Probabilities and Potential

Claude Dellahcerie and Paul-André Meyer
typed by Yûki Hirai

February 25, 2018

Contents

Chapter 1

Measurable Spaces

1.1 σ-Fields and Random Variables

Definition of σ-fields

1

Definition. Let Ω be a set. A σ-field on Ω is a family of subsets of Ω which contains the empty set and which is closed under the operation $(\bigcup c, \bigcap c, \quad ^c)$.

The word σ-algebra is synonym frequently used for σ-field [1]. The ordered pair $(\Omega, \mathscr{F})$ consisting of a set Ω and a σ-field $\mathscr{F}$ of subsets of Ω is called a measurable space; the elements of $\mathscr{F}$ are called measurable or $\mathscr{F}$-measurable sets. In the language of probability theory, they are called events.

The set Ω is then called the "sure event" and the empty set the "impossible even"; the operation of taking complements is called passing to the "opposite event". We sometimes use phrases such as: "the event A occurs", "the events A and B occur simultaneously", "the events A and B are incompatible", to express the assertions which we would write respectively in set theoretic terms: "$\omega \in A$", "$\omega \in A \cap B$", "$A \cap B = \emptyset$". The reader will no doubt very quickly become familiar with such statements.

To understand this language and use it fully as an aid to intuition, the points of Ω can be conceived of as the possible results of drawing lots. Every subsets A of Ω is then associated with an "event" in the usual sense of this term, a physical phenomenon which occurs when the result ω "falls in A". Among all possible "events", the σ-field $\mathscr{F}$ contains those which are sufficiently simple to have a probability attributes to them.

Definition of random variables

2

[1] A (Boolean) algebra is a set of subsets of Ω which contains $\emptyset$ and is closed under $(\bigcup f, \bigcap f, \quad ^c)$.

Definition. Let $(\Omega, \mathscr{F})$, $(E, \mathscr{E})$ be two measurable spaces. A mapping f of Ω into E is called measurable if

$$f^{-1}(A) \in \mathscr{F} \quad \text{for all } A \in \mathscr{E}.$$

In the language of probability theory, we also say that f is a random variable (r.v.). If there is any danger of ambiguity, we say explicitly: a random variable on $(\Omega, \mathscr{F})$ with values in $(E, \mathscr{E})$.

3

Example. (1) Let E be the set consisting of the two numbers 0, 1, with the σ-field $\mathscr{E}$ of all subsets of E. A subset A of Ω is an event if and only if its characteristic function χ_A (equal to 1 on A and 0 on $\Omega \setminus A$) is a random variable. Probabilist prefer to call this function the indicator of A and to denote it by I_A. We shall use this terminology.

(2) Let $(\Omega, \mathscr{F})$ be a measurable space and $\mathscr{G}$ is a sub σ-field of $\mathscr{F}$. The identity mapping of $(\Omega, \mathscr{F})$ onto $(\Omega, \mathscr{G})$ then is a random variable.

4

Theorem. Let $(\Omega, \mathscr{F})$, $(G, \mathscr{G})$ and $(E, \mathscr{E})$ are three measurable spaces and $u : \Omega \to G$, $v : G \to E$ are two random variables. The composite mapping $v \circ u$ then is a random variable.

5

Definition. (a) Let Ω be a set and $\mathscr{A}$ a family of subsets of Ω. The σ-field generated by $\mathscr{A}$, denoted by $\sigma(\mathscr{A})$, is the smallest σ-field of subsets of Ω containing $\mathscr{A}$.

(b) Let Ω be a set and $(f_i)_{i \in I}$ a family of mappings of Ω into measurable spaces $(E_i, \mathscr{E}_i)_{i \in I}$. The σ-field generated by the mappings f_i, denoted by $\sigma(f_i,\, i \in I)$, is the smallest σ-field of subsets of Ω with respect to which all the mappings f_i are measurable.

The existence of σ-field described in these two definitions is obvious: on just takes the intersection (in $\mathfrak{P}(\Omega)$) of all σ-fields fir which the sets or functions considered are measurable - there us at least one such σ-field, namely the σ-field consisting of all subsets of Ω.

There is moreover a close relation between parts (a) and (b) of 5: the σ-field generated by a set of subsets is also generated by the indicators of these subsets; the σ-field generated by the mappings f_i is also generated by the family of subsets of the form $f_i^{-1}(A_i)$, where for all i, A_i belongs to $\mathscr{E}_i$

6

Remark. (a) Let $(E, \mathscr{E})$ be a measurable space and f a mapping of E into Ω; then f is measurable with respect to the σ-field $\sigma(\mathscr{A})$ (5(a)) if and only if $f^{-1}(A) \in \mathscr{E}$ for all $A \in \mathscr{A}$ (the family of subsets B of Ω such that $f^{-1}(B) \in \mathscr{E}$ is a σ-field which contains $\mathscr{A}$ and hence also $\sigma(\mathscr{A})$). Similarly, f is measurable with respect to the σ-field $\sigma(f_i,\, i \in I)$ on Ω (5(b)), if and only if each one of mappings $f_i \circ f$ is measurable

(b) The σ-field generated by a family of functions $(f_i)_{i \in I}$ is identical to the union (in $\mathfrak{P}(\Omega)$) of all the σ-fields $\sigma(f_i,\, i \in J)$, with J running through the family of all countable subsets of I.

7

Example. (a) Let E be a topological space. The Borel σ-field of E, denoted by $\mathscr{B}(E)$, is the σ-field generated by the open sets of E. If F is subspace of E, the elements of $\mathscr{B}(F)$ are the traces on F of the elements of $\mathscr{B}(E)$. When E is a space with a countable base $\mathscr{H}$, every open set is a union of sequence of elements of $\mathscr{H}$ and $\mathscr{B}(E)$ is therefore generated by $\mathscr{H}$. For example, the Borel σ-field if the real line is generated by the open intervals with rational endpoints. If E and F are two topological spaces, a Borel mapping of E into F is any mapping of E into F which is measurable from $(E, \mathscr{B}(E))$ to $(F, \mathscr{B}(F))$. Every continuous mapping is Borel. When we consider a topological space E as a measurable space without specifying the σ-field, we always mean the σ-field $\mathscr{B}(E)$.

(b) The continuous real valued functions on E generate on E a smaller σ-field than $\mathscr{B}(E)$, called the Baire σ-field of E. It coincides with $\mathscr{B}(E)$ when E is metrizable: foe if d is a distance on E and F is a closed set, then $F = \{x : f(x) = 0\}$, where f is the continuous function $x \mapsto d(x, F)$.

(c) The family of all subsets of $\mathbb{R}$ which are measurable in the Lebesgue sense is a σ-field richer than $\mathscr{B}(\mathbb{R})$.

(d) The family of all subsets of $\mathbb{R}$ which are either countable or have a countable complement, is a σ-field (generated by the sets $\{x\}$, $x \in \mathbb{R}$). We shall see in Chapter III that this σ-field presents *pathological* characteristics (??).

Product σ-fields

8

Definition. Let $(E_i, \mathscr{E}_i)_{i \in I}$ be a family of measurable spaces; let E denote the product set $\prod_{i \in I} E_i$ and $X_i\,(i \in I)$ the coordinate mappings. The σ-field $\sigma(X_i,\, i \in I)$ is called the product σ-field of the σ-fields $\mathscr{E}_i$ and denoted by $\prod_{i \in I} \mathscr{E}_i$.

We denote the product of two σ-fields by $\mathscr{E}_1 \times \mathscr{E}_2$. Many authors write $\mathscr{E}_1 \otimes \mathscr{E}_2$, $\bigotimes_{i \in I} \mathscr{E}_i$.

Remark. (a) The product σ-field is also generated by the subsets of E of the form $\prod_{i \in I} A_i$, where $A_i \in \mathscr{E}_i$ for all i and $A_i = E_i$ except for a finite number of indices.

(b) Let f_i $(i \in I)$ be mappings of a set Ω into measurable spaces $(E_i, \mathscr{E}_i)$ and f be the mappings $(f_i)_{i \in I}$ of Ω into the product set $\prod_{i \in I} E_i = E$. We give this set the product σ-field: we then have $\sigma(f_i, i \in I) = \sigma(f)$. With the same notation suppose that Ω has been given a σ-field $\mathscr{F}$ and that each of the mappings f_i of Ω into E_i is measurable; then the mapping f of Ω into E is measurable.

