borel space, then the space $\mathcal{P}(\mathbb{X})$ of probability distributions over $\mathbb{X}$ is a Borel space.

---

[1] A metrizable topological space is *separable* if it contains a countable dense set.

[2] A *homeomorphism* φ between topological spaces $(\mathbb{X}, \mathcal{T})$ and $(\mathbb{X}', \mathcal{T}')$ is one-to-one and continuous, and φ^{-1} is continuous on $\varphi(\mathbb{X})$ with the relative topology.

B.6.2 Stochastic Kernels

Stochastic Kernels and Parametric Disintegration

Definition B.14 ([89]). *Let* $(\mathbb{X}, \mathcal{X})$ *and* $(\mathbb{Y}, \mathcal{Y})$ *be two measurable spaces. A* stochastic kernel *from* $(\mathbb{X}, \mathcal{X})$ *to* $(\mathbb{Y}, \mathcal{Y})$ *is a mapping* $p : \mathbb{X} \times \mathcal{Y} \to [0, 1]$ *such that*

- *for any* $F \in \mathcal{Y}$, $p(\cdot, F)$ *is* $\mathcal{X}$*-measurable;*
- *for any* $x \in \mathbb{X}$, $p(x, \cdot)$ *is a probability on* $\mathcal{Y}$.

A random measure *is a stochastic kernel from* $(\Omega, \mathcal{A})$ *to* $(\Omega, \mathcal{A})$.

A stochastic kernel may equivalently be seen as a measurable mapping from $(\mathbb{X}, \mathcal{X})$ to $\mathcal{P}(\mathbb{Y})$.Thus, as for notation and terminology, we shall speak of a *stochastic kernel* $p(x, \mathrm{d}y)$ *from* $\mathbb{X}$ *to* $\mathbb{Y}$ or of a *stochastic kernel* $p(\mathrm{d}y \mid x)$ *on* $\mathbb{Y}$ *given* $\mathbb{X}$.

Here is a composition operation on stochastic kernels.

Definition B.15 ([89]). *Let* $(\mathbb{X}, \mathcal{X})$, $(\mathbb{Y}, \mathcal{Y})$ *and* $(\mathbb{Z}, \mathcal{Z})$ *be three measurable spaces. Consider two stochastic kernels,* $p(\mathrm{d}y \mid x)$ *on* $\mathbb{Y}$ *given* $\mathbb{X}$ *and* $q(\mathrm{d}z \mid y)$ *on* $\mathbb{Z}$ *given* $\mathbb{Y}$. *Then, the following expression defines a stochastic kernel* $p \otimes q$ *on* $\mathbb{Z}$ *given* $\mathbb{X}$*:*

$$(p \otimes q)(F \mid x) := \int_{\mathbb{Y}} p(\mathrm{d}y \mid x) \int_F q(\mathrm{d}z \mid y) , \quad \forall F \in \mathcal{Z} . \tag{B.38}$$

The following proposition establishes that one can decompose a probability measure on a product $\mathbb{Y} \times \mathbb{Z}$ of Borel spaces as a marginal on $\mathbb{Y}$ and a stochastic kernel on $\mathbb{Z}$ given $\mathbb{Y}$. Moreover, this property remains valid when a measurable dependence w.r.t. a parameter is admitted.

Proposition B.16 (Parametric disintegration [21]). *Let* $(\mathbb{X}, \mathcal{X})$ *be a measurable space,* $\mathbb{Y}$ *and* $\mathbb{Z}$ *be Borel spaces and* $q(\mathrm{d}y\, \mathrm{d}z \mid x)$ *be a stochastic kernel on* $\mathbb{Y} \times \mathbb{Z}$ *given* $\mathbb{X}$. *Then, there exists a stochastic kernel* $r(\mathrm{d}z \mid x, y)$ *on* $\mathbb{Z}$ *given* $\mathbb{X} \times \mathbb{Y}$ *and a stochastic kernel* $s(\mathrm{d}y \mid x)$ *on* $\mathbb{Y}$ *given* $\mathbb{X}$ *such that* $q = r \otimes s$*, i.e.:*

$$q(\mathrm{d}y\, \mathrm{d}z \mid x) = r(\mathrm{d}z \mid x, y) s(\mathrm{d}y \mid x) . \tag{B.39}$$

