Note 1

Review of measure spaces, measures and integration.

see also A. Bovier's script for SS17, Chapter 1 [pdf].

A class $\mathscr{A} \subseteq \mathscr{P}(\Omega)$ is an *algebra* iff contains Ω and is closed wrt. complements and finite unions. Is a *σ–algebra* if it is also closed under countable unions.

The pair $(\Omega, \mathscr{F})$ where Ω is a set and $\mathscr{A}$ a σ-algebra of subsets of Ω is a *measurable space*.

A (positive) *measure* μ on the measurable space $(\Omega, \mathscr{F})$ is a map $\mu: \mathscr{F} \to [0, \infty]$ such that $\mu(\emptyset) = 0$ and

$$\mu(\cup_n A_n) = \sum_n \mu(A_n),$$

for any countable $(A_n)_n$ family of disjoint elements of $\mathscr{F}$. The measure μ is finite if $\mu(\Omega) < \infty$ and σ-finite if there exists $(\Omega_n)_n \subseteq \mathscr{A}$ such that $\mu(\Omega_n) < \infty$ for all n and $\Omega = \cup_n \Omega_n$. A triple $(\Omega, \mathscr{F}, \mu)$ is called a measure space.

A probability space is a measure space $(\Omega, \mathscr{F}, \mathbb{P})$ such that $\mathbb{P}(\Omega) = 1$, the measure $\mathbb{P}$ is then called a probability measure. The set Ω is called the set of elementary events and $\mathscr{F}$ is the σ-algebra of all the events.

In general, given a family $\mathscr{G}$ of subsets of Ω we can consider the smallest σ-algebra containing $\mathscr{G}$ and we will denote it with $\sigma(\mathscr{G})$.

If Ω is a topological space then we consider $\mathscr{B}(\Omega)$ the Borel σ-algebra, which is the smallest σ-algebra of Ω containing all open sets.

1 Carathéodory's construction of a measure

To construct a measure one has to overcome the sheer complexity of the structure of a σ-algebra.

In order to *describe* a measure is useful to start with a simpler object, namely a σ-additive (positive) set-function which is a function $\nu: \mathscr{U} \to [0, \infty]$ defined over an arbitrary family $\mathscr{U}$ of subsets of Ω satisfying $\nu(\emptyset) = 0$ and $\nu(\cup_k A_k) = \sum_k \nu(A_k)$ for all countable families $(A_k)_k \subseteq \mathscr{U}$ such that $\cup_k A_k \in \mathscr{U}$ and made of mutually disjoint sets.

Theorem 1. (Carathéodory) *Let Ω be a set, $\mathscr{U}$ an algebra of subsets of Ω and μ_0 a positive σ-additive set-function on $\mathscr{U}$. Then there exists a measure μ on $\sigma(\mathscr{U})$ such that $\mu|_{\mathscr{U}} = \mu_0$. If μ_0 is σ-finite then μ is unique.*

Let us introduce some useful intermediate object to the construction.

A map $\mu^*: \mathscr{P}(\Omega) \to [0, \infty]$ is an *outer measure* iff $\mu^*(\emptyset) = 0$, μ^* is non-decreasing for the inclusion order and μ^* is σ-subadditive, i.e. $\mu^*(\cup_k A_k) \leqslant \sum_k \mu^*(A_k)$ for all countable families $(A_k)_k \subseteq \mathscr{P}(\Omega)$.

If μ^* is an outer measure, then a set $B \subseteq \Omega$ is μ^*-measurable iff $\mu^*(A) = \mu^*(B \cap A) + \mu^*(B^c \cap A)$ for all $A \subseteq \Omega$. We let $\mathscr{M}(\mu^*)$ the family of all measurable subsets of μ^*.

Theorem 2.

i. $\mathscr{M}(\mu^*)$ is a σ-algebra that contains all $B\subseteq\mathscr{P}(\Omega)$ such that $\mu^*(B)=0$;

ii. The restriction μ of μ^* to $\mathscr{M}(\mu^*)$ is a measure.