(c) If (E_n) is a finite or infinite sequence of topological spaces with countable bases $\mathscr{L}_n$ (in particular, separable metrizable spaces) then $\mathscr{B}(\prod_n E_n) = \prod_n \mathscr{B}(E_n)$. We may indeed assume that $E_n \in \mathscr{L}_n$ for all n. Then the sets of the form $\prod_n U_n$, where $U_n \in \mathscr{L}_n$ for all n, $U_n = E_n$ except for a finite number of indices, form a countable vase of open sets of the topology pf $\prod_n E_n$ and hence $\mathscr{B}(\prod_n E_n)$. On the other hand, they generate $\prod_n \mathscr{B}(E_n)$.

Atoms; separable σ-fields

9 Let $(\Omega, \mathscr{F})$ be a measurable space. The atom of $\mathscr{F}$ are the equivalence classes in Ω for the relation

$$I_A(\omega) = I_A(\omega') \quad \text{for all } A \in \mathscr{F}. \tag{9.1}$$

Every real-valued measurable mapping on Ω (or, more generally, with valued in a separable metrizable space), being a limit of elementary functions, is constant on atoms. The measurable space $(\Omega, \mathscr{F})$ is called Hausdorff [2] if the atoms of $\mathscr{F}$ are the points of Ω. If $(\Omega, \mathscr{F})$ is not Hausdorff, we defined the associated Hausdorff space $(\dot{\Omega}, \dot{\mathscr{F}})$ as follows: $\dot{\Omega}$ is the quotient space of Ω by the relation (9.1) and $\dot{\mathscr{F}}$ is the σ-field consisting of the images in $\dot{\Omega}$ of the elements of $\mathscr{F}$ under the canonical mapping of Ω onto $\dot{\Omega}$.

10 The measurable space $(\Omega, \mathscr{F})$ is called separable (or $\mathscr{F}$ alone is said to be separable) if there exists a sequence [3] of elements of $\mathscr{F}$ which generate $\mathscr{F}$. If the σ-field $\mathscr{F}$ is separable, generated by a sequence (A_n), the atom of $\mathscr{F}$ which contains the point $\omega \in \Omega$ is the intersection of those A_n or A_n^c which contains ω: the atoms are therefore measurable. Note that the Hausdorff space associated with separable space is also separable.

[2] In French "espace séparé". It seems that no "official" English terminology exists.

[3] The closure under $(\bigcup f, \bigcap f, \quad ^c)$ of a countable family of subsets still is countable hence a separable σ-field is generated by a countable Boolean algebra.

11 Two measurable spaces are said to be isomorphic if there exists a bijection between them, which is measurable and has measurable inverse (such a bijection is measurable isomorphism; between topological spaces given their Borel fields, it is also called a Borel isomorphism). Clearly a measurable space isomorphic to a separable metrizable space (with its Borel σ-field) is a separable Hausdorff space. Conversely,

Theorem. Let $(E, \mathscr{E})$ be a separable Hausdorff measurable space. Then $(E, \mathscr{E})$ is isomorphic to a (not necessarily Borel) subspace of $\mathbb{R}$ with its Borel σ-field. More precisely, if (A_n) is a sequence of subsets of E generating $\mathscr{E}$, the mapping [4] f defined by

$$f(x) = \sum_n \frac{1}{3^n} 1_{A_n}(x) \quad \text{for all } x \in E$$

is a measurable isomorphism of $(E, \mathscr{E})$ onto $(f(E), \mathscr{B}(f(E)))$.

Proof. Clearly f is a measurable bijection of E onto $f(E)$. To show that its inverse is measurable, it suffices to show that theσ-field generated by f, which is contained in $\mathscr{E}$, is equal to $\mathscr{E}$, or equivalently that it contains each of the A_n. This follows from the fact that A_n is the inverse image under f of the $y \in [0, 2]$ such that the n-th digit of the expansion of y in base 3. 3 is equal to 1. □

Let us show that, under suitable separability hypotheses, some useful sets are measurable.

12

Theorem. Let $(\Omega, \mathscr{F})$ be a measurable space and $(E, \mathscr{E})$ be a separable Hausdorff measurable spaces.

(a) The diagonal of $E \times E$ belongs to the product σ-field $\mathscr{E} \times \mathscr{E}$.

(b) If f is a measurable mapping of Ω into E, the graph of f in $\Omega \times E$ belongs to the product σ-field $\mathscr{F} \times \mathscr{E}$.

(c) If f and g are measurable mappings of Ω into E, the set $\{f = g\}$ belongs to $\mathscr{F}$.

Proof. Let $\mathscr{A} = (A_n)$ be a countable Boolean algebra generating $\mathscr{E}$ and let D be the set of all $(m, n) \in \mathbb{N} \times \mathbb{N}$ such that $A_m \cap A_n = \emptyset$. Since $\mathscr{E}$ is Hausdorff, the diagonal of $E \times E$ is the complement of the union of the $A_m \times A_n$, where (m, n) runs through D, and hence belongs to $\mathscr{E} \times \mathscr{E}$. Assertion (c) follows from the fact that the set $\{f = g\}$ is the inverse image of the diagonal of $E \times E$ under measurable mapping $\omega \mapsto (f(\omega), g(\omega))$ of $(\Omega, \mathscr{F})$ into $(E \times E, \mathscr{E} \times \mathscr{E})$. Assertion (b) follows from (c) applied to $(\Omega', \mathscr{F}') = (\Omega \times E, \mathscr{F} \times \mathscr{E})$ and to the measurable mappings of Ω' into E, $f' : (\omega, x) \mapsto f(\omega)$ and $g' : (\omega, x) \mapsto x$; the graph of f is equal to the set $\{f' = g'\}$. □

[4] The mapping $(2/3)f$, which takes values in the classical Cantor set, is sometimes called the Marczewski indicator of the sequence (A_n).

In fact, this theorem reduced by 11 to the classical special case where E is separable metrizable space and $\mathscr{E} = \mathscr{B}(E)$. We shall adopt the more intuitive topological terminology, whenever possible, and say for instance, "let E be a separable metrizable space..." rather than "let $(E, \mathscr{E})$ be a separable Hausdorff measurable space..."

1.2 Real-Valued Random Variables

13 As usual, functions taking their values in $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$ are called extended real-valued function, while real-valued functions aren't allowed the values $\pm\infty$.

The word elementary random variable is used currently to describe a random variable taking either countably many values or more precisely finitely many values. The meanig below is the former, unless explicitly stated.

First properties

The functions f, g,... below are assumed to be defined on the same measurable space $(\Omega, \mathscr{F})$.

14 Let f and g be two extended real-valued random variables. Then the functions $f \wedge g$, $f \vee g$, $f + g$, fg (if every where defied) are random variables [5].

15 Let $(f_n)_{n\in\mathbb{N}}$ be a pointwise convergent sequence of extended real-valued random variables and let $f = \lim_n f_n$. Then the function f is a random variables.

This property extends to random variables with values in a metric spaces E: if d is the distance and F is closed in E, the set $\{f \in E\}$ is equal to $\{x : d(f_n(x), F) \xrightarrow[n\to\infty]{} 0\}$.

16 Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of extended real-valued random variables. The convergence set of the sequence (f_n) is measurable.

This property extends to random variables with values in subspace E of a complete metric space F if E is Borel in F: the set A of $\omega \in \Omega$ such that $(f_n(\omega))$ is a Cauchy sequence in F belongs $\mathscr{F}$. The mapping f of Ω into F defined by $f(\omega) = \lim_n f_n(\omega)$ if $\omega \in A$ and $f(\omega) = x_0$ if $\omega \in A$, where x_0 is any point of $F \setminus E$, is measurable; the convergence set of f_n in E is equal to $A \cap f^{-1}(E)$.

17 An extended real-valued function f is measurable, if and only if there exists a sequence (f_n) of measurable elementary functions which increases to to

[5] This is a chance to recall that the convention $0 \cdot \infty = 0/0 = 0$ is universally adopted in integration theory.

f. The following explicit sequence is often useful [6]. It is known as "Lebesgue's approximation of f"

$$f_n = \sum_{k\in\mathbb{Z}} \frac{k}{2^n} I_{\left\{\frac{k}{2^n} < f \le \frac{k+1}{2^n}\right\}} + (-\infty) I_{\{f=-\infty\}}$$

Observe that it converges uniformly to f. If we replace the word "increase" by "converges uniformly", 17 extends to random variables with values in a separable metric space E, given its Borel σ-field (here generated by open balls). For let $(x_n)_{n\in\mathbb{N}}$ be a sequence dense in E and let B_n be the open ball of centre x_n and radius ε. Set $C_n = B_n \setminus (\bigcup_{p<n} B_p)$; the sets C_n then are Borel and disjoint and cover E. If g denotes the elementary function equal to x_n on $f^{-1}(C_n)$, g is measurable and its distance to f is less than ε.