The stochastic kernel $s(\mathrm{d}y \mid x)$ *is given by*

$$\int_F s(\mathrm{d}y \mid x) = \int_{F \times \mathbb{Z}} q(\mathrm{d}y\, \mathrm{d}z \mid x) , \quad \forall F \in \mathcal{Z} . \tag{B.40}$$

Corollary B.17. *Let* $\mathbb{X}$, $\mathbb{Y}$ *and* $\mathbb{Z}$ *be Borel spaces and* $q(\mathrm{d}y\, \mathrm{d}z \mid x)$ *be a stochastic kernel over* $\mathbb{Y} \times \mathbb{Z}$ *knowing* $\mathbb{X}$. *Then, there exists a stochastic kernel* $r(\mathrm{d}z \mid x, y)$ *over* $\mathbb{Z}$ *knowing* $\mathbb{X} \times \mathbb{Y}$ *and a stochastic kernel* $s(\mathrm{d}y \mid x)$ *over* $\mathbb{Y}$ *knowing* $\mathbb{X}$ *such that* $q = r \otimes s$ *as in* (B.39).

Corollary B.18 ([21]). *Let $\mathbb{Y}$ and $\mathbb{Z}$ be Borel spaces and $q \in \mathcal{P}(\mathbb{Y} \times \mathbb{Z})$. Then, there exists a stochastic kernel $r(\,\mathrm{d}z \mid y)$ over $\mathbb{Z}$ knowing $\mathbb{Y}$ such that $q = r \otimes s$:*

$$q(\,\mathrm{d}y\ \mathrm{d}z) = r(\,\mathrm{d}z \mid y)s(\,\mathrm{d}y) \ \textit{where}\ s(\,\mathrm{d}y) = \int_{\mathbb{Z}} q(\,\mathrm{d}y\ \mathrm{d}z)\,. \tag{B.41}$$

Regular Conditional Laws and Disintegration

Definition B.19 ([30, 89]). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $\boldsymbol{X}$ be a random variable taking values in a measurable space $(\mathbb{X}, \mathcal{X})$.*

1. *Let $\mathcal{G}$ be a subfield of $\mathcal{A}$. A* regular conditional law of the random variable $\boldsymbol{X}$ knowing $\mathcal{G}$ *is a stochastic kernel p from $(\Omega, \mathcal{A})$ to $(\mathbb{X}, \mathcal{X})$, such that, for any $F \in \mathcal{X}$, $p(\cdot, F)$ is a version*[3] *of $\mathbb{P}(\boldsymbol{X} \in F \mid \mathcal{G})$.*
2. *Let $\boldsymbol{Y} : (\Omega, \mathcal{A}) \to (\mathbb{Y}, \mathcal{Y})$ be a random variable taking values in a measurable space $(\mathbb{Y}, \mathcal{Y})$. A* regular conditional law of the random variable $\boldsymbol{X}$ knowing the random variable $\boldsymbol{Y}$ *is a stochastic kernel p from $(\mathbb{Y}, \mathcal{Y})$ to $(\mathbb{X}, \mathcal{X})$, such that, for any $F \in \mathcal{X}$, $p(\boldsymbol{Y}(\cdot), F)$ is a version of $\mathbb{P}(\boldsymbol{X} \in F \mid \boldsymbol{Y})$.*

When $(\mathbb{X}, \mathcal{X}) = (\Omega, \mathcal{A})$ and $\boldsymbol{X} = \mathrm{I}_{\Omega} : (\Omega, \mathcal{A}) \to (\Omega, \mathcal{A})$ is the identity mapping, we obtain the following particular cases. A regular conditional law of $\mathbb{P}$ knowing $\mathcal{G}$ *is a random measure P such that, for any $F \in \mathcal{A}$, $P(\cdot, F)$ is a version of $\mathbb{P}(F \mid \mathcal{G})$. A* regular conditional law of $\mathbb{P}$ knowing $\boldsymbol{Y}$ *is a stochastic kernel p such that, for any $F \in \mathcal{A}$, $p(\boldsymbol{Y}(\cdot), F)$ is a version of $\mathbb{P}(F \mid \boldsymbol{Y})$.*