Proof. By subadditivity we have always $\mu^*(A)\leqslant\mu^*(B\cap A)+\mu^*(B^c\cap A)$, so if $\mu^*(B)=0$ then $\mu^*(A)\leqslant\mu^*(B^c\cap A)\leqslant\mu^*(A)$ and $B\in\mathscr{M}(\mu^*)$. Moreover if $B\in\mathscr{M}(\mu^*)$ then by symmetry of the definition also $B^c\in\mathscr{M}(\mu^*)$. Assume $B_1,B_2\in\mathscr{M}(\mu^*)$ then $\mu^*(A)=\mu^*(B_1\cap A)+\mu^*(B_1^c\cap A)=\mu^*(B_1\cap B_2\cap A)+\mu^*(B_1\cap B_2^c\cap A)+\mu^*(B_1^c\cap B_2\cap A)+\mu^*(B_1^c\cap B_2^c\cap A)\geqslant\mu^*((B_1\cap B_2)\cap A)+\mu^*((B_1\cap B_2)^c\cap A)$ since $(B_1\cap B_2)^c=(B_1\cap B_2^c)\cup(B_1^c\cap B_2)\cup(B_1^c\cap B_2^c)$ and subadditivity. Therefore $B_1\cap B_2\in\mathscr{M}(\mu^*)$ and by stability under complement we have also that $\mathscr{M}(\mu^*)$ is stable under finite unions. Let now $(B_k)_{k\geqslant1}$ a countable family in $\mathscr{M}(\mu^*)$ made of disjoint sets. Let $\bar{B}_n=\cup_{k\geqslant n}B_k$. Then

$$\mu^*(A)=\mu^*((B_1\cup\cdots\cup B_n)\cap A)+\mu^*((B_1\cup\cdots\cup B_n)^c\cap A)$$

$$\geqslant\mu^*((B_1\cup\cdots\cup B_n)\cap A)+\mu^*((\cup_{k\geqslant1}B_k)^c\cap A)=\sum_{k=1}^{n}\mu^*(B_k\cap A)+\mu^*((\cup_{k\geqslant1}B_k)^c\cap A).$$

Taking $n\to\infty$ we obtain, by σ-subadditivity,

$$\mu^*(A)\geqslant\sum_{k=1}^{\infty}\mu^*(B_k\cap A)+\mu^*((\cup_{k\geqslant1}B_k)^c\cap A)\geqslant\mu^*((\cup_{k\geqslant1}B_k)\cap A)+\mu^*((\cup_{k\geqslant1}B_k)^c\cap A),$$

while the reverse inequality holds also for σ-subadditivity. Then $(\cup_{k\geqslant1}B_k)\in\mathscr{M}(\mu^*)$.

In order to prove that the restriction of μ^* to $\mathscr{M}(\mu^*)$ is a measure one simply observes that, if $(B_k)_k\subseteq\mathscr{M}(\mu^*)$ is a countable family made of mutually disjoint sets

$$\mu^*(\cup_{k\geqslant1}B_k)=\mu^*((\cup_{k\geqslant1}B_k)\cap B_1)+\mu^*((\cup_{k\geqslant1}B_k)\cap B_1^c)=\mu^*(B_1)+\mu^*(\cup_{k\geqslant2}B_k)$$

therefore $\mu^*(\cup_{k\geqslant1}B_k)\geqslant\sum_k\mu^*(B_k)$, so the equality holds by σ-subadditivity. Note that $\mu^*(\emptyset)=0$, so the claim is proved. □

Lemma 3. *Let $\mathscr{U}$ be an algebra of subsets of Ω and μ a σ-additive set function on $\mathscr{U}$, define μ^*: $\mathscr{P}(\Omega)\to[0,\infty]$ by*

$$A\in\mathscr{P}(\Omega)\mapsto\mu^*(A):=\inf\left\{\sum_k\mu(F_k):(F_k)_k\subseteq\mathscr{U}\text{ and }A\subseteq\cup_kF_k\right\}$$

Then μ^* is an outer measure. Moreover $\mathscr{U}\subseteq\mathscr{M}(\mu^*)$.

Proof. Is clear that $\mu^*(\emptyset)=0$ and that μ^* is non-decreasing. Let $(A_k)_k$ be a countable family of sets in Ω, then find families $(F_{k,\ell})$ such that $A_k\subseteq\cup_\ell F_{k,\ell}$ and $\mu^*(A_k)\geqslant\sum_\ell\mu(F_{k,\ell})+\varepsilon2^{-k}$. Then $\cup_kA_k\subseteq\cup_{k,\ell}F_{k,\ell}$ and

$$\mu^*(\cup_kA_k)\leqslant\sum_k\mu(\cup_\ell F_{k,\ell})\leqslant\sum_{k,\ell}\mu(F_{k,\ell})\leqslant\sum_k\mu^*(A_k)-\varepsilon\sum_k2^{-k}$$

and taking $\varepsilon \to 0$ we obtain the σ-subadditivity of μ^*, so μ^* is an outer measure. Let us show now that $\mathscr{U} \subseteq \mathscr{M}(\mu^*)$. Let $A \subset \Omega$, and consider again $(F_k)_k \subseteq \mathscr{U}$ such that $\mu^*(A) \geqslant \sum_k \mu(F_k) + \varepsilon$. Now for all $B \in \mathscr{U}$ we have

