Jan van Neerven

Measure, Integral, and Conditional Expectation

Handout

May 2, 2012

1

Measure and Integral

In this lecture we review elements of the Lebesgue theory of integration.

1.1 σ-Algebras

Let Ω be a set. A *σ-algebra* in Ω is a collection $\mathscr{F}$ of subsets of Ω with the following properties:

(1) $\Omega \in \mathscr{F}$;
(2) $A \in \mathcal{F}$ implies $\complement A \in \mathscr{F}$;
(3) $A_1, A_2, \dots \in \mathscr{F}$ implies $\bigcup_{n=1}^{\infty} A_n \in \mathscr{F}$.

Here, $\complement A = \Omega \setminus A$ is the complement of A.

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

The above three properties express that $\mathscr{F}$ is non-empty, closed under taking complements, and closed under taking countable unions. From

$$\bigcap_{n=1}^{\infty} A_n = \complement\Big(\bigcup_{n=1}^{\infty} \complement A_n\Big)$$

it follows that $\mathscr{F}$ is closed under taking countable intersections. Clearly, $\mathscr{F}$ is closed under finite unions and intersections as well.

Example 1.1. In every set Ω there is a minimal σ-algebra, $\{\varnothing, \Omega\}$, and a maximal σ-algebra, the set $\mathscr{P}(\Omega)$ of all subsets of Ω.

Example 1.2. If $\mathscr{A}$ is any collection of subsets of Ω, then there is a smallest σ-algebra, denoted $\sigma(\mathscr{A})$, containing $\mathscr{A}$. This σ-algebra is obtained as the intersection of all σ-algebras containing $\mathscr{A}$ and is called the *σ-algebra generated by $\mathscr{A}$*.

Example 1.3. Let $(\Omega_1, \mathscr{F}_1)$, $(\Omega_2, \mathscr{F}_2)$, ... be a sequence of measurable spaces. The *product*

$$(\prod_{n=1}^{\infty} \Omega_n, \prod_{n=1}^{\infty} \mathscr{F}_n)$$

is defined by $\prod_{n=1}^{\infty} \Omega_n = \Omega_1 \times \Omega_2 \times \ldots$ (the cartesian product) and $\prod_{n=1}^{\infty} \mathscr{F}_n = \mathscr{F}_1 \times \mathscr{F}_2 \times \ldots$ is defined as the σ-algebra generated by all sets of the form

$$A_1 \times \cdots \times A_N \times \Omega_{N+1} \times \Omega_{N+2} \times \ldots$$

with $N = 1, 2, \ldots$ and $A_n \in \mathscr{F}_n$ for $n = 1, \ldots, N$. Finite products $\prod_{n=1}^{N} (\Omega_n, \mathscr{F}_n)$ are defined similarly; here one takes the σ-algebra in $\Omega_1 \times \cdots \times \Omega_N$ generated by all sets of the form $A_1 \times \cdots \times A_N$ with $A_n \in \mathscr{F}_n$ for $n = 1, \ldots, N$.

Example 1.4. The *Borel σ-algebra* of $\mathbb{R}^d$, notation $\mathscr{B}(\mathbb{R}^d)$, is the σ-algebra generated by all open sets in $\mathbb{R}^d$. The sets in this σ-algebra are called the *Borel sets* of $\mathbb{R}^d$. As an exercise, check that

$$(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d)) = \prod_{n=1}^{d} (\mathbb{R}, \mathscr{B}(\mathbb{R})).$$

Hint: Every open set in $\mathbb{R}^d$ is a countable union of rectangles of the form $[a_1, b_1) \times \cdots \times [a_d, b_d)$.

Example 1.5. Let $f : \Omega \to \mathbb{R}$ be any function. For Borel sets $B \in \mathscr{B}(\mathbb{R})$ we define

$$\{f \in B\} := \{\omega \in \Omega : \ f(\omega) \in B\}.$$

The collection

$$\sigma(f) = \big\{\{f \in B\} : \ B \in \mathscr{B}(\mathbb{R})\big\}$$

is a σ-algebra in Ω, the *σ-algebra generated by f*.

1.2 Measurable functions

Let $(\Omega_1, \mathscr{F}_1)$ and $(\Omega_2, \mathscr{F}_2)$ be measurable spaces. A function $f : \Omega_1 \to \Omega_2$ is said to be *measurable* if $\{f \in A\} \in \mathscr{F}_1$ for all $A \in \mathscr{F}_2$. Clearly, compositions of measurable functions are measurable.