18 The following theorem is due to Doob [] It shows that the notion of a $\sigma(f)$-measurable random variable may be replaced by the less abstract notion of a measurable function of f. By 16 and 17, this extends easily to functions g taking values in a complete separable measurable space: details are left to the reader.

Theorem. Let f be a random variable defied on $(\Omega, \mathscr{F})$ with values in a measurable space $(E, \mathscr{F})$ and g be a real-valued function defied on Ω. Then g is $\sigma(f)$-measurable, if and only if there exits on E a real valued random variable h such that $g = h \circ f$.

Proof. The condition is obviously sufficient. To show that it is also necessary, we begin with the case where g assumes only countably many values a_n $(n \in \mathbb{N})$. Since the sets $A_n = \{g = a_n\}$ are $\sigma(f)$- measurable, they are of the form $f^{-1}(B_n)$ with $B_n \in \mathscr{E}$. Let $C_n = B_n \setminus \left(\bigcup_{p<n} B_p\right)$: these sets belongs to $\mathscr{E}$ and are disjoint. On the other hand

$$f^{-1}(C_n) = A_n \setminus \left(\bigcup_{p<n} A_p\right) = A_n.$$

Then let h be the function on E which takes the values a_n on C_n and 0 (for example) on $E \setminus (\bigcup_n C_n)$. It is obvious that $h \circ f = g$.

We now pass to the general case. By 17, there exists a sequence (g_n) of elementary $\sigma(f)$-measurable random variables converging to g, and g_n is of the form $h_n \circ f$ by the above. Let H be the convergence set of the sequence (h_n): H is $\mathscr{E}$-measurable and contains $f(\Omega)$. We set

$$h(\omega) = \lim_n h_n(\omega) \quad \text{for } \omega \in H$$
$$h(\omega) = 0 \quad \text{for } \omega \in H^c.$$

The function thus constructed is the required function. □

[6] If f is positive, $f_n = 2^{-n} \sum_{k>0} I_{\{f > k2^{-n}\}}$

The "Monotone Class Theorem"

The two theorems 19 and 21 are extremely useful.

19

Theorem. Let $\mathscr{C}$ be a family of subsets of Ω containing $\emptyset$ and closed under $(\bigcup f, \bigcap f)$. Let $\mathscr{M}$ be a family of subsets of Ω containing $\mathscr{C}$, and closed under $(\bigcup mc, \bigcap mc)$ ($\mathscr{M}$ is a "monotone class"). Then $\mathscr{M}$ contains the closure $\mathscr{S}$ of $\mathscr{C}$ under $(\bigcup c, \bigcap c)$.

If in addition $\mathscr{C}$ is closed under the operation $\complement$, $\mathscr{M}$ contains the σ-field generated by $\mathscr{C}$.

Proof. [7] To abbreviate, we call any set of subsets closed under $(\bigcup f, \bigcap f)$ a horde. Let $\mathscr{H}$ be a maximal horde among the hordes contained in $\mathscr{M}$ and containing $\mathscr{C}$ (Zorn's lemma). We show that $\mathscr{H}$ is closed under $(\bigcup c, \bigcap c)$. Let (A_n) be a decreasing sequence [8] of elements of $\mathscr{H}$ and let $A = \bigcap A_n$; the family of all subsets of the form $(H \cap A) \cup H'$, with $H \in \mathscr{H} \cup \Omega$ and $H' \in \mathscr{H}$, is a horde containing $\mathscr{H}$ (take $H = \emptyset$) and A (take $H = \Omega$, $H' = \emptyset$) and contained in $\mathscr{M}$. Since $\mathscr{M}$ is maximal, this horde is identical to $\mathscr{H}$ and hence $A \in \mathscr{H}$. In other words $\mathscr{H}$ is closed under $(\bigcap c)$, so that $\mathscr{H}$ contains the closure of $\mathscr{C}$ under $(\bigcup c)$. The argument is similar for $(\bigcup c)$.

Let $\mathscr{T}$ be the set of all $A \in \mathscr{S}$ such that $A \in \mathscr{S}^c$; if rge complement if every element of $\mathscr{C}$ belongs to $\mathscr{C}$ (or more generally to $\mathscr{S}$), $\mathscr{T}$ contains $\mathscr{C}$ and, in particular, $\emptyset \in \mathscr{T}$. Obviously $\mathscr{T}$ is a σ-field contained in $\mathscr{M}$, and the last sentence in the statement of theorem follows. □

The following example (which encroaches slightly on the beginning of chapter II) illustrates Theorem 19.

20

Theorem. Let $\mathscr{F}_0$ be a set of subsets of Ω, closed under the operation $(\bigcup f, \quad ^c)$ [9]; Let $\mathbb{P}$ and $\mathbb{P}'$ be two probability laws on $\mathscr{F} = \sigma(\mathscr{F}_0)$ such that $\mathbb{P}(A) = \mathbb{P}'(A)$ for all $A \in \mathscr{F}_0$. Then $\mathbb{P}$ and $\mathbb{P}'$ are equal on $\mathscr{F}$.

Proof. Apply 19 with $\mathscr{C} = \mathscr{F}_0$ and with $\mathscr{M}$ denoting the set of elements A of $\mathscr{F}$ such that $\mathbb{P}(A) = \mathbb{P}'(A)$. □

Here is the functional form of the monotone class theorem. We first give the statement we use most often, and then some variants of it.

---

[7] For proofs not using Zorn's lemma, see Chung [Chung˙1968], p.17.

[8] We need only consider such a sequence since $\mathscr{H}$ is closed under $\bigcap f$.

[9] That is, a Boolean algebra 1.

21

Theorem. Let $\mathscr{H}$ be a vector space of bounded real-valued functions defined on Ω, which contains the constant, is closed under uniform convergence and has the following property: for every uniformly bounded increasing sequence of positive functions $f_n \in \mathscr{H}$, the function $f = \lim_n f_n$ belongs to $\mathscr{H}$.

Let $\mathscr{C}$ be a subset of $\mathscr{H}$ which is closed under multiplication. The space $\mathscr{H}$ then contains all bounded functions measurable with respect to the σ-field $\sigma(\mathscr{C})$.

Proof. Let $\mathscr{C}'$ be the algebra generated by the function 1 and the element of $\mathscr{C}$; clearly $\mathscr{C}' \subset \mathscr{H}$. Zorn's lemma allows us to choose a maximal element $\mathscr{A}_0$ of the set of algebras $\mathscr{A}$ satisfying the inclusion $\mathscr{C}' \subset \mathscr{A} \subset \mathscr{H}$. It is known [10] that the function $x \mapsto |x|$ can be uniformly approximated by polynomials on every compact interval of $\mathbb{R}$; the algebra $\mathscr{A}_0$ is obviously closed under uniform convergence and contains the constants, hence it is closed under the operation $f \mapsto |f|$, and therefore also under the operation $\wedge$ and $\vee$. Let g be the limit of a uniformly bounded increasing sequence of positive elements of $\mathscr{A}_0$; it is easily verified that the algebra generated by $\mathscr{A}_0$ and g is contained in $\mathscr{H}$: hence it is identical to $\mathscr{A}_0$ and $g \in \mathscr{A}_0$.

Let $\mathscr{S}$ be the family of all subsets of Ω whose indicators belong to $\mathscr{A}_0$; since $\mathscr{A}_0$ is an algebra, $\mathscr{S}$ is closed under the operation $(\bigcap f, \quad ^c)$. The closure of $\mathscr{A}_0$ under monotone convergence implies that $\mathscr{G}$ is a σ-field and, by virtue of 17, that $\mathscr{A}_0$ contains all the $\mathscr{S}$-measurable bounded functions. It remains to show that $\mathscr{S}$ contains $\sigma(\mathscr{C})$; it obviously suffices to show that $\mathscr{S}$ contains, for every function $f \in \mathscr{A}_0$, the set $B = \{\omega : f(\omega) \geq 1\}$. But the function $g = (f \wedge 1)^+$ belongs to $\mathscr{A}_0$ and the indicator of B is the limit of the decreasing sequence of functions g^n. This concludes the proof. □

22 Variants of Theorem 21 The "monotone class theorem" is one of the basic results of probability theory. To understand the intersect of its variants, let us assume we want to show that the some property $\mathscr{P}$ is true of all bounded functions measurable with respect to some σ-field $\mathscr{T}$. We know how to prove $\mathscr{P}$ for some class $\mathscr{C}$ of functions which generates $\mathscr{T}$. We also know how $\mathscr{P}$ behaves under monotone convergence. Then letting $\mathscr{H}$ denote the set of bounded $\mathscr{T}$-measurable functions which satisfy $\mathscr{P}$, we expect that some variant of the monotone class theorem will tell us that $\mathscr{H}$ contains all the $\mathscr{T}$-measurable functions.