$$\mu^*(A \cap B) + \mu^*(A \cap B^c) \leqslant \sum_k (\mu(F_k \cap B) + \mu(F_k \cap B^c)) = \sum_k \mu(F_k) \leqslant \mu^*(A) - \varepsilon$$

and taking $\varepsilon \to 0$ we have that $B \in \mathscr{M}(\mu^*)$. □

We need now also a characterization for measures which is conveniently provided by Dynkin's $\pi - \lambda$ theorem. We say that a family Λ of subsets of Ω is a λ-system if it contains $\emptyset$, is closed under complement and countable disjoint unions. Alternatively a λ-system can be characterized by saying that it contains Ω, is stable under differences (i.e. if $A \subset B$ and $A, B \in \Lambda$ then $A \backslash B \in \Lambda$) and under increasing limits (i.e. $(A_k)_k \subseteq \Lambda$ with $A_k \subseteq A_{k+1}$, then $\lim_k A_k \in \Lambda$). We say that another family Π of subsets of Ω is a π-system if it is closed under finite intersection.

Theorem 4. (Dynkin's $\boldsymbol{\pi - \lambda}$ theorem) *If Π is a π-system and Λ a λ-system then $\Pi \subseteq \Lambda$ implies that $\sigma(\Pi) \subseteq \Lambda$.*

Proof. (of Theorem 1) Existence of the required extension is an immediate consequence of Lemma 3 and Theorem 2. Dynkin's theorem allows to prove that the measure $\bar{\mu}$ obtained by restriction of μ^* to $\sigma(\mathscr{U})$ is the unique extension of μ from $\mathscr{U}$ to $\sigma(\mathscr{U})$ if μ is σ-finite. Let us describe the argument for μ a finite measure. It is left then to the reader to generalize to σ-finiteness. Assume that $\hat{\mu}$ is another such extension which coincide with μ on $\mathscr{U}$. Let Λ be the set of elements $B \in \sigma(\mathscr{U})$ such that $\bar{\mu}(B) = \hat{\mu}(B)$. Then Λ is a λ-system since $\{\emptyset, \Omega\} \subseteq \mathscr{U} \subseteq \Lambda$, $\bar{\mu}(B_1^c) = \bar{\mu}(\Omega) - \bar{\mu}(B_1) = \hat{\mu}(\Omega) - \hat{\mu}(B_1) = \hat{\mu}(B_1^c)$, and if $(B_k)_k \subseteq \Lambda$ is a pairwise disjoint family we have $\bar{\mu}(\cup_k B_k) = \sum_k \bar{\mu}(B_k) = \sum_k \hat{\mu}(B_k) = \hat{\mu}(\cup_k B_k)$. Also $\mathscr{U}$ is a π-system, so by Dynkin's theorem we have that $\sigma(\mathscr{U}) \subseteq \Lambda$ but then $\Lambda = \sigma(\mathscr{U})$ and therefore $\bar{\mu}$ and $\hat{\mu}$ coincide on $\sigma(\mathscr{U})$. □

Remark 5. Lebesgue measure on $\mathbb{R}$ can be constructed starting from the additive set function $\lambda: \mathscr{U} \to [0, \infty]$ defined on the family $\mathscr{U}$ of sets obtained via finite unions of intervals of the form $(a, b]$, $(-\infty, b]$, $(a, +\infty) \subset \mathbb{R}$ by letting $\lambda(\cup_k (a_k, b_k]) = \sum_k (b_k - a_k)$. One can prove that $\mathscr{U}$ is an algebra and λ a σ-additive, σ-finite set function. Then its extension λ^* to $\mathscr{M}(\lambda^*)$ defines Lebesgue measure on $\sigma(\mathscr{U}) = \mathscr{B}(\Omega)$ the Borel σ-field of $\mathbb{R}$. Extension of this construction to higher dimension is straightforward.

Remark 6. About the necessity of σ-finiteness for uniqueness. The set-function ν defined on $\mathscr{U}$ (see previous Remark) by $\nu((a_k, b_k]) = +\infty$ has an extension to $\sigma(\mathscr{U})$ which is always infinite, however this extension is not unique since for example the counting measure $\hat{\nu}$ (i.e. the measure which assigns to a set B its cardinality) has the same restriction to $\mathscr{U}$. Later on we will see that product measures also provide another counterexample.