As a special case, when $\boldsymbol{X}$ is $\mathcal{G}$-measurable, we may take $p(\omega, F) = \mathbf{1}_F(\boldsymbol{X}(\omega))$ as a version of $\mathbb{P}(\boldsymbol{X} \in F \mid \mathcal{G})$. This gives $p(\omega, \mathrm{d}\omega') = \delta_{\boldsymbol{X}(\omega)}(\,\mathrm{d}\omega')$

When a regular conditional law of $\boldsymbol{X}$ knowing $\mathcal{G}$ exists, we say that it is unique $\mathbb{P}$-a.s. if any two candidates P and Q are almost surely equal, in the sense that $\mathbb{P}(\{\omega \in \Omega \mid P(\omega, \cdot) = Q(\omega, \cdot)\}) = 1$. In that case, we denote it by $\mathbb{P}^{\mathcal{G}}_{\boldsymbol{X}}(\omega, \mathrm{d}x)$ or by $\mathbb{P}^{\mathcal{G}}(\omega, \boldsymbol{X} \in \mathrm{d}x)$. The regular conditional distribution of $\boldsymbol{X}$ knowing $\mathcal{G}$ is such that, for any measurable function $\varphi : \mathbb{X} \to \mathbb{R}$ satisfying $\mathbb{E}\left(\left|\varphi(\boldsymbol{X})\right|\right) < +\infty$, we have that:

$$\mathbb{E}\left(\varphi(\boldsymbol{X}) \mid \mathcal{G}\right)(\cdot) = \int_{\mathbb{X}} \varphi(x)\mathbb{P}^{\mathcal{G}}_{\boldsymbol{X}}(\cdot, \mathrm{d}x) \quad \mathbb{P}\text{-a.s.}\,. \tag{B.42}$$

In the same way, *the regular conditional distribution of $\mathbb{P}$ knowing $\mathcal{G}$* is denoted by $\mathbb{P}^{\mathcal{G}}(\omega, \mathrm{d}\omega')$. It is such that, for any integrable random variable $\boldsymbol{X} : \Omega \to \mathbb{R}$,

$$\mathbb{E}\left(\boldsymbol{X} \mid \mathcal{G}\right)(\cdot) = \int_{\Omega} \boldsymbol{X}(\omega')\mathbb{P}^{\mathcal{G}}(\cdot, \mathrm{d}\omega')\,, \ \ \mathbb{P}\text{-a.s.}\,. \tag{B.43}$$

In the same vein, *the regular conditional distribution of $\boldsymbol{X}$ knowing $\boldsymbol{Y}$* is denoted by $\mathbb{P}^{\boldsymbol{Y}}_{\boldsymbol{X}}(y, \mathrm{d}x)$, by $\mathbb{P}^{\boldsymbol{Y}}(y, \boldsymbol{X} \in \mathrm{d}x)$ or by $\mathbb{P}(\boldsymbol{X} \in \mathrm{d}x \mid \boldsymbol{Y} = y)$. It is

[3] This means that $p(\cdot, F)$ and $\mathbb{P}(\boldsymbol{X} \in F \mid \mathcal{G})$ are almost surely equal w.r.t. $\mathbb{P}$.

such that, for any measurable function $\varphi : \mathbb{X} \to \mathbb{R}$ satisfying $\mathbb{E}\left(\left|\varphi(\boldsymbol{X})\right|\right) < +\infty$, we have that:

$$\mathbb{E}\left(\varphi(\boldsymbol{X}) \mid \boldsymbol{Y}\right)(\cdot) = \int_{\mathbb{X}} \varphi(x)\mathbb{P}_{\boldsymbol{X}}^{\boldsymbol{Y}}(\boldsymbol{Y}(\cdot),\, \mathrm{d}x)\ , \ \ \mathbb{P}\text{-a.s.}\ . \tag{B.44}$$

Example B.20. The following expressions are well known and may be easily verified. If $\mathcal{G}$ is an atomic σ-field generated by a countable partition $\{\Omega_n\}_{n\in N}$, we have that:

$$\mathbb{P}^{\mathcal{G}}(\omega,\, \mathrm{d}\omega') = \sum_{n, \mathbb{P}(\Omega_n)>0} \mathbf{1}_{\Omega_n}(\omega)\mathbb{P}_{|\Omega_n}(\,\mathrm{d}\omega')\ . \tag{B.45}$$