The statement of 21 is appropriate to linear properties: if $\mathscr{P}$ is true for two functions f, g, it is still true for $af + bg$. In this case, little is assumed about $\mathscr{C}$. We repeat the hypotheses:

(22.1) (a) $\mathscr{H}$ is a vector space of bounded functions, closed under bounded monotone convergence and uniform convergence and containing 1:

[10] The Taylor series of $(1-z)^{1/2}$ converges uniformly on $[-1, +1]$. The function $|x| = (1-(1-x^2))^{1/2}$ can therefore be uniformly approximated by polynomials on $[-1, +1]$ (write $z = (1-x^2)$).

(b) $\mathscr{C}$ is closed under multiplication.

The same proof yields the same conclusion under the following hypotheses:

(22.2) (a) $\mathscr{H}$ is a set of bounded functions, closed under bounded monotone convergence and uniform convergence;

(b) $\mathscr{C}$ is an algebra and $1 \in \mathscr{C}$;

(The condition $1 \in \mathscr{C}$ can in fact be weakened by assuming for example that there exists $f_n \in \mathscr{C}$ which increase to 1).

Uniform convergence just serves to pass from closure under multiplication to closure under the operation $\wedge$, $\vee$. Hence a variant without uniform convergence:

(22.3) (a) $\mathscr{H}$ is a set which is closed under bounded monotone convergence;

(b) $\mathscr{C}$ is a $\wedge$-closed vector space and $1 \in \mathscr{C}$.

Here we must change the proof a little, replacing algebras $\mathscr{A}$ containing $\mathscr{C}$ by $\wedge$-closed vector spaces containing $\mathscr{C}$. If $\mathscr{A}_0$ is such a maximal space in $\mathscr{H}$, $\mathscr{A}_0$ is closed under monotone convergence. To conclude, it is necessary to know that if $g \in \mathscr{A}_0$ is positive, them $g^n \in \mathscr{A}_0$: this is very easy, since the convex function $x \mapsto x^n$ is the upper envelope of a sequence of affine functions.

As for 19, we now illustrate Theorem 21 by applications which anticipate Chapter II.

23

Theorem. Let E be a metric space given its Borel σ-field.

(a) Let $\mathbb{P}$ and $\mathbb{P}'$ be two probability laws such that $\int f\mathbb{P} = \int f\mathbb{P}'$ for every bounded continuous functions f. Then $\mathbb{P} = \mathbb{P}'$.

(b) For every bounded Borel function f and all $\varepsilon > 0$, there exists two bounded functions f' and f'' which are respectively u.s.c. and l.s.c., such that $f' \leq f \leq f''$ and $\int (f'' - f')\mathbb{P} < \varepsilon$.

Proof. For (a), apply 21 with $\mathscr{C}$ the algebra of bounded continuous functions and $\mathscr{H}$ the set of bounded Borel functions f such that $\int f\mathbb{P} = \int f\mathbb{P}'$. We know that $\mathscr{C}_b(E)$ generates $\mathscr{B}(E)$(15).

Here the convenient property certainly is closure under multiplication, as is shown by the special case where $E = \mathbb{R}^n$ and $\mathscr{C}_b(E)$ is replaces by the set of bounded infinitely differentiable functions with compact support: in the case of two probability laws $\mathbb{P}$ and $\mathbb{P}'$, or more generally of two locally bounded measures μ and μ' on $\mathbb{R}^n$.

For (b), we take for $\mathscr{C}$ the st of all bounded continuous functions and for $\mathscr{H}$ the set of all bounded Borel functions possessing the above stated approximation property. We then apply the form (22.3) of the theorem, which avoids uniform convergence. To show that $\mathscr{H}$ is closed under bounded monotone convergence, we consider responding u.s.c. functions f'_n l.s.c. functions f''_n such that $f'(n) \leq f_n \leq f''_n$ and $\int (f''_n - f'_n)\mathbb{P} < \varepsilon 2^{-n-2}$. We write $f = \lim_n f_n$, $f'' = \sup_n f''_n$, $f'_1 = \sup_n f'_n$ and verify that $f'_1 \leq f \leq f''$, $\int (f'' - f'_1)\mathbb{P} < \varepsilon/2$. The function f'' is l.s.c. but the function f'_1 is note u.s.c.: it is necessary to take f' to be a function $\sup_{n \leq N} f'_n$, where N is chosen sufficiently large so that $\int (f'_1 - f')\mathbb{P} < \varepsilon/2$. □

Here is another example of the use of Theorem 21, useful in the theory of Markov processes.

24

Theorem. Let $(\Omega, \mathscr{F}, P)$ be a probability space and X and Y two random variables with values in a separable metric space E. To check that $X = Y$ $\mathbb{P}$-a.s., it suffices to check that, for every pair (f, g) of bounded continuous functions on E,

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[g(X)f(X)] \tag{24.1}$$

Proof. Let $\mathscr{H}$ be the set of all bounded Borel functions $h(x, y)$ on $E \times E$ such that $\mathbb{E}[h(X, X)] = \mathbb{E}[h(X, Y)]$: $\mathscr{H}$ is a vector space closed under bounded monotone convergence and uniform convergence. Let $\mathscr{C}$ be the set (closed under multiplication) of all functions of the form $(x, y) \mapsto f(x)g(y)$, where f and g are continuous and bounded on E [11]. Formula (24.1) tell that $\mathscr{C} \subset \mathscr{H}$ and we know that $\mathscr{C}$ generates the σ-field $\mathscr{B}(E) \times \mathscr{B}(E) = \mathscr{B}(E \times E)$. By 21, $\mathscr{H}$ contains all bounded Blrel functions. We concludes by taking $h(x, y)$ to be the indicator of the complement of the diagonal. □

[11] The function $(x, y) \mapsto f(x)g(y)$ is frequently denoted by $f \otimes g$.

Chapter 2

Probability Laws and Mathematical Expectations

As said in the introduction, we assume that our reader is familiar with the more classical parts if measure theory. The first part of this chapter is therefore simply a summary, intended to present the terminology of probability theory. we resume giving complete proofs in the paragraph devoted to uniform integrability.

2.1 A Summary of Integration Theory

Probability law

1

Definition. A probability law on a measurable space $(\Omega, \mathscr{F})$ is a measure $\mathbb{P}$ defined on $\mathscr{F}$, which is positive and has a total mass of 1. The triple $(\Omega, \mathscr{F}, P)$ is called a probability space.

2 In other words, $\mathbb{P}$ is a positive function defined on $\mathscr{F}$ such that $\mathbb{P}(\Omega) = 1$, which satisfies the following property ("countable additivity"): $\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n)$ for every sequence $(\mathscr{A}_n)_{n\in\mathbb{N}}$ of disjoint events.

The number $\mathbb{P}(A)$ us called probability of the event A. An event whose probability is equal to 1 is said to almost sure. Let f and g BE two random variables defined on $(\Omega, \mathscr{F})$ with values in the same measurable space $(E, \mathscr{E})$. If the set $\{\omega : f(\omega) = g(\omega)\}$ is an event[1] of probability 1, we write

$$f = g \quad \text{a.s.}$$

where "a.s." is an abbreviation of "almost surely". Similarly, we shall write "$A = B$ a.s." to express that two events A and B differ only by a set of zero

[1]This is the case if $(E, \mathscr{E})$ is separable and Hausdorff (12).

probability. More generally, we use the expression "almost surely" in the same way as people use "almost everywhere" in measure theory. In fact probabilists freely use the vocabulary of measure theory alongside their own: this enables them to avoid repetition and makes their books very pleasant to read.

3 A probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called complete if every subsets A of Ω which is contained in a $\mathbb{P}$-negligible set belongs to the σ-field $\mathscr{F}$ (and then necessarily $\mathbb{P}(A) = 0$). We shall return to this notion in 32 and prove there that any probability space can be completed.

4

Example. (a) Let I be the interval $[0, 1]$. Let us set, for every $A \in \mathscr{B}(I)$:

$$\mathbb{P}(A) = \int_A dx \quad \text{(Lebesgue measure.)}$$

Then $\mathbb{P}$ is a probability law on I. $\mathbb{P}$ is not complete; it becomes so when extended to the σ-field of Lebesgue measurable sets.

(b) Let $(\Omega, \mathscr{F})$ be a measurable space and x be a point of Ω. We denote by ε_x the probability law defined by:

$$\varepsilon_X(A) = I_A(x) \quad (A \in \mathscr{F}).$$

This law is also called the degenerate law at x or the unit mass at x. More generally a law $\mathbb{P}$ on a measurable space $(\Omega, \mathscr{F})$ is said to be degenerate if $\mathbb{P}(A) = 0$ or 1 for all $A \in \mathscr{F}$. Every real-valued random variable then is a.s. equal to constant.