Theorem 7. *A probability measure $\mathbb{P}: \mathscr{B}(\Omega) \to [0, 1]$ on a compact Hausdorff metrizable space Ω is inner regular (or tight), that is $\mathbb{P}(F) = \sup_{K \subseteq F} \mathbb{P}(K)$ where K runs over all the compacts K contained in F.*

Proof. Let $\mathscr{U}$ the subfamily of $\mathscr{B}(\Omega)$ made of sets B such that, for any $\varepsilon > 0$, there exists a compact $K \subseteq B$ and an open set $O \supseteq B$ such that $\mathbb{P}(B \backslash K), \mathbb{P}(O \backslash B) \leqslant \varepsilon$. Is not difficult to prove that $\mathscr{U}$ is a σ-algebra on general grounds (exercise). We want now to prove that $\mathscr{U} = \mathscr{B}(\Omega)$. Here we use the assumption that Ω is metrizable and let ρ a metric which generates the topology of Ω. For any compact K let $G_n = \{\omega \in \Omega: \rho(\omega, K) < 1/n\} \in \mathscr{B}(\Omega)$ which is open, decreasing in n and converging to K, in the sense that $K = \cap_n G_n$. Then $\mathbb{P}(G_n) \downarrow \mathbb{P}(K)$ therefore $K \in \mathscr{U}$. By the compactness of Ω we have also that any closed set B is compact, therefore $B \in \mathscr{U}$. Since all closed sets are in $\mathscr{U}$ and $\mathscr{U}$ is a σ-algebra, we conclude that $\mathscr{B}(\Omega) \subseteq \mathscr{U}$ but this implies that $\mathscr{U} = \mathscr{B}(\Omega)$ since the other inclusion is trivial. Now for any $F \in \mathscr{B}(\Omega)$ we have a compact $K \subseteq F$ such that $\mathbb{P}(F) - \mathbb{P}(K) = \mathbb{P}(F \backslash K) \leqslant \varepsilon$ therefore $\mathbb{P}(F) = \sup_{K \subseteq F} \mathbb{P}(K)$ as claimed. □

(recall that on Hausdorff spaces compact sets are closed and that closed sets in a compact topological space are compact, and that a space is compact iff every open cover has a finite sub-cover)

2 Random variables and integrals

A function $f: (\Omega, \mathscr{F}) \to (E, \mathscr{E})$ between two measurable spaces is *measurable* iff $f^{-1}(A) \in \mathscr{F}$ for all $A \in \mathscr{E}$.

The interest of this definition is that given such a measurable function and a measure μ on $(\Omega, \mathscr{F})$ we can construct the *induced* measure μ_f on $(E, \mathscr{E})$ by letting $\mu_f(A) = \mu(f^{-1}(A))$ for all $A \in \mathscr{E}$.

A (real valued) *random variable* on $(\Omega, \mathscr{F})$ is a measurable function from this measure space to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. If $(E, \mathscr{E})$ is another measure space then an E-valued random variable is a measurable map from $(\Omega, \mathscr{F})$ to $(E, \mathscr{E})$. If E is a topological space, usually the Borel σ-algebra $\mathscr{B}(E)$ is used and the random variable is called *E-valued*. If E is a metric space then the random variable is called *Borel*.

Given a random variable $f: (\Omega, \mathscr{F}) \to (E, \mathscr{E})$ we call $\sigma(f)$ the smallest sub-σ-algebra of $\mathscr{F}$ which still makes $f: (\Omega, \sigma(f)) \to (E, \mathscr{E})$ measurable.

Measurable functions are stable under upper and lower limits, therefore if $f_n \to f$ and each f_n is measurable, then also the limiting f is measurable.

Given a family $\mathscr{U}$ of subsets of Ω we would like to characterize all the functions which are $\sigma(\mathscr{U})$ measurable. This is the purpose of the following theorem:

Theorem 8. **(Monotone class theorem)** *Let $\mathscr{H}$ be a class of bounded functions on Ω to $\mathbb{R}$ such that*

i. *$\mathscr{H}$ is a vector space over $\mathbb{R}$,*

ii. $1 \in \mathscr{H}$,

iii. *if $f_n \geqslant 0$ and $f_n \uparrow f$ with f bounded, then $f \in \mathscr{H}$.*

Then if $\mathscr{H}$ contains the indicator functions of every element of a π-system $\mathscr{U}$ then $\mathscr{H}$ contains every bounded $\sigma(\mathscr{U})$-measurable function.

Proof. Consider the set Λ of subsets B of Ω such that $\mathbb{1}_B \in \mathscr{H}$. This is a λ-system (check). Then by Dynkin's theorem $\mathscr{H}$ contains the indicator function of every element in $\sigma(\mathscr{U})$. Let now $f \geqslant 0$ be a bounded $\sigma(\mathscr{U})$ measurable positive function and let K an upper bound for f. Define functions f_n as

$$f_n = \sum_{k=0}^{\lceil K \rceil 2^n - 1} k 2^{-n} \mathbb{1}_{\{k 2^{-n} \leqslant f \leqslant (k+1) 2^{-n}\}}.$$