Proposition 1.6. *If $\mathscr{A}_2$ is a subset of $\mathscr{F}_2$ with the property that $\sigma(\mathscr{A}_2) = \mathscr{F}_2$, then a function $f : \Omega_1 \to \Omega_2$ is measurable if and only if*

$$\{f \in A\} \in \mathscr{F}_1 \textit{ for all } A \in \mathscr{A}_2.$$

Proof. Just notice that $\{A \in \mathscr{F}_2 : \ \{f \in A\} \in \mathscr{F}_1\}$ is a sub-σ-algebra of $\mathscr{F}_2$ containing $\mathscr{A}_2$. $\square$

In most applications we are concerned with functions f from a measurable space $(\Omega, \mathscr{F})$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Such functions are said to be *Borel measurable*. In what follows we summarise some measurability properties of Borel measurable functions (the adjective 'Borel' will be omitted when no confusion is likely to arise).

Proposition 1.6 implies that a function $f : \Omega \to \mathbb{R}$ is Borel measurable if and only if $\{f > a\} \in \mathscr{F}$ for all $a \in \mathbb{R}$. From this it follows that linear combinations, products, and quotients (if defined) of measurable functions are measurable. For example, if f and g are measurable, then $f + g$ is measurable since

$$\{f + g > a\} = \bigcup_{q \in \mathbb{Q}} \{f > q\} \cap \{g > a - q\}.$$

This, in turn, implies that fg is measurable, as we have

$$fg = \frac{1}{2}[(f + g)^2 - (f^2 + g^2)].$$

If $f = \sup_{n \geqslant 1} f_n$ pointwise and each f_n is measurable, then f is measurable since

$$\{f > a\} = \bigcup_{n=1}^{\infty} \{f_n > a\}.$$

It follows from $\inf_{n \geqslant 1} f_n = -\sup_{n \geqslant 1}(-f_n)$ that the pointwise infimum of a sequence of measurable functions is measurable as well. From this we get that the pointwise limes superior and limes inferior of measurable functions is measurable, since

$$\limsup_{n \to \infty} f_n = \lim_{n \to \infty} \Big(\sup_{k \geqslant n} f_k \Big) = \inf_{n \geqslant 1} \Big(\sup_{k \geqslant n} f_k \Big)$$

and $\liminf_{n \to \infty} f_n = -\limsup_{n \to \infty}(-f_n)$. In particular, the pointwise limit $\lim_{n \to \infty} f_n$ of a sequence of measurable functions is measurable.

In these results, it is implicitly understood that the suprema and infima exist and are finite pointwise. This restriction can be lifted by considering functions $f : \Omega \to [-\infty, \infty]$. Such a function is said to be measurable if the sets $f \in B\}$ as well as the sets $\{f = \pm\infty\}$ are in $\mathscr{F}$. This extension will also be useful later on.

1.3 Measures

Let $(\Omega, \mathscr{F})$ be a measurable space. A *measure* is a mapping $\mu : \mathscr{F} \to [0, \infty]$ with the following properties:

(1) $\mu(\varnothing) = 0$;
(2) For all disjoint set $A_1, A_2, \ldots$ in $\mathscr{F}$ we have $\mu\big(\bigcup_{n=1}^{\infty} A_n\big) = \sum_{n=1}^{\infty} \mu(A_n)$.

A triple $(\Omega, \mathscr{F}, \mu)$, with μ a measure on a measurable space $(\Omega, \mathscr{F})$, is called *measure space*. A measure space $(\Omega, \mathscr{F}, \mathbb{P})$ is called *finite* if μ is a *finite* measure, that is, if $\mu(\Omega) < \infty$. If $\mu(\Omega) = 1$, then μ is called a *probability measure* and $(\Omega, \mathscr{F}, \mu)$ is called a *probability space*. In probability theory, it is customary to use the symbol $\mathbb{P}$ for a probability measure.