Mathematical expectation

5

Definition. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and f be an integrable real-valued random variable. The integral $\int_\Omega f(\omega)\mathbb{P}(d\omega)$ is called the mathematical expectation of the random variable f and is denoted by the symbol $\mathbb{E}[f]$.

We shall henceforth omit the adjective "mathematical".

We give few details on integration theory proper. We just state the two theorems which are most often used and make a few remarks.

6

Theorem (The dominated convergence theorem). Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of real-valued random variables which converges almost surely[2], and let f be a random variable a.s. equal to $\lim_n f_n$. If the f_n are bounded in absolutely value by some integrable function, f is integrable and $\mathbb{E}[f] = \lim_n \mathbb{E}[f_n]$.

[2]Or even only in probability (see 10)

Given a positive random variable f, finite or not, which is not integrable, we use the convention $\mathbb{E}[f] = +\infty$. Then the following theorem holds.

7

Theorem (Fatou's lemma). Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of positive random variables; then we have:

$$\mathbb{E}\left[\liminf_n f_n\right] \leq \liminf_n \mathbb{E}[f_n].$$

This inequality can be replaced by equality when the sequence is increasing, whether the integrals are finite or not. This last result is known as Lebesgue's monotone convergence theorem.

8 In conformity with Bourbaki's notation, we denote by $\mathscr{L}^p(\Omega, \mathscr{F}, \mathbb{P})$ (or simply $\mathscr{L}^p$) the vector space of real-valued random variables whose p-th power is integrable ($1 \leq p < \infty$) and by L^p the quotient space of $\mathscr{L}^p$ by the equivalence relation defined by almost surely equality. For every real-valued measurable function f, we set

$$\|f\|_p = (\mathbb{E}[|f|^p])^{1/p} \quad \text{(possibly } +\infty).$$

Similarly, we denote by $\mathscr{L}^\infty(\Omega, \mathscr{F})$ the space (independence of $\mathbb{P}$) of bounded random variables, with the norm of uniform convergence, and by $L^\infty(\Omega, \mathscr{F}, \mathbb{P})$ the quotient space of $\mathscr{L}^\infty$ by the same equivalence relation. The norm of an element of L^∞ (the essential supremum of $|f|$) is denoted by $\|f\|_\infty$.

We shall use without further reference the following properties of the spaces L^p: the fact that L^p is a Banach space (see for example Dunford-Schwartz [], p.146); Höllder's inequality (ibid. p.119); the fact that the dual of L^1 is L^∞ (ibid. p.289). Another necessary result is the Radon-Nikodym theorem (ibid. p.176), which will also established in chapter V as an application of martingale theory.

9 The following two remarks are useful.

(a) Let f be an integrable random variable which is measurable with respect to a sub σ-field $\mathscr{G}$ of $\mathscr{F}$. Then f is a.s. positive, if and only if

$$\int_A f(\omega)\mathbb{P}(d\omega) \geq 0 \quad \text{for all } A \in \mathscr{G}.$$

(Take A be the event $\{f < 0\}$).

It follows in particular that two integrable random variables f and g which are both $\mathscr{G}$-measurable and have the same integral on every set on $\mathscr{G}$ are a.s. equal.

(b) Let f and g be two integrable random variables; we say that f and g are orthogonal if the product $f \cdot g$ is integrable and has zero expectation. Let $\mathscr{G}$ denote a sub-σ-field of $\mathscr{F}$, U be the closed subspace of L^1 consisting of all classes of $\mathscr{G}$-measurable random variables, and V be the subspace of L^∞ consisting of all classes of bounded random variables orthogonal to every element of U. It follows from the Hahn-Banach theorem that every random variable $f \in \mathscr{L}^1$ orthogonal to every element of V is a.s. equal to a $\mathscr{G}$-measurable function.

Convergence of random variables

10 We now recall, restricting our selves to the case of sequences, the main type convergence of real-valued random variables[3].

Let (f_n) be a sequence of random variables defined on $(\Omega, \mathscr{F}, P)$. We say that the sequence (f_n) converges to a random variable f:

- almost surely if $\mathbb{P}\{\omega : f_n(\omega) \to f(\omega)\} = 1$,
- in probability if $\lim_n \mathbb{P}\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\} = 0$ for all $\varepsilon > 0$,
- in the strong sense in L^p if the f_n and f belong to $\mathscr{L}^p$ and $\lim_n \mathbb{E}[|f_n - f|^p] = 0$,
- in the weak sense in L^1 (or alternatively: in the sense of topology $\sigma(L^1, L^\infty)$) if the f_n and f belong to $\mathscr{L}^1$ and, for every random variable g in $\mathscr{L}^\infty$, $\lim_n \mathbb{E}[f_n \cdot g] = \mathbb{E}[f \cdot g]$,
- in the weak sense in L^2 (or alternatively; in the sense of topology $\sigma(L^2, L^2)$) if the f_n and f belong to $\mathscr{L}^2$ and, for every random variable g in $\mathscr{L}^2$, $\lim_n \mathbb{E}[f_n \cdot g] = \mathbb{E}[f \cdot g]$.

We shall return to weak convergence in L^1 in the section concerning uniform integrability. We just recall here that almost sure convergence and strong convergence in L^p imply convergence in probability and that every sequence which converges in probability contains a subsequences which converges almost surely. More precisely, let us set for every real-valued random variable f

$$\pi[f] = \mathbb{E}[|f| \wedge 1].$$

Then the function $(f, g) \mapsto \pi[f - g]$ is a pseudo-metric which defines convergence in probability; if the sequence (f_n) satisfies the property

$$\sum_n \pi[f_n - f_{n+1}] < \infty$$

it converges in probability and almost surely (see for example: Dunford and Schwartz [], p.150).

[3] Or a.s. finite extended real-valued. The definitions relating to convergence in probability need not slight modification for r.v. which are not a.s. finite.

Image laws

11

Definition. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $(E, \mathscr{E})$ be a measurable space and f be a random variable from Ω to E. The image law of $\mathbb{P}$ under f, denoted by $f(\mathbb{P})$, is the law $\mathbb{Q}$ on $(E, \mathscr{E})$ defined by:

$$\mathbb{Q}(A) = \mathbb{P}(f^{-1}(A)) \quad (A \in \mathbb{E}).$$

This law is also called the law of or the distribution of f.

Let g be a measurable mapping of $(E, \mathscr{E})$ into a measurable space $(G, \mathscr{G})$. We have the obvious equation:

$$g(f(\mathbb{P})) = (g \circ f)(\mathbb{P})$$

("transitivity of image laws").

12

Theorem. Let h be a real-valued random variable on $(E, \mathscr{E})$; h is $\mathbb{Q}$-integrable if and only if $h \circ f$ is $\mathbb{P}$-integrable and then:

$$\int_E h(x)\mathbb{Q}(dx) = \int_\Omega (h \circ f)(\omega)\mathbb{P}(d\omega).$$

Integration of probability laws; Fubini's Theorem

13

Definition. Let $(\Omega, \mathscr{F})$ and $(E, \mathscr{E})$ be two measurable spaces. A family $(P_x)_{x \in E}$ of probability laws on $(\Omega, \mathscr{F})$ is said to be $\mathscr{E}$-measurable if the function $x \mapsto P_x(A)$ is $\mathscr{E}$-measurable for all $A \in \mathscr{F}$.

Given such a family $(\mathbb{P}_x)_{x \in E}$, we have the following statement:

14

Theorem (Fubini's Theorem). Let $\mathbb{Q}$ be a probability law on $(E, \mathscr{E})$. Let $(U, \mathscr{U})$ denoted the measurable space $(E \times \Omega, \mathscr{E} \times \mathscr{F})$.

(1) Let f be a real-valued random variable defined on $(U, \mathscr{U})$. Each one of the partial mappings $x \mapsto f(x, \omega)$, $\omega \mapsto f(x, \omega)$ is corresponding factor space.

(2) There exists one and only one probability law $\mathbb{S}$ on $(U, \mathscr{U})$ such that, for all $A \in \mathscr{E}$ and $B \in \mathscr{F}$,

$$\mathbb{S}(A \times B) = \int_A \mathbb{P}_x(B)\mathbb{Q}(dx). \tag{14.1}$$

(3) Let f be a positive[4] random variable on $(U, \mathscr{U})$. The function

$$x \longmapsto \int_\Omega f(x,\omega)\mathbb{P}_x(dw).$$

is $\mathscr{E}$-measurable and:

$$\int_U f(x,\omega)\mathbb{S}(dx,d\omega) = \int_E \mathbb{Q}(dx)\int_\Omega f(x,\omega)\mathbb{P}(d\omega). \tag{14.2}$$

This relation still holds true if f is $\mathbb{S}$-integrable; but one can then only assert that $\omega \mapsto f(x,\omega)$ is $\mathbb{P}_x$ integrable for $\mathbb{Q}$-almost all $x \in E$.