If $\boldsymbol{Y}$ is a discrete random variable, we have that (see (B.29) and (B.30)):

$$\mathbb{P}^{\boldsymbol{Y}}(y,\, \mathrm{d}\omega') = \mathbb{P}_{|\{\boldsymbol{Y}=y\}}(\,\mathrm{d}\omega')\ . \tag{B.46}$$

If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are discrete random variables, we have that:

$$\mathbb{P}_{\boldsymbol{X}}^{\boldsymbol{Y}}(y,\, \mathrm{d}x) = \sum_{x' \in \boldsymbol{X}(\Omega)} \mathbb{P}(\boldsymbol{X} = x' \mid \boldsymbol{Y} = y)\delta_{x'}(\,\mathrm{d}x)\ . \tag{B.47}$$

If the couple $(\boldsymbol{X}, \boldsymbol{Y})$ has a density $f_{(\boldsymbol{X},\boldsymbol{Y})} > 0$ w.r.t. the Lebesgue measure on $\mathbb{R}^2$, we have that:

$$\mathbb{P}_{\boldsymbol{X}}^{\boldsymbol{Y}}(y,\, \mathrm{d}x) = \frac{f_{(\boldsymbol{X},\boldsymbol{Y})}(x,y)}{\int_{\mathbb{X}} f_{(\boldsymbol{X},\boldsymbol{Y})}(x',y)\,\mathrm{d}x'}\,\mathrm{d}x\ . \tag{B.48}$$

$\triangle$

Proposition B.21 ([30, 89]). *Let $\boldsymbol{X}$ be a random variable taking values in a Borel space. If $\boldsymbol{Y}$ is another random variable, there exists a regular conditional distribution of $\boldsymbol{X}$ knowing $\boldsymbol{Y}$, and it is unique $\mathbb{P}_{\boldsymbol{Y}}$-a.s..*

Using the previous proposition with $\boldsymbol{X} = \mathrm{I}_{\Omega}$ and when $(\Omega, \mathcal{A})$ is a Borel space we obtain as a corollary that there exists a regular conditional distribution of $\mathbb{P}$ knowing $\boldsymbol{Y}$.

The following disintegration formula is widely used.

Proposition B.22 (Disintegration[89]). *Let $\boldsymbol{X} : (\Omega, \mathcal{A}) \to (\mathbb{X}, \mathcal{X})$ be a random variable taking values in a measurable space $(\mathbb{X}, \mathcal{X})$, $\mathcal{G}$ be a subfield of $\mathcal{A}$, and $\boldsymbol{Y} : (\Omega, \mathcal{A}) \to (\mathbb{Y}, \mathcal{Y})$ be a $\mathcal{G}$-measurable random variable. Let also f be a measurable function on $\mathbb{X} \times \mathbb{Y}$ such that $\mathbb{E}\left(\left|f(\boldsymbol{X}, \boldsymbol{Y})\right|\right) < +\infty$. If $\boldsymbol{X}$ has a regular conditional distribution $\mathbb{P}_{\boldsymbol{X}}^{\mathcal{G}}(\omega,\, \mathrm{d}x)$ knowing $\mathcal{G}$, we have that:*

$$\mathbb{E}\left(f(\boldsymbol{X}, \boldsymbol{Y}) \mid \mathcal{G}\right) = \int_{\mathbb{X}} \mathbb{P}_{\boldsymbol{X}}^{\mathcal{G}}(\cdot,\, \mathrm{d}x) f(x, \boldsymbol{Y}(\cdot))\ , \ \ \mathbb{P}\text{-a.s.}\ . \tag{B.49}$$

If $\boldsymbol{X}$ has a regular conditional distribution $\mathbb{P}_{\boldsymbol{X}}^{\boldsymbol{Y}}(\boldsymbol{Y}(\cdot),\,\mathrm{d}x)$ knowing $\boldsymbol{Y}$, we have that:

$$\mathbb{E}\left(f(\boldsymbol{X},\boldsymbol{Y}) \mid \boldsymbol{Y}\right) = \int_{\mathbb{X}} \mathbb{P}_{\boldsymbol{X}}^{\boldsymbol{Y}}(\boldsymbol{Y}(\cdot),\,\mathrm{d}x) f(x,\boldsymbol{Y}(\cdot))\ ,\ \ \mathbb{P}\text{-}a.s.\ . \tag{B.50}$$