The following properties of measures are easily checked:

(a) if $A_1 \subseteq A_2$ in $\mathscr{F}$, then $\mu(A_1) \leqslant \mu(A_2)$;
(b) if $A_1, A_2, \dots$ in $\mathscr{F}$, then

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leqslant \sum_{n=1}^{\infty} \mu(A_n);$$

Hint: Consider the disjoint sets $A_{n+1} \setminus (A_1 \cup \dots \cup A_n)$.
(c) if $A_1 \subseteq A_2 \subseteq \dots$ in $\mathscr{F}$, then

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} \mu(A_n);$$

Hint: Consider the disjoint sets $A_{n+1} \setminus A_n$.
(d) if $A_1 \supseteq A_2 \supseteq \dots$ in $\mathscr{F}$ and $\mu(A_1) < \infty$, then

$$\mu\Big(\bigcap_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} \mu(A_n).$$

Hint: Take complements.

Note that in c) and d) the limits (in $[0, \infty]$) exist by monotonicity.

Example 1.7. Let $(\Omega, \mathscr{F})$ and $(\tilde{\Omega}, \tilde{\mathscr{F}})$ be measurable spaces and let μ be a measure on $(\Omega, \mathscr{F})$. If $f : \Omega \to \tilde{\Omega}$ is measurable, then

$$f_*(\mu)(\tilde{A}) := \mu\{f \in \tilde{A}\}, \quad \tilde{A} \in \tilde{\mathscr{F}},$$

defines a measure $f_*(\mu)$ on $(\tilde{\Omega}, \tilde{\mathscr{F}})$. This measure is called the *image* of μ under f. In the context where $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space, the image of $\mathbb{P}$ under f is called the *distribution* of f.

Concerning existence and uniqueness of measures we have the following two fundamental results.

Theorem 1.8 (Caratheodory). *Let $\mathscr{A}$ be a non-empty collection of subsets of a set Ω that is closed under taking complements and finite unions. If $\mu : \mathscr{A} \to [0, \infty]$ is a mapping such that properties 1.2(1) and 1.2(2) hold, then there exists a unique measure $\widetilde{\mu} : \sigma(\mathscr{A}) \to [0, \infty]$ such that*

$$\widetilde{\mu}|_{\mathscr{A}} = \mu.$$

The proof is lengthy and rather tedious. The typical situation is that measures are given, and therefore the proof is omitted. The uniqueness part is a consequence of the next result, which is useful in many other situations as well.

Theorem 1.9 (Dynkin). *Let μ_1 and μ_2 be two finite measures defined on a measurable space $(\Omega, \mathscr{F})$. Let $\mathscr{A} \subseteq \mathscr{F}$ be a collection of sets with the following properties:*

(1) *$\mathscr{A}$ is closed under finite intersections;*
(2) *$\sigma(\mathscr{A})$, the σ-algebra generated by $\mathscr{A}$, equals $\mathscr{F}$.*

If $\mu_1(A) = \mu_2(A)$ for all $A \in \mathscr{A}$, then $\mu_1 = \mu_2$.

Proof. Let $\mathscr{D}$ denote the collection of all sets $D \in \mathscr{F}$ with $\mu_1(D) = \mu_2(D)$. Then $\mathscr{A} \subseteq \mathscr{D}$ and $\mathscr{D}$ is a *Dynkin system*, that is,

- $\Omega \in \mathscr{D}$;
- if $D_1 \subseteq D_2$ with $D_1, D_2 \in \mathscr{D}$, then also $D_2 \setminus D_1 \in \mathscr{D}$;
- if $D_1 \subseteq D_2 \subseteq \dots$ with all $D_n \in \mathscr{D}$, then also $\bigcup_{n=1}^{\infty} D_n \in \mathscr{D}$.

The finiteness of μ_1 and μ_2 is used to prove the second property.

By assumption we have $\mathscr{D} \subseteq \mathscr{F} = \sigma(\mathscr{A})$; we will show that $\sigma(\mathscr{A}) \subseteq \mathscr{D}$. To this end let $\mathscr{D}_0$ denote the smallest Dynkin system in $\mathscr{F}$ containing $\mathscr{A}$. We will show that $\sigma(\mathscr{A}) \subseteq \mathscr{D}_0$. In view of $\mathscr{D}_0 \subseteq \mathscr{D}$, this will prove the lemma.

Let $\mathscr{C} = \{D_0 \in \mathscr{D}_0 : \ D_0 \cap A \in \mathscr{D}_0 \text{ for all } A \in \mathscr{A}\}$. Then $\mathscr{C}$ is a Dynkin system and $\mathscr{A} \subseteq \mathscr{C}$ since $\mathscr{A}$ is closed under taking finite intersections. It follows that $\mathscr{D}_0 \subseteq \mathscr{C}$, since $\mathscr{D}_0$ is the smallest Dynkin system containing $\mathscr{A}$. But obviously, $\mathscr{C} \subseteq \mathscr{D}_0$, and therefore $\mathscr{C} = \mathscr{D}_0$.