Remark. (a) If f is neither positive nor $\mathbb{S}$-integrable, the right-hand side of (14.2) may be meaningful without the left-hand side being so.

(b) If all the $\mathbb{P}_x$ are equal to the same law $\mathbb{P}$, the law $\mathbb{S}$ is called the product (law) od $\mathbb{Q}$ and $\mathbb{P}$ denoted by $\mathbb{Q}\otimes\mathbb{P}$. The probability space $(U, \mathscr{U}, \mathbb{Q}\otimes\mathbb{P})$ is not complete in general. Fubini's Theorem is often stated for product laws only and slightly different form: assume that the factor spaces are complete and that f is measurable on the completed product space; assertion (1) the is no longer true, but still the partial mappings $x \mapsto f(x,\omega)$ (resp. $\omega \mapsto f(x,\omega)$) are $\mathscr{E}$-measurable (resp. $\mathscr{F}$-measurable) for $\mathbb{Q}$-almost all $x \in E$ (resp. for $\mathbb{P}$-almost all $\omega \in \Omega$).

(c) The definition of the product of finitely many probability laws is obvious. We do not study here infinite products, which, however, are examples of inverse limit of probability laws, see Chapter III.

15

Definition. In the notion of 14, the integral of family $\mathbb{P}_x$ with respect to $\mathbb{Q}$, denoted by $\int_E \mathbb{P}_x\mathbb{Q}(dx)$, is the image law of $\mathbb{S}$ under the projection mapping of $E\times\Omega$ onto Ω.

By combining 12 and 14 we get the following theorem.

16

Theorem. Let $\mathbb{P}$ denote the law $\int_E \mathbb{P}_x\mathbb{Q}(dx)$ and f be a positive random variable on $(\Omega, \mathscr{F})$. Then the function $x \mapsto \int_\Omega f(\omega)\mathbb{P}(d\omega)$ is $\mathscr{E}$ measurable and

$$\int_\Omega f(\omega)\mathbb{P}(d\omega) = \int_E \mathbb{Q}(dx)\int_\Omega f(\omega)\mathbb{P}_x(d\omega).$$

This assertion is also true for every $\mathbb{P}$-integrable random variable f; However f is $\mathbb{P}_x$-integrable only for $\mathbb{Q}$-almost all $x \in E$, so that $\int_\Omega f(\omega)\mathbb{P}_x(d\omega)$ is defined $\mathbb{Q}$-a.s., and no longert on the whole E.

[4] Recall that the integral has been defined for all positive measurable function (cf. 6)

2.2 Supplement of Integration

Uniformly integrable random variables

All the random variables considered in this section are real-valued and defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$ [5].

17

Definition. Let $\mathscr{H}$ be a subset of the space $\mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P})$. $\mathscr{H}$ is called a uniformly integrable set if the integral

$$\int_{\{|f|>c\}} |f(\omega)| \mathbb{P}(d\omega) \quad (f \in \mathscr{H}) \tag{17.1}$$

tend uniformly to 0 as the positive number c tends to $+\infty$.

Notation. Let f be a random variable. We denote by f^c the function

$$f^c(\omega) = f(\omega) \quad \text{for } |f(\omega)| \leq c$$
$$f^c(\omega) = 0 \quad \text{for } |f(\omega)| > c.$$

We write $f_c = f - f^c$. Definition 17 then take the following form: $\mathscr{H}$ is uniformly integrable if and only if, for every $\varepsilon > 0$, a number c exists so that $\|f_c\|_1 < \varepsilon$ for every $f \in \mathscr{H}$.

18

Remark. (a) Every family of random variables dominated in absolute value by a fixed integrable function (in particular, every finite subset of $\mathscr{L}^1$) is uniformly integrable.

(b) Definition 17 is obviously compactible with a.s. equality of random variables [6]. It only involves the latter through their absolute values; so we may often restrict ourselves to positive random variables.

19

Theorem. Let $\mathscr{H}$ be a subset of $\mathscr{L}^1$; for $\mathscr{H}$ be uniformly integrable, it is necessary and sufficient that the following conditions hold:

(a) the expectation $\mathbb{E}[|f|]$, $f \in \mathscr{H}$, are uniformly bounded [7];

[5] For the case of non-bounded measure, see Dunford-Schwartz []

[6] We can thus speak of uniformly integrable subsets of L^1.

[7] It can be proved that (a) is a consequence of (b) if the law $\mathbb{P}$ is diffuse (i.e. has no atomic part).

(b) for every $\varepsilon > 0$, there exists a number $\delta > 0$ such that the conditions $A \in \mathscr{F}$, $\mathbb{P}(A) \leq \delta$, imply the inequality

$$\int_A |f(\omega)| \mathbb{P}(d\omega) \leq \varepsilon \quad (f \in \mathscr{H}) \tag{19.1}$$

.

Proof. To establish the necessity of conditions (a) and (b), we note that, for every integrable function f and every set $A \in \mathscr{F}$,

$$\int_A |f(\omega)| \mathbb{P}(d\omega) \leq c\mathbb{P}(A) + \mathbb{E}\left[|f_c|\right]. \tag{19.2}$$

Suppose that $\mathscr{H}$ is uniformly integrable and choose c so largest that

$$\mathbb{E}\left[|f_c|\right] < \frac{\varepsilon}{2} \quad (f \in \mathscr{H}).$$

We first obtain (a) by taking $A = \Omega$, then (b) choosing $\delta = \varepsilon/2c$.

Conversely ,suppose that properties (a) and (b) hold, and let $\varepsilon > 0$ be given. Choose some $\delta > 0$ satisfying (b) and let $c = \sup_{f \in \mathscr{H}} E[|f|]/\delta$, (finite by virtues of (a)). Apply (19.1), taking for A the set $\{|f| > c\}$, whose probability is less that δ according to the inequality

$$\mathbb{P}\left\{|f| \geq c\right\} \leq \frac{1}{c}\mathbb{E}\left[|f|\right];$$

we get

$$\int_{\{|f| \geq c\}} |f(\omega)| \mathbb{P}(d\omega) \leq \varepsilon \quad (f \in \mathscr{H})$$

and $\mathscr{H}$ indeed is uniformly integrable. □

20

Theorem. Let $\mathscr{H}$ be a uniformly integrable set; the closed convex hull of $\mathscr{H}$ in $\mathscr{L}^1$ is also uniformly integrable.

Proof. We begin by noting that the closure of a uniformly integrable set in $\mathscr{L}^1$ us also uniformly integrable: this is an immediate consequence of theorem 19. Hence it suffices to show that the convex hull of $\mathscr{H}$ is uniformly integrable. We check conditions (a) and (b) of 19. The first one is obvious. Let us choose δ such that (19.1) holds for every $f \in \mathscr{H}$; let $f_1, \ldots, f_n$ be elements of $\mathscr{H}$, $t_1, \ldots, t_n$ numbers ≥ 0 such that $t_1 + \cdots + t_n = 1$ and A a measurable set such that $\mathbb{P} \leq \delta$. Then

$$\int_A |t_1 f_1 + \ldots t_n f_n| \mathbb{P} \leq t_1 \int_0^t |f_1| \, \mathbb{P} + \cdots + t_n \int_0^t |f_n| \, \mathbb{P} \leq \varepsilon.$$

Hence condition (b) us satisfied. □

Remark. Let H and K be two uniformly integrable subsets of L^1; there union $H \cup K$ obviously is uniformly integrable and so is its convex hull; it then follows from inclusion:

$$\frac{1}{2}(H+K) \subset \text{convex hull of } H \cup K. \tag{20.1}$$

that the sum $H+K$ is uniformly integrable. This result can also be deduced simply from 19.

21 The following result generalizes the dominated convergence theorem.

Theorem. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of integrable random variables which converges almost everywhere[8] to a random variable f in the strong sense in L^1, if and only if the f_n are uniformly integrable. If the random variables f_n are positive, it is also necessary and sufficient that:

$$\lim_n \mathbb{E}[f_n] = \mathbb{E}[f] < \infty.$$

Proof. Assume first that the f_n converge to f in L^1 (which supposes the integrability of f); we show that conditions (a) and (b) of 19 are satisfied. We have for $A \in \mathscr{F}$

$$\int_A |f_n(\omega)| \mathbb{P}(d\omega) \leq \int_A |f(\omega)| \mathbb{P}(d\omega) + \|f_n - f\|_1. \tag{21.1}$$

Condition (a) follows immediately. We choose an integer N such that $\|f_n - f\| \leq \varepsilon/2$ for all $n > N$ and a number δ such that the inequality $\mathbb{P}(A) \leq \delta$ implies $\int_A |g| \mathbb{P} \leq \varepsilon/2$, when g runs through the finite set $\{f_1, f_2, \ldots, f_N, f\}$. The left-hand side of (21.1) is then at most ε for all n provided $\mathbb{P}(A) \leq \delta$, and condition (b) is satisfied.