Equation (B.50) is usually written under the form

$$\mathbb{E}\left(f(\boldsymbol{X},\boldsymbol{Y}) \mid \mathcal{G}\right)(\omega) = \mathbb{E}\left(f(\boldsymbol{X},y) \mid \mathcal{G}\right)(\omega)_{|y=\boldsymbol{Y}(\omega)}\ ,\ \ \mathbb{P}\text{-a.s.} \tag{B.51}$$

whenever $\boldsymbol{Y}$ is $\mathcal{G}$-measurable. As a corollary, we have that

$$\mathbb{E}\left(f(\boldsymbol{X},\boldsymbol{Y})\right) = \int_{\Omega} \mathbb{P}(\,\mathrm{d}\omega) \int_{\mathbb{X}} \mathbb{P}_{\boldsymbol{X}}^{\mathcal{G}}(\omega,\,\mathrm{d}x) f(x,\boldsymbol{Y}(\omega))\ . \tag{B.52}$$

B.7 Monte Carlo Method

The knowledge of a random phenomenon arises from experiments, which often consist of a set of independent observations. In this section, we analyze this intuitive idea. We first recall what is the product of probability spaces. We then introduce the notion of sample, and we recall the construction of the underlying probability space. We present the celebrated "Strong Law of Large Numbers" and the "Central Limit Theorem". We conclude this section by presenting numerical experiments and practical considerations.

B.7.1 Infinite-Dimensional Product of Probability Spaces

Let $\{(\mathbb{X}_n, \mathcal{X}_n, \mu_n)\}_{n\in\mathbb{N}}$ be a sequence of probability spaces. We denote by $\mathbb{X}_\infty$ the *product space*, that is the Cartesian product of $\mathbb{X}_n$

$$\mathbb{X}_\infty = \prod_{n\in\mathbb{N}} \mathbb{X}_n\ ,$$

and we define the sequence $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ of *coordinate mappings*, namely

$$\begin{aligned} \boldsymbol{X}_n\ :\quad & \mathbb{X}_\infty \to \mathbb{X}_n \\ & (x_0,\dots,x_n,\dots) \mapsto x_n\ . \end{aligned}$$

Our first objective is to equip the product space with a σ-field.

Definition B.23. *The σ-field $\sigma\big(\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}\big)$ generated by the sequence of coordinate mappings is defined to be the smallest σ-field relative to which all $\boldsymbol{X}_n$ are measurable. It is also called the* product σ-field *with components $\mathcal{X}_n$, and is denoted*

$$\mathcal{X}_\infty = \bigotimes_{n\in\mathbb{N}} \mathcal{X}_n\ .$$

In the product space $\mathbb{X}_\infty$, sets of the form $\prod_{n\in\mathbb{N}} A_n$ with $A_n \in \mathcal{X}_n$ are called *cylinders with finite dimensional basis* if $A_n = \mathbb{X}_n$ for all but a finite number of indices n. For such a cylinder A, the (finite) subset of $\mathbb{N}$ for which $A_n \neq \mathbb{X}_n$ is denoted $\mathbb{B}_A$.

Proposition B.24. *The product σ-field $\mathcal{X}_\infty$ is the smallest σ-field containing all cylinders with finite dimensional bases of $\mathbb{X}_\infty$.*

For a proof, see [37, §1.3].

We are now able to define a probability measure on $(\mathbb{X}_\infty, \mathcal{X}_\infty)$.

Theorem B.25. *There exists a unique probability distribution μ_∞ defined on the product σ-field $\mathcal{X}_\infty$ such that for every cylinder A with finite basis $\mathbb{B}_A$,*

$$\mu_\infty(A) = \prod_{i\in\mathbb{B}_A} \mu_i(A_i) \ .$$

For a proof, see [37, §6.4].