Note that $\{k2^{-n} \leqslant f \leqslant (k+1)2^{-n}\} \in \sigma(\mathscr{U})$, $f_n \leqslant f$, $f_n \uparrow f$ and $f_n \in \mathscr{H}$. Therefore we conclude $f \in \mathscr{H}$. In order to deal with general f we decompose it into its positive and negative parts and reason as above. □

In order to define integrals we need the notion of simple function. A function $f: (\Omega, \mathscr{F}) \to \mathbb{R}$ is *simple* if can be written as

$$f = \sum_i w_i \mathbb{1}_{A_i},$$

where $(A_i)_i \subset \mathscr{F}$ is a partition of Ω and $(w_i)_i \subset \mathbb{R}$ are distinct numbers. The space of simple functions (on the measure space $(\Omega, \mathscr{F})$) is denoted $\mathscr{E}$ and we denote by $\mathscr{E}_+ \subseteq \mathscr{E}$ the subspace of positive simple functions.

Simple functions are of course measurable. Given a measure μ on $(\Omega, \mathscr{F})$ we can define the *integral* of $f = \sum_i w_i \mathbb{1}_{A_i} \in \mathscr{E}$ wrt. μ as

$$\int_\Omega f \mathrm{d}\mu = \sum_i w_i \mu(A_i),$$

provided $f \in \mathscr{E}_+$ or provided the measure μ is finite. For general measurable functions $f: \Omega \to \mathbb{R}$ we define the integral as follows. If $f \geqslant 0$ then

$$\int_\Omega f \mathrm{d}\mu = \sup_{g \in \mathscr{E}_+, g \leqslant f} \int_\Omega g \mathrm{d}\mu,$$

otherwise we let $f = f_+ - f_-$ where $f_+ = \max(f, 0)$ and $f_- = \max(-f, 0)$ and we let

$$\int_\Omega f \mathrm{d}\mu = \int_\Omega f_+ \mathrm{d}\mu - \int_\Omega f_- \mathrm{d}\mu,$$

provided either of the integrals in the r.h.s. is finite and leave the integral undefined otherwise.

A function f is *absolutely integrable* if $\int_\Omega |f| \mathrm{d}\mu < +\infty$. (in probability theory they are usually called simply *integrable*).

The integral is linear, and monotone. The basic convergence results for the (Lebesgue) integral are (without proofs)

Theorem 9.

i. (Monotone convergence) If $(f_n)_n$ is an increasing sequences of measurable non-negative functions such that $f_n \uparrow f$. Then

$$\lim_n \int_\Omega f_n \mathrm{d}\mu = \int_\Omega f \mathrm{d}\mu.$$

ii. (Fatou's lemma) If $(f_n)_n$ is a sequence of measurable non-negative functions, then

$$\liminf_n \int_\Omega f_n \mathrm{d}\mu \geqslant \int_\Omega \left(\liminf_n f_n\right) \mathrm{d}\mu.$$

iii. *(Lebesgue's dominated convergence) Let $(f_n)_n$ be a sequence of absolutely integrable function, such that $f_n \to f$ and let g another absolutely integrable function such that $|f_n(\omega)| \leqslant g(\omega)$ for μ-almost all ω and for all n. Then*

$$\lim_n \int_\Omega f_n \mathrm{d}\mu = \int_\Omega f \mathrm{d}\mu.$$

Using monotone convergence is possible to prove a more intuitive representation of the integral of a non-negative measurable function f, namely that

$$\int f \mathrm{d}\mu = \lim_n \left[\sum_{k=0}^{n2^n-1} k2^{-n} \mu(k2^{-n} \leqslant f < (k+1)2^{-n}) + n\mu(f \geqslant n) \right].$$

(this was actually the original definition given by Lebesgue).

When dealing with a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ we denote with $\mathbb{E}$ the corresponding integral, namely if $X: \Omega \to \mathbb{R}$ is a random variable, then

$$\mathbb{E}[X] = \int_\Omega X \mathrm{d}\mathbb{P} = \int_\Omega X(\omega) \mathbb{P}(\mathrm{d}\omega),$$

where sometimes we will use the notation on the r.h.s. in order to highlight the role of of the "integration" variable.

An important concept related to integration, is the following.

A family $(X_\alpha)_\alpha$ of random variables is *uniformly integrable* if for any $\varepsilon > 0$ there exists $L > 0$ such that

$$\sup_\alpha \mathbb{E}[|X_\alpha| \mathbb{1}_{|X_\alpha| > L}] < \varepsilon.$$

In particular, it holds that $\sup_\alpha \mathbb{E}[|X_\alpha|] < \infty$.

A single random variable is uniformly integrable due to the monotone convergence theorem (exercise). A finite family of random variable is also easily seen to be uniformly integrable.