Conversely, suppose that the functions f_n are uniformly integrable. Then the expectations $\mathbb{E}[|f_n|]$ are uniformly bounded and Fatou's Lemma implies that $\mathbb{E}[|f|] < \infty$. Let us show that f_n converges to f in L^1. We have

$$\mathbb{E}[|f_n - f|] \leq \mathbb{E}[|f_n^c - f^c|] + \mathbb{E}[f_{nc}] + \mathbb{E}[|f_c|]. \tag{21.2}$$

Let $\varepsilon > 0$ be given. Choose c so large that the last two expectations are bounded by $\varepsilon/3$ for all n, and such that $\mathbb{P}\{|f| = c\} = 0$ (which is possible, since there are only countably many t such that $\mathbb{P}\{|f| = t\} > 0$.) Next we can choose n so large that the first expectation is bounded by $\varepsilon/3$, according to Lebesgue' Theorem, since the function $|f_n^c - f|$ are uniformly bounded and converge almost everywhere to 0. The left-hand side of (21.2) the is at most ε, and convergence in norm is established.

It remains to show that the convergence of $\mathbb{E}[f_n]$ to $\mathbb{E}[f] < \infty$ implies, when f_n are positive, the convergence of $\mathbb{E}[|f_n - f|] \to 0$ (and consequently the uniform integrability of the f_n). To this end, we write:

$$f + f_n = (f \vee f_n) + (f \wedge f_n).$$

[8] Or only in probability.

$\mathbb{E}[f \wedge f_n]$ tends to $\mathbb{E}[f]$ by Lebesgue's Theorem. On the other hand, $\mathbb{E}[f + f_n]$ tends to $2\mathbb{E}[f]$ by hypothesis. It follows that $\mathbb{E}[f \vee f_n]$ tends to $\mathbb{E}[f]$. We then deduce from the relation

$$|f - f_n| = f \vee f_n - f \wedge f_n$$

that $\mathbb{E}[|f - f_n|]$ tends to 0. □

We give a complete proof of the following theorem (due to la Vallée-Poussin), because it helps to understand the signification of uniform integrability. However, the most useful part of it is the implication (2) $\Longrightarrow$ (1), which is also the easier to establish. For example, every bounded subset of L^2 is uniformly integrable (take $G(t) = t^2$).

22

Theorem. Let $\mathscr{H}$ be a subset of $\mathscr{L}^1$. The following properties are equivalent:

(a) $\mathscr{H}$ is uniformly integrable.

(b) There exists a positive function $G(t)$ defined on $\mathbb{R}_+$ such that $\frac{G(t)}{t} = +\infty$ and[9]

$$\sup_{f \in \mathscr{H}} \mathbb{E}\left[G \circ |f|\right] < \infty. \tag{22.1}$$

Proof. The establish that (2) $\Longrightarrow$ (1), let $\varepsilon > 0$ be given and let $a = \frac{M}{\varepsilon}$, where M is the value of the left-hand side of 22.1. We choose c so large that $\frac{G(t)}{t} \geq a$ for all $t \geq c$. Then we have $|f| \leq \frac{G \circ |f|}{a}$ on the set $\{|f| \geq c\}$ and consequently

$$\int_{\{|f| \geq c\}} |f| \, \mathbb{P} \leq \frac{1}{a} \int_{\{|f| \geq c\}} G \circ |f| \, \mathbb{P} \leq \frac{1}{a} M = \varepsilon$$

for every function $f \in \mathscr{H}$. Definition 17 is therefore satisfied.

We now establish the converse by constructing a function $G(t)$ of the form $\int_0^t g(s)ds$, where g is an increasing function equal to zero at $t = 0$, which tends to ∞ with t and takes a constant value g_n on each interval $[n, n+1[$ $(n \in \mathbb{N})$. We write, for each function $f \in \mathscr{H}$.

$$a_n(f) = \mathbb{P}\{|f| > n\}.$$

Since $g_0 = 0$, we have

$$\begin{aligned} \mathbb{E}\left[G \circ |f|\right] &\leq g_1 \cdot \mathbb{P}\{1 < |f| \leq 2\} + (g_1 + g_2) \cdot \mathbb{P}\{2 < |f| \leq 3\} + \dots \\ &= \sum_{n=1}^{\infty} g_n a_n(f). \end{aligned}$$

[9]The function G which we construct is also convex.

Hence it remains to show that it is possible to choose coefficients g_n which tend to infinity as n increases, such that the sums $\sum g_n \cdot a_n(f)$ are uniformly bounded. We choose an increasing sequence of integers c_n, which tend to infinity, such that

$$\int_{|f|\geq c_n} |f|\,\mathbb{P} \leq 2^{-n} \quad (f \in \mathscr{H})$$

according to our assumption of uniform integrability. We have:

$$\int_{\{|f|\geq c_n\}} |f|\,\mathbb{P} \geq \sum_{k=c_n}^{\infty} k\mathbb{P}\{k < |f| \leq k+1\} \geq \sum_{m=c_n}^{\infty} \mathbb{P}\{|f| > m\} = \sum_{m=c_n}^{\infty} a_m(f).$$

It follows that the sum $\sum_n \sum_{c_n}^{\infty} a_m(f)$ is uniformly bounded for $f \in \mathscr{H}$; but this sum is of the form $\sum_m g_m \cdot a_m(f)$, where g_m denotes the number of integers n such that $c_n \leq m$. The theorem is established. □

Weak topologies

We now give some results on the weak topology $\sigma(L^1, L^\infty)$ closely related in fact to uniform integrability. W make some use of the conditional expectation operators, which will only be defined later (??), but this involves of course no circularity.

We first recall a well known theorem:

23

Theorem (Vitali-Hahn-Saks). Let (μ_n) be a sequence of bounded measures, not necessarily positive, on a measurable space $(\Omega, \mathscr{F})$ and let λ be a bounded positive measure such that the μ_n are absolutely continuous with respect to λ. Suppose that for all $A \in \mathscr{A}$ the limit $\mu(A) = \lim_n \mu_n(A)$ exists and is finite. Then

(1) μ is a bounded measure.

(2) For every $\varepsilon > 0$, there exists a $\eta > 0$ such that the inequality $\lambda(A) \leq \eta$ implies $\sup_n |\mu_n(A)|(A) \leq \varepsilon$. Further, the mass $\|\mu_n\|$ are uniformly bounded.

Proof. We note first that the existence of λ such that μ_n are absolutely continuous with respect to λ is not a restriction: it suffices to take $\lambda = \sum |\mu_n|/2^n \|\mu_n\|$. Then comparing (2) and (19), we may state (2) in a different way: the densities μ_n/λ are uniformly integrable with respect to λ.

Let Φ be the subset of $L^1(\lambda)$ consisting of the equivalent classes of indicators of elements of $\mathscr{F}$ (we shall denoted the classes by the elements of $\mathscr{F}$ they present). Φ is closed in $L^1(\lambda)$, hence Φ is a complete metric space. The function $A \to \mu_n(A)$ are continuous on Φ and converge pointwise to $A \mapsto \mu(A)$.

Let $\alpha > 0$ and let

$$L_j = \{U \in \Phi \colon \forall m \geq j,\ \forall n \geq j,\ |\mu_n(U) - \mu_m(U)| \leq \alpha\}.$$

L_j is a closed subset of Φ and the union of the L_j is the whole of Φ. By Baire's Theorem, there exists a j such that L_j has an interior point A. In other words, there exists an integer j and a number $h > 0$ such that the relations $n \geq j$, $m \geq j$, $\lambda(B \triangle A) \leq h$ imply $|\mu_n(B) - \mu_n(A)| \leq \alpha$. $\square$

Chapter 3

Complements to Measure Theory

Thanks go Hunt [Hunt˙1957], Choquet's theorem on capacitablility has become one of the fundamental tools of probability htoery. This theorem is proved in paragraph 2 and constitutes the core of the chapter. Paragraph 1 contains the elements of analytic set theory necessary to prove Choquet's theorem and other results useful to probabilists (Blackwell's theorem for instance). Paragraph 3 is devoted to bounded Radon measures.

We have tried to restrict ourselves to really useful result, either for probability theory or for potential theory except in (the appendix, which contains some luxury theorems). But this does not mean they are all equally important. The reader that looks for essentials may limit himself to nos 1-13, 27-32 and 44.