The following notation is usual for the product distribution:

$$\mu_\infty = \bigotimes_{n\in\mathbb{N}} \mu_n \ .$$

The probability space $(\mathbb{X}_\infty, \mathcal{X}_\infty, \mu_\infty)$ is called the *infinite-dimensional product probability space* associated with the sequence $\{(\mathbb{X}_n, \mathcal{X}_n, \mu_n)\}_{n\in\mathbb{N}}$. When $(\mathbb{X}_n, \mathcal{X}_n, \mu_n) = (\mathbb{X}, \mathcal{X}, \mu)$ for all $n \in \mathbb{N}$, the following notation is used:

$$(\mathbb{X}_\infty, \mathcal{X}_\infty, \mu_\infty) = \left(\mathbb{X}^{\mathbb{N}}, \mathcal{X}^{\otimes\mathbb{N}}, \mu^{\otimes\mathbb{N}}\right) \ .$$

B.7.2 Samples and Realizations

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $\boldsymbol{X}$ be a random variable, that is a measurable mapping, defined on Ω taking its values in a space $\mathbb{X}$ equipped with a σ-field $\mathcal{X}$:

$$\boldsymbol{X} : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{X}, \mathcal{X}) \ .$$

For the sake of simplicity, we restrict ourselves to the finite-dimensional case $\mathbb{X} = \mathbb{R}^p$, $\mathcal{X} = \mathcal{B}^{\mathrm{o}}_{\mathbb{R}^p}$ being the associated Borel σ-field. We use the term “random variable” for $\boldsymbol{X}$ whatever the dimension p is (see §B.2). We denote by μ the probability distribution of $\boldsymbol{X}$, that is the probability distribution induced by $\boldsymbol{X}$:

$$\mu(A) = \mathbb{P}(\boldsymbol{X}^{-1}(A)) \ , \ \ \forall A \in \mathcal{X} \ .$$

Here we study problems which involve carrying out a sequence of observations of a random phenomenon. More precisely, we are interested in (possibly infinite-dimensional) sequences of independent observations of $\boldsymbol{X}$, that is, sequences of independent random variables which have the same probability distribution as $\boldsymbol{X}$. We first define the notion of sample and realization.

Definition B.26. *A* n-sample *from the probability distribution* μ *is a sequence* $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ *of independent random variables with the same probability distribution* μ.

This definition easily extends to infinite-dimensional samples, namely sequences $\{\boldsymbol{X}_k\}_{k \in \mathbb{N}}$ of *independent identically distributed* (i.i.d.) random variables.

Before using samples, we have to know if it is always possible to get a (possibly infinite-dimensional) sample from the random variable $\boldsymbol{X}$ defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The answer is usually negative, the original probability space being not "big enough" to support independent random variables.[4] There is, however, a *canonical* way to define samples. Let the infinite product of the probability spaces $(\mathbb{X}, \mathcal{X}, \mu)$ become the probability space under consideration

$$(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}) = \left(\mathbb{X}^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}}, \mu^{\otimes \mathbb{N}}\right) ,$$

and consider the coordinate mappings

$$\begin{array}{rccl} \boldsymbol{X}_n : & \mathbb{X}^{\mathbb{N}} & \to & \mathbb{X} \\ & (x_1, \ldots, x_n, \ldots) & \mapsto & x_n \ . \end{array}$$

They are measurable (by definition of $\mathcal{X}^{\otimes \mathbb{N}}$), independent by construction (since $\mu^{\otimes \mathbb{N}}$ is a product probability), and their common probability distribution is μ. They thus constitute an infinite-dimensional sample of $\boldsymbol{X}$. Owing to a change of the probability space, $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ replacing $(\Omega, \mathcal{A}, \mathbb{P})$, it is possible to generate samples of arbitrary size.

Consider now the products of n probability spaces $(\mathbb{X}, \mathcal{X}, \mu)$, namely $\left(\mathbb{X}^n, \mathcal{X}^{\otimes n}, \mu^{\otimes n}\right)$. The projection mappings

$$\begin{array}{rccl} \boldsymbol{\Pi}_n : & \mathbb{X}^{\mathbb{N}} & \to & \mathbb{X}^n \\ & (x_1, \ldots, x_n, \ldots) & \mapsto & (x_1, \ldots, x_n) \ , \end{array}$$

are measurable. Let $\widetilde{\mathcal{F}}_n$ be the σ-field generated by $\boldsymbol{\Pi}_n$:

$$\widetilde{\mathcal{F}}_n = \sigma(\boldsymbol{\Pi}_n) = \sigma(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \ .$$

Then $\{\widetilde{\mathcal{F}}_n\}_{n \in \mathbb{N}}$ is the filtration associated with the sample $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \ldots)$. Otherwise stated, considering $\boldsymbol{X}_n$ as the observation of $\boldsymbol{X}$ delivered at stage n, $\widetilde{\mathcal{F}}_n$ is the σ-field generated by all observations prior to n.