An alternative characterization of uniform integrability is given by:

Lemma 10. *A family $(X_\alpha)_\alpha$ of random variables is uniformly integrable iff $\sup_\alpha \mathbb{E}[|X_\alpha|] < \infty$ and for all $\varepsilon > 0$ there exists $\delta > 0$ such that, for all $A \in \mathscr{F}$ for which $\mathbb{P}(A) < \delta$ we have $\sup_\alpha \mathbb{E}[|X_\alpha| \mathbb{1}_A] < \varepsilon$.*

Proof. Note that, for all $K > 0$, $|X_\alpha| \leqslant (|X_\alpha| - K)_+ + K$ therefore, if the family is uniformly integrable we have

$$\mathbb{E}[|X_\alpha| \mathbb{1}_A] \leqslant \mathbb{E}[(|X_\alpha| - K)_+ \mathbb{1}_A] + K\mathbb{P}(A) \leqslant \mathbb{E}[(|X_\alpha| - K) \mathbb{1}_{|X_\alpha| > K}] + K\mathbb{P}(A) \leqslant \varepsilon/2 + K\mathbb{P}(A)$$

by choosing $K = K(\varepsilon)$ appropriately, then is enough to take δ small enough. The reverse implication follows observing that $K\mathbb{P}(|X_\alpha| > K) \leqslant \sup_\alpha \mathbb{E}[|X_\alpha|] < \infty$ therefore by choosing $K = K(\varepsilon)$ large enough we have $\sup_\alpha \mathbb{P}(|X_\alpha| > K) < \delta(\varepsilon)$ and as a consequence $\sup_\alpha \mathbb{E}[|X_\alpha| \mathbb{1}_{|X_\alpha| > K}] < \varepsilon$. □

Uniform integrability is the best possible condition for the convergence of integrals, as the following theorem shows. In this respect it more general that Lebesgue's dominated convergence theorem.

Theorem 11. (Uniform integrability) *Let* $(X_n)_n$ *and* X *be integrable random variables, then* $\mathbb{E}[|X_n - X|] \to 0$ *(convergence in average) iff*

a) $X_n \to X$ *in probability, i.e.* $\lim_n \mathbb{P}(|X_n - X| > \varepsilon) = 0$ *for all* $\varepsilon > 0$*;*

b) the family $(X_n)_n$ *is uniformly integrable.*

Proof. Let us show the reverse implication. Let $\phi_K(x) = (K \wedge x) \vee (-K)$ and observe that $|\phi_K(x) - y| \leqslant |x - y|$ so if $X_n \to X$ in probability, then also $\phi_K(X_n) \to \phi_K(X)$ in probability. Now by uniform integrability we have

$$\mathbb{E}[|\phi_K(X_n) - X_n|] \leqslant \mathbb{E}[(|X_n| - K)\mathbb{1}_{|X_n| > K}] \leqslant \varepsilon / 2$$

and a similar statement for X, provided that $K = K(\varepsilon)$ is suitably chosen. On the other hand

$$\mathbb{E}[|\phi_K(X_n) - \phi_K(X)|] = \mathbb{E}[|\phi_K(X_n) - \phi_K(X)|\mathbb{1}_{|X_n - X| \leqslant \varepsilon}] + \mathbb{E}[|\phi_K(X_n) - \phi_K(X)|\mathbb{1}_{|X_n - X| > \varepsilon}]$$

$$\leqslant \varepsilon \mathbb{P}(|X_n - X| \leqslant \delta) + 2K\mathbb{P}(|X_n - X| > \delta) \leqslant \varepsilon + 2K\mathbb{P}(|X_n - X| > \varepsilon) \to \varepsilon,$$

therefore we conclude that $\lim_n \mathbb{E}[|X_n - X|] \leqslant 3\varepsilon$ and since ε is arbitrary $\mathbb{E}[|X_n - X|] \to 0$ follows. In order to deduce the direct implication, note that convergence in average implies easily convergence in probability, moreover that we need to have $\mathbb{E}[|X_n - X|] < \varepsilon / 2$ eventually, that is for $n \geqslant n_0$ for some n_0. The finite family $(X_n)_{n < n_0}$ is easily seen uniformly integrable, while if $n \geqslant n_0$ we have

$$\mathbb{E}[|X_n|\mathbb{1}_{|X_n| > K}] \leqslant \mathbb{E}[|X_n - X|\mathbb{1}_{|X_n| > K}] + \mathbb{E}[|X|\mathbb{1}_{|X_n| > K}] \leqslant \varepsilon / 2 + \mathbb{E}[|X|\mathbb{1}_{|X_n| > K}]$$

moreover, Markov's inequality gives

$$\mathbb{P}(|X_n| > K) \leqslant K^{-1}\mathbb{E}[|X_n|] \leqslant K^{-1}(\mathbb{E}[|X|] + \mathbb{E}[|X - X_n|]) \leqslant K^{-1}(\mathbb{E}[|X|] + \varepsilon / 2).$$