3.1 Analytic Sets

Pavings

1 Let E be a set. A paving on E is any family of subsets of E which contains the empty set; the pair $(E, \mathscr{E})$ consisting of a set E and a paving $\mathscr{E}$ on E is called a paved set. This terminology is used only in this chapter and the applications which depend on it.

Let $(E_i, \mathscr{E}_i)_{i\in I}$ be a family of paved set. The product paving of the $\mathscr{E}_i$ (resp. the sum of paving[1] of the $\mathscr{E}_i$) is the paving on the set $\prod_{i\in I} E_i$ (resp. $\sum_{i\in I} E_i$) consisting of the subsets of the form $\prod_{i\in I} A_i$ (resp. $\sum_{i\in I} A_i$) where $A_i \subset E_i$ belongs to $\mathscr{E}_i$ for all i (and, in the case of sum, differs from $\emptyset$ only for finitely many indices).

> The first edition of this book gave a different definition of the product paving, analogous to that of the sum paving, insisting that $A_i = E_i$ except for finite number of indices. It then follows, when this number

[1] Recall that the sum of the E_i (denoted by $\sum_{i\in I} E_i$ or $\coprod_{i\in I} E_i$) is the union of the sets $E_i \times \{i\}$.

> is equal to 0, that the whole space belongs to every product paving, which causes some inconveniences. The present definition is better, given that we only consider countable product or sum.

It should be noted that, when the $\mathscr{E}_i$ are σ-fields, the product paving of the $\mathscr{E}_i$ is not the same as the product σ-field of the $\mathscr{E}_i$ (the latter being generated by the product paving when I is countable). Hence the is some ambiguity in using notations such as $\prod_{i\in I}\mathscr{E}_i$ or $\mathscr{E}\times\mathscr{F}$ to denote a product paving. We shall nevertheless use them, in this chapter only[2], being explicit when necessary.

Compact and Semi-compact Pavings

2 Let $(E,\mathscr{E})$ be a paved set and $(K_i)_{i\in I}$ be a family of elements of $\mathscr{E}$. We say that this family has the finite intersection property if $\bigcap_{i\in I_0} K_i = \emptyset$ for every finite subsets $I_0 \subset I$. This amounts to saying that the sets K_i belongs to a filter or also, by the ultrafilter theorem[3], that there exists an ultrafilter $\mathfrak{U}$ such that $K_i \in \mathfrak{U}$ for all $i \in I$.

3

Definition. Let $(E,\mathscr{E})$ be a paved set. The paving $\mathscr{E}$ is said to be compact (resp. semi-compact) if every family (resp. evry countable family) of elements of $\mathscr{E}$, which has the finite intersection property, has a non-empty intersection[4].

For instance, if E is a Hausdorff topological space, the paving consisting of the compact subsets of E (henceforth denoted by $\mathscr{K}(E)$) is compact. Abstract compact pavings are seldom found: an interesting example is that of the "islets" of $\mathbb{N}^{\mathbb{N}}$ (no 77 in the appendix).

Let $\mathscr{E}$ be a compact (resp. semi-compact) paving on E; then the paving $\mathscr{E}\cup\{E\}$ is compact (resp. semi-compact).

> The definition of analytic sets given in this edition no longer uses semi-compact pavings. Hence the reader can omit every reference to it. The reason for retaining it are of a purely aesthetic nature.

4

Theorem. Let E be a set with a compact (resp. semi-compact) paving $\mathscr{E}$ and let $\mathscr{E}'$ be the closure of $\mathscr{E}$ under $(\bigcup f, \bigcap c)$. Then the paving $\mathscr{E}'$ is compact (resp. semi-compact).

Proof. Let $\mathscr{F}$ be a closure of $\mathscr{E}$ under $(\bigcup f)$. Then $\mathscr{E}'$ is the closure of $\mathscr{F}$ under $(\bigcap a)$ (reap. $(\bigcap c)$). The latter closure obviously preserves compactness and

[2] The best solution consists in using the symbol $\otimes$ for σ-fields, as does Neveu [Neveu˙1964]

[3] Bourbaki [Bourbaki˙1966] (3rd edition), §6 no 4, Theorem 1.

[4] Here is a simple example of a non-compact semi-compact paving: on a non-countable set, the paving consisting of all finite subsets and all subsets with a countable complement.

hence it will suffice to show that $\mathscr{F}$ is compact (resp. semi-compact). So let us consider a family (resp. countable family) $(K_i)_{i\in I}$ of elements of $\mathscr{F}$, which has the finite intersection property; Let $\mathfrak{U}$ be ultrafilter such that $K_i \in \mathfrak{U}$ for all i. Each set K_i is a union $\bigcup_{j\in J_i} K_{ij}$ of elements of $\mathscr{E}$, where J_i is a finite set. Hence there exists an index $j_i \in J_i$ such that $K_{ij_i} \in \mathfrak{U}$ [5]. The family $(K_{ij_i})_{i\in I}$ therefore has the finite intersection property, hence its intersection is non-empty and so a fortiori is the intersection of the family $(K_i)_{i\in I}$. □

5

Theorem. Let $(E_i, \mathscr{E}_i)_{i\in I}$ be a family of paved sets. If each of the pavings $\mathscr{E}_i$ is compact (resp. semi-compact) so are the product $\prod_{i\in I} \mathscr{E}_i$ and the sum of paving $\sum_{i\in I} \mathscr{E}_i$.

Proof. The proof is immediate as far as the product paving is concerned. Let $\mathscr{H}$ be the paving on the sum set $\sum_{i\in I} E_i$ consisting of all subsets of the form $\sum_{i\in I} A_i$ such that $A_i \neq \emptyset$ for all indices except at most one i, for which A_i belongs to $\mathscr{E}_i$. This paving is obviously compact (semi-compact). It then suffices to note that the sum paving is the closure of $\mathscr{H}$ under $(\bigcup f)$. □

The is no need to attach any importance to the “semi”-compact nature of the paving in the following statement: the gain in generality is illusory.

6

Theorem. Let $(E, \mathscr{E})$ be a paved set and let f be a mapping of E into a set F. Suppose that , for all $x \in F$, the paving consisting of the sets $f^{-1}(\{x\}) \cap A$, $A \in \mathscr{E}$, is semi-compact. Then, for every decreasing sequence $(A_n)_{n\in\mathbb{N}}$ of elements of $\mathscr{E}$.

$$f\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \bigcap_{n\in\mathbb{N}} f(A_n). \tag{6.1}$$

Proof. It suffices to show that we can associate to every $x \in f\left(\bigcap_{n\in\mathbb{N}} A_n\right)$ and element $y \in \bigcap_n A_n$ such that $f(y) = x$. Now the family of sets of the form $f^{-1}(\{x\}) \cap A_n$ has the finite intersection property, hence it has a non-empty intersection and we just choose y in this intersection. □

$\mathscr{F}$-analytic sets

7

Definition. Let $(F, \mathscr{F})$ be a paved set. A subset A of F is called $\mathscr{F}$-analytic if there exists an auxiliary compact metrizable space E and a subset $B \subset E \times F$ belonging to $(\mathscr{K}(E) \times \mathscr{F})_{\sigma\delta}$, such that A is the projection of B onto F. The paving on F consisting of all $\mathscr{F}$-analytic sets is denoted by $\mathscr{A}(\mathscr{F})$.

[5]Bourbaki [] (3rd edition), §6, no 4, Proposition 5. This proof was communicated to us by G. Mokobodzki.

It follows immediately from the definition that every $A \in \mathscr{A}(\mathscr{F})$ is contained in some element of $\mathscr{F}_\sigma$. In particular, the whole space F is $\mathscr{F}$-analytic if and only if it belongs to $\mathscr{F}_\sigma$ (8 below).

> Definition 7 involves a variable compact space E. We show in the appendix that replacing E by the fixed compact space $\overline{\mathbb{N}}^{\mathbb{N}}$ ($\overline{\mathbb{N}}$ being the one point compactification of $\mathbb{N}$), or by $\overline{\mathbb{R}}$, leads ti the same class of analytic sets. The same is true, on the other hand, if E is replaced by a variable semi-compact paved space $(E, \mathscr{E})$, as was done in the first edition. Finally, the $\mathscr{F}$-analytic sets are those which are constructed from Souslin's operation (A) applied to elements of $\mathscr{F}$.

8

Theorem. $\mathscr{F} \subset \mathscr{A}(\mathscr{F})$; the paving $\mathscr{A}(\mathscr{F})$ is closed under $(\bigcup c, \bigcap c)$.

Souslin Measurable Spaces, etc.

Index