[4] Consider for example a coin toss, and let $\Omega = \{H, T\}$ and $\mathcal{A} = \big\{\emptyset, \{H\}, \{T\}, \Omega\big\}$. If the game is modeled using a real-valued random variable defined on Ω, every potential random variable representing the game can be obtained (by composition with a deterministic function) from the *unique* random variable $(H \mapsto 0, T \mapsto 1)$.

B.7.3 Monte Carlo Simulation

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $\boldsymbol{X}$ be a random variable defined on Ω taking its values in the Borel space $\left(\mathbb{R}^p, \mathcal{B}^o_{\mathbb{R}^p}\right)$. We denote the probability distribution of $\boldsymbol{X}$ by μ, and we consider an infinite-dimensional sample $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \ldots)$ of $\boldsymbol{X}$. According to the previous paragraph, such a sequence exists up to a change of probability space. We suppose from now that $(\Omega, \mathcal{A}, \mathbb{P})$ is "big enough" for such a sequence to exist.

We first recall a classical convergence theorem (Strong Law of Large Numbers).

Theorem B.27. *Let $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \ldots)$ be a sequence of i.i.d. random variables, and let $\boldsymbol{M}_n = (1/n)(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)$. We suppose that $\mathbb{E}\left(\left\|\boldsymbol{X}_1\right\|\right) < +\infty$. Then, the random variable $\boldsymbol{M}_n$ almost surely converges to $\mathbb{E}\left(\boldsymbol{X}_1\right)$ as n goes to infinity:*

$$\boldsymbol{M}_n \xrightarrow{a.s.} \mathbb{E}\left(\boldsymbol{X}_1\right) .$$

A second classical theorem (Central Limit Theorem) gives some indication about the convergence rate of the estimator $\boldsymbol{M}_n$.

Theorem B.28. *Let $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \ldots)$ be a sequence of i.i.d. random variables and let $\boldsymbol{M}_n = (1/n)(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)$. We suppose that $\mathbb{E}\left(\left\|\boldsymbol{X}_1\right\|^2\right) < +\infty$, and we denote $M = \mathbb{E}\left(\boldsymbol{X}_1\right)$ and $\Sigma = \mathrm{Var}\left(\boldsymbol{X}_1\right)$ the mean and the covariance matrix of $\boldsymbol{X}_1$. Then, the sequence of probability distributions of the random variables $\sqrt{n}\big(\boldsymbol{M}_n - M\big)$ narrowly converges towards the centered normal distribution with covariance matrix Σ:*

$$\sqrt{n}\Big(\boldsymbol{M}_n - M\Big) \xrightarrow{\mathcal{D}} \mathcal{N}_p(0, \Sigma) .$$

Otherwise stated, it means that the covariance matrix of $\boldsymbol{M}_n$ is asymptotically equal to Σ/n: the convergence rate is $1/n$, and *it does not depend on the dimension of the space $\mathbb{R}^p$.*

The proof of these two celebrated theorems can be found in any textbook on Probability (e.g. [37]). Other results for the rate of convergence are available, e.g. from Large Deviations Theory [56].

B.7.4 Numerical Considerations

We are now interested in the computational point of view, that is the manipulation of random variables on a computer. As a matter of fact, if the convergence analysis of algorithms involving samples has to be carried out on random variables, their implementation on a computer is done using numerical values. The following definition is useful for numerical considerations.

Definition B.29. *A* realization $(x_1, \dots, x_n)$ *of the* n*-sample* $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)$ *is a value taken by the sample at some* $\omega \in \Omega$*:*

$$(x_1, \dots, x_n) = \big(\boldsymbol{X}_1(\omega), \dots, \boldsymbol{X}_n(\omega)\big) \ .$$

This definition is extended without difficulty to infinite-dimensional sample $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n, \dots)$. Such realizations are obtained using a *pseudo-random number generator*, that is a computational device designed to generate numbers that approximate the properties of random numbers.[5] Such a generator usually delivers the components of the realization $(x_1, \dots, x_n)$ one by one: given a realization $(x_1, \dots, x_{n-1})$ of a $(n-1)$-sample of $\boldsymbol{X}$, a further call to the generator produces a value x_n such that $(x_1, \dots, x_n)$ is a realization of a n-sample of $\boldsymbol{X}$.