Now let $\mathscr{C}' = \{D_0 \in \mathscr{D}_0 : \ D_0 \cap D \in \mathscr{D}_0 \text{ for all } D \in \mathscr{D}_0\}$. Then $\mathscr{C}'$ is a Dynkin system and the fact that $\mathscr{C} = \mathscr{D}_0$ implies that $\mathscr{A} \subseteq \mathscr{C}'$. Hence $\mathscr{D}_0 \subseteq \mathscr{C}'$, since $\mathscr{D}_0$ is the smallest Dynkin system containing $\mathscr{A}$. But obviously, $\mathscr{C}' \subseteq \mathscr{D}_0$, and therefore $\mathscr{C}' = \mathscr{D}_0$.

It follows that $\mathscr{D}_0$ is closed under taking finite intersections. But a Dynkin system with this property is a σ-algebra. Thus, $\mathscr{D}_0$ is a σ-algebra, and now $\mathscr{A} \subseteq \mathscr{D}_0$ implies that also $\sigma(\mathscr{A}) \subseteq \mathscr{D}_0$. □

Example 1.10 (Lebesgue measure). In $\mathbb{R}^d$, let $\mathscr{A}$ be the collection of all finite unions of rectangles $[a_1, b_1) \times \cdots \times [a_d, b_d)$. Define $\mu : \mathscr{A} \to [0, \infty]$ by

$$\mu([a_1, b_1) \times \cdots \times [a_d, b_d)) := \prod_{n=1}^{d} (b_n - a_n)$$

and extend this definition to finite disjoint unions by additivity. It can be verified that $\mathscr{A}$ and μ satisfy the assumptions of Caratheodory's theorem. The resulting measure μ on $\sigma(\mathscr{A}) = \mathscr{B}(\mathbb{R}^d)$, the Borel σ-algebra of $\mathbb{R}^d$, is called the *Lebesgue measure*.

Example 1.11. Let $(\Omega_1, \mathscr{F}_1, \mathbb{P}_1)$, $(\Omega_2, \mathscr{F}_2, \mathbb{P}_2), \dots$ be a sequence of probability spaces. There exists a unique probability measure $\mathbb{P} = \prod_{n=1}^{\infty} \mathbb{P}_n$ on $\mathscr{F} = \prod_{n=1}^{\infty} \mathscr{F}_n$ which satisfies

$$\mathbb{P}(A_1 \times \cdots \times A_N \times \Omega_{N+1} \times \Omega_{N+2} \times \dots) = \prod_{n=1}^{N} \mathbb{P}_n(A_n)$$

for all $N \geqslant 1$ and $A_n \in \mathscr{F}_n$ for $n = 1, \dots, N$. Finite products of arbitrary measures $\mu_1, \dots, \mu_N$ can be defined similarly.

For example, the Lebesgue measure on $\mathbb{R}^d$ is the product measure of d copies of the Lebesgue measure on $\mathbb{R}$.

As an illustration of the use of Dynkin's lemma we conclude this section with an elementary result on independence. In probability theory, a measurable function $f : \Omega \to \mathbb{R}$, where $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space, is called a *random variable*. Random variables $f_1, \dots, f_N : \Omega \to \mathbb{R}$ are said to be *independent* if for all $B_1, \dots, B_N \in \mathscr{B}(\mathbb{R})$ we have

$$\mathbb{P}\{f_1 \in B_1, \dots, f_N \in B_N\} = \prod_{n=1}^{N} \mathbb{P}\{f_1 \in B_1\}.$$

Proposition 1.12. *The random variables $f_1, \dots, f_N : \Omega \to \mathbb{R}$ are independent if and only if $\nu = \prod_{n=1}^{N} \nu_n$, where ν is the distribution of $(f_1, \dots, f_N)$ and ν_n is the distribution of f_n, $n = 1, \dots, N$.*

Proof. By definition, $f_1, \dots, f_N$ are independent if and only if $\nu(B) = \prod_{n=1}^{N} \nu_n(B_n)$ for all $B = B_1 \times \cdots \times B_N$ with $B_n \in \mathscr{B}(\mathbb{R})$ for all $n = 1, \dots, N$. Since these sets generate $\mathscr{B}(\mathbb{R}^N)$, by Dynkin's lemma the latter implies $\nu = \prod_{n=1}^{N} \nu_n$. □

1.4 The Lebesgue integral

Let $(\Omega, \mathscr{F})$ be a measurable space. A *simple function* is a function $f : \Omega \to \mathbb{R}$ that can be represented in the form

$$f = \sum_{n=1}^{N} c_n 1_{A_n}$$

with coefficients $c_n \in \mathbb{R}$ and disjoint sets $A_n \in \mathscr{F}$ for all $n = 1, \dots, N$.