Therefore if K is chosen big enough we have $\mathbb{P}(|X_n| > K) < \delta = \delta(\varepsilon)$ and uniform integrability of X allows to conclude that $\mathbb{E}[|X_n|\mathbb{1}_{|X_n| > K}] \leqslant \varepsilon$ also for $n \geqslant n_0$. □

3 L^p spaces

Given a measure space $(\Omega, \mathscr{F}, \mu)$ one can introduce a family of semi-norms on measurable functions indexed by $p \geqslant 1$:

$$\|f\|_p := \left(\int_\Omega |f|^p \mathrm{d}\mu \right)^{1/p}.$$

That this function satisfies the triangle inequality is the content of *Minkowski's inequality*:

$$\|f + g\|_p \leqslant \|f\|_p + \|g\|_p.$$

Moreover, Hölder's inequality holds:

$$\left| \int_\Omega fg \mathrm{d}\mu \right| \leqslant \|f\|_p \|g\|_q$$

for $1 \leqslant p, q \leqslant \infty$ such that $p^{-1} + q^{-1} = 1$, where $\|f\|_\infty = \text{esssup}\,|f| = \inf\{L \geqslant 0 \colon \mu(|f| > L) = 0\}$.

Both inequalities can be deduced from Jensen's inequality (in the case of probability measures):

Lemma 12. (Jensen's inequality) *Let $(\Omega, \mathscr{F}, \mathbb{P})$ a probability space, X an absolutely integrable random variable and $\varphi \colon \mathbb{R} \to \mathbb{R}$ a convex function, then*

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

Proof. Since φ is convex, for any $z \in \mathbb{R}$ there exists $m \in \mathbb{R}$ such that $\varphi(x) \geqslant m(x - z) + \varphi(z)$. Take $z = \mathbb{E}[X]$, then

$$\mathbb{E}[\varphi(X)] \leqslant \mathbb{E}[m(X - \mathbb{E}[X]) + \varphi(\mathbb{E}[X])] = \varphi(\mathbb{E}[X]). \quad \square$$

For all $p \in [1, \infty]$, the space of all function f such that $\|f\|_p < \infty$ is called $\mathscr{L}^p = \mathscr{L}^p(\Omega, \mathscr{F}, \mu)$. On this linear space $\|\cdot\|_p$ is only a semi-norm, since there exists functions $f \neq 0$ such that $\|f\|_p = 0$. Introducing the equivalence relation $f \sim f'$ iff $\|f - f'\|_p = 0$ we can consider the set $L^p = L^p(\Omega, \mathscr{F}, \mu)$ of equivalence classes of functions in $\mathscr{L}^p$ modulo functions which are non-zero on sets of measure zero. On L^p the function $\|*\|_p$ is a norm. Moreover if $p = 2$ then we have also $|\int_\Omega fg \mathrm{d}\mu| \leqslant \|f\|_2 \|g\|_2$, therefore L^2 is an Hilbert space when endowed with the scalar product $\langle f, g \rangle = \int_\Omega fg \mathrm{d}\mu$. This is fully justified by the following completeness result.

Theorem 13. *The spaces L^p are Banach spaces.*

Proof. We need to show completeness. Let $(f_n)_n$ be a Cauchy sequence in L^p. We can choose $(n_k)_k$ increasing such that for all $i, j \geqslant n_k$ we have $\|f_i - f_j\| \leqslant 2^{-k-k/p}$. Now let $F \coloneqq \sum_k 2^{kp} |f_{n_{k+1}} - f_{n_k}|^p$ and observe that, on the one hand

$$\int F \mathrm{d}\mu = \sum_k 2^{kp} \int |f_{n_{k+1}} - f_{n_k}|^p \mathrm{d}\mu \leqslant \sum_k 2^{-k} < \infty,$$

so F is μ-almost everywhere finite, while on the other hand, $|f_{n_{k+1}} - f_{n_k}| \leqslant 2^{-k} F^{1/p}$ which implies that $(f_{n_k})_k$ is a Cauchy sequence everywhere F is finite. Let $f = \lim_k f_{n_k}$ if $F < \infty$ and $f = 0$ when $F = \infty$. Observe that

$$|f - f_{n_k}| \leqslant \sum_{m \geqslant k} |f_{n_{m+1}} - f_{n_m}| \leqslant 2^{-k} F^{1/p}$$

therefore $\|f - f_{n_k}\|_p \to 0$ and as a consequence $f \in L^p$ and $f_n \to f$ in L^p. $\square$