Let us illustrate the previous notions with help of a basic example, namely the numerical computation of the expectation of a random variable $\boldsymbol{X}$. Using a infinite dimensional sample $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n, \dots)$ of $\boldsymbol{X}$, we know from Theorem B.27 that $\boldsymbol{M}_n = (1/n)(\boldsymbol{X}_1 + \dots + \boldsymbol{X}_n)$ almost surely converges to $M = \mathbb{E}\big(\boldsymbol{X}\big)$. Moreover, the last summation can be written recursively, namely

$$\boldsymbol{M}_n = \boldsymbol{M}_{n-1} - \frac{1}{n}\big(\boldsymbol{M}_{n-1} - \boldsymbol{X}_n\big) \ .$$

From a realization $(x_1, \dots, x_n, \dots)$ of the sample, we deduce the realization m_n of the estimator $\boldsymbol{M}_n$:

$$m_n = \frac{1}{n}(x_1 + \dots + x_n) \ .$$

Note that the strong law of large numbers asserts that, except on some $\Omega_0 \subset \Omega$ such that $\mathbb{P}(\Omega_0) = 0$, each realization $(x_1, \dots, x_n, \dots)$ of the sample satisfies

$$\lim_{n \to +\infty} m_n = M \ .$$

Using the recursive formulation, the sequence $\{m_n\}_{n \in \mathbb{N}}$ is obtained by a computer as follows.

Algorithm B.30 (Recursive Monte Carlo Estimation).

1. Set $m_0 = 0$ and $n = 1$.
2. Draw a realization x_n of $\boldsymbol{X}$.
3. Compute $m_n = m_{n-1} - (1/n)(m_{n-1} - x_n)$.
4. Set $n = n + 1$ and go to step 2.

As already explained, the value x_n is obtained in such a way that $(x_1, \dots, x_n)$ is a realization of a n-sample of the random variable $\boldsymbol{X}$. The algorithm is stopped after a given number N of iterations (say a few thousands). Outputs of the algorithm are shown at Figure B.1, in the specific case $M = 0$. We have

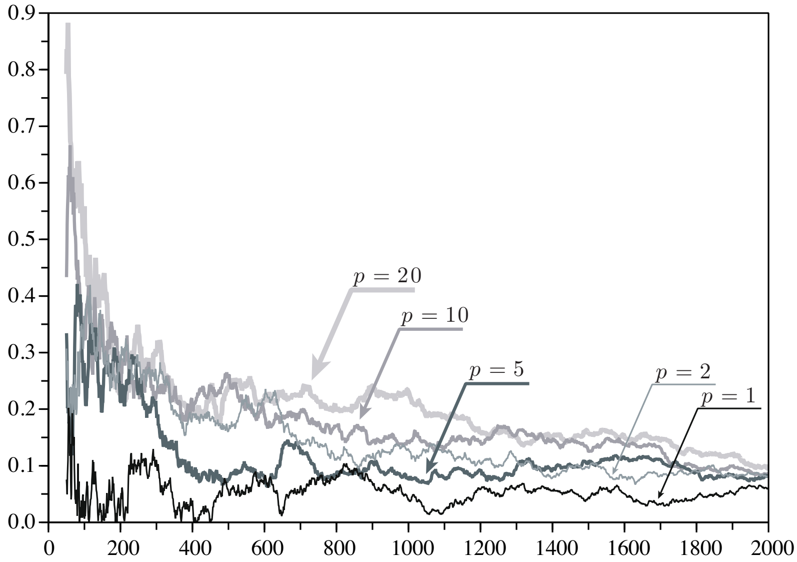

Fig. B.1. Estimation by the Monte Carlo Method

represented the variation of $\|m_n\|$ over the iterations, for different values p of the dimension of the space $\mathbb{X}$.

Let us conclude with two remarks about this algorithm.

1. The output m_N of Algorithm B.30 is a *single* realization of the random variable $\boldsymbol{M}_N$. We have to perform multiple runs of the code in order to obtain statistical conclusions on the output.
2. It is clear from the experiments presented at Figure B.1 that, at least asymptotically, the rate of convergence does not depend on the dimension p of the space $\mathbb{X}$.

[5] sequence of numbers that lacks any pattern, in the computer science terminology.