4 Product measures and integrals

Given two σ-finite measurable spaces $(\Omega_i, \mathscr{F}_i, \mu_i)_{i=1,2}$ we can consider their product $(\Omega, \mathscr{F}, \mu)$ where $\Omega = \Omega_1 \times \Omega_2$ (product of sets), $\mathscr{F} = \mathscr{F}_1 \otimes \mathscr{F}_2$ is the σ-algebra on Ω generated by the family $\mathscr{U}$ of sets of the form $A \times B$ with $A \in \mathscr{F}_1$ and $B \in \mathscr{F}_2$ and $\mu = \mu_1 \otimes \mu_2$ is the measure defined by $\mu(A \times B) = \mu_1(A) \mu_2(B)$ on $\mathscr{U}$ (with $0 \cdot \infty = 0$). Existence and uniqueness of such a measure follows from Caratheodory's extension theorem.

On product spaces, sections are measurable.

Proposition 14. *If* $A\in\mathscr{F}_1\otimes\mathscr{F}_2$ *and we let* $A_x := \{y\in\Omega_2 : (x,y)\in A\}$ *then we have* $A_x\in\mathscr{F}_2$.

Proof. Let $\mathscr{G}$ the family of sets A for which $A_x\in\mathscr{F}_1$. Clearly measurable rectangles are in $\mathscr{G}$ and one observes that $\mathscr{G}$ is a σ-algebra. Therefore $\mathscr{G}\subseteq\mathscr{F}_1\otimes\mathscr{F}_2\subseteq\mathscr{G}$. □

Remark 15. We have $\mathscr{B}(\mathbb{R}^n)\otimes\mathscr{B}(\mathbb{R}^m)=\mathscr{B}(\mathbb{R}^{n+m})$. Recall that the Borel σ-algebra of $\mathbb{R}^n$ can be generated by rectangles.

In this setting we have the following results about integration with respect to the product measure $\mu_1\otimes\mu_2$.

Theorem 16. (Fubini-Tonelli) *If* $f:\Omega\to\mathbb{R}$ *is a non-negative measurable function we have*

$$\int_{\Omega_1\times\Omega_2} f(x,y)(\mu_1\otimes\mu_2)(\mathrm{d}x\mathrm{d}y)=\int_{\Omega_1} f_1(x)\mu_1(\mathrm{d}x)=\int_{\Omega_2} f_2(y)\mu_2(\mathrm{d}y) \tag{1}$$

where

$$f_1(x):=\int_{\Omega_2} f(x,y)\mu_2(\mathrm{d}y),\quad f_2(y):=\int_{\Omega_1} f(x,y)\mu_1(\mathrm{d}x),$$

are functions which are measurable wrt. $\mathscr{F}_1$ *and* $\mathscr{F}_2$ *respectively.*

If $f:\Omega\to\mathbb{R}$ *is a* μ*–absolutely integrable function, then* f *is absolutely integrable wrt. to each variable separately,* f_1, f_2 *defined as above are well defined, except possibly for a set of measure zero (wrt.* μ_1 *resp.* μ_2*) and the equality of integrals in (1) holds.*

Remark 17. Again σ-finiteness is a necessary condition to be able to identify the product measure $\mu_1\otimes\mu_2$ uniquely using the condition that $(\mu_1\otimes\mu_2)(A\times B)=\mu_1(A)\mu_2(B)$ for $A\in\mathscr{F}_1$ and $B\in\mathscr{F}_2$. Indeed consider $\Omega_1=\Omega_2=[0,1]$ with the Borel σ-algebra and μ_1 given by the Lebesgue measure while μ_2 given by the counting measure. Note that the set $D=\{(x,y): 0\leqslant x=y\leqslant 1\}\subseteq[0,1]\times[0,1]$ is measurable wrt. $\mathscr{B}([0,1])\otimes\mathscr{B}([0,1])=\mathscr{B}([0,1]^2)$ (this equality holds since $[0,1]^2$ is separable and we can approximate open sets with balls and balls with rectangles). In this case there are many possible product measures since one can take for example the measures ν_1,ν_2 such that $\nu_1(F)=\int_{\Omega_2}\mu_1(\{x\in\Omega_1:(x,y)\in F\})\mu_2(\mathrm{d}y)$ and $\nu_2(F)=\int_{\Omega_1}\mu_2(\{y\in\Omega_1:(x,y)\in F\})\mu_1(\mathrm{d}x)$ (one would have to prove that such sections are measurable, of course). Both measures are product measures but they are clearly different since $\nu_1(D)=0$ and $\nu_2(D)=1$. The extension ν_C given by Carathéodory's theorem (starting from a σ-additive set function on the algebra of sets generated by measurable rectangles) satisfies $\nu_C(D)=+\infty$ since any finite cover of D by elements of the algebra necessarily have infinite measure.