Proposition 1.13. *A function $f : \Omega \to [-\infty, \infty]$ is measurable if and only if it is the pointwise limit of a sequence of simple functions. If in addition f is non-negative, we may arrange that $0 \leqslant f_n \uparrow f$ pointwise.*

Proof. Take, for instance, $f_n = \sum_{m=-2^{2n}}^{2^{2n}} \frac{m}{2^n} 1_{\{f \in [\frac{m}{2^n}, \frac{m+1}{2^n})\}}$. □

1.4.1 Integration of simple functions

Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. For a non-negative simple function $f = \sum_{n=1}^{N} c_n 1_{A_n}$ we define

$$\int_\Omega f \, d\mu := \sum_{n=1}^{N} c_n \mu(A_n).$$

We allow $\mu(A_n)$ to be infinite; this causes no problems because the coefficients c_n are non-negative (we use the convention $0 \cdot \infty = 0$). It is easy to check that this integral is well-defined, in the sense that it does not depend on the particular representation of f as a simple function. Also, the integral is linear with respect to addition and multiplication with non-negative scalars,

$$\int_\Omega a f_1 + b f_2 \, d\mu = a \int_\Omega f_1 \, d\mu + b \int_\Omega f_2 \, d\mu,$$

and monotone in the sense that if $0 \leqslant f_1 \leqslant f_2$ pointwise, then

$$\int_\Omega f_1 \, d\mu \leqslant \int_\Omega f_2 \, d\mu.$$

1.4.2 Integration of non-negative functions

In what follows, a *non-negative function* is a function with values in $[0, \infty]$.

For a non-negative measurable function f we choose a sequence of simple functions $0 \leqslant f_n \uparrow f$ and define

$$\int_\Omega f \, d\mu := \lim_{n \to \infty} \int_\Omega f_n \, d\mu.$$

The following lemma implies that this definition does not depend on the approximating sequence.

Lemma 1.14. *For a non-negative measurable function f and non-negative simple functions f_n and g such that $0 \leqslant f_n \uparrow f$ and $g \leqslant f$ pointwise we have*

$$\int_\Omega g \, d\mu \leqslant \lim_{n \to \infty} \int_\Omega f_n \, d\mu.$$

Proof. First consider the case $g = 1_A$. Fix $\varepsilon > 0$ arbitrary and let $A_n = \{1_A f_n \geqslant 1 - \varepsilon\}$. Then $A_1 \subseteq A_2 \subseteq \ldots \uparrow A$ and therefore $\mu(A_n) \uparrow \mu(A_n)$. Since $1_A f_n \geqslant (1 - \varepsilon) 1_{A_n}$,

$$\lim_{n \to \infty} \int_\Omega 1_A f_n \, d\mu \geqslant (1 - \varepsilon) \lim_{n \to \infty} \mu(A_n) = (1 - \varepsilon)\mu(A) = (1 - \varepsilon) \int_\Omega g \, d\mu.$$

This proves the lemma for $g = 1_A$. The general case follows by linearity. $\square$

The integral is linear and monotone on the set of non-negative measurable functions. Indeed, if f and g are such functions and $0 \leqslant f_n \uparrow f$ and $0 \leqslant g_n \uparrow g$, then for $a, b \geqslant 0$ we have $0 \leqslant af_n + bg_n \uparrow af + bg$ and therefore

$$\begin{aligned}\int_\Omega af + bg\, d\mu &= \lim_{n\to\infty} \int_\Omega af_n + bg_n\, d\mu \\ &= a \lim_{n\to\infty} \int_\Omega f_n\, d\mu + b \lim_{n\to\infty} \int_\Omega g_n\, d\mu \\ &= a \int_\Omega f\, d\mu + b \int_\Omega g\, d\mu.\end{aligned}$$

If such f and g satisfy $f \leqslant g$ pointwise, then by noting that $0 \leqslant f_n \leqslant \max\{f_n, g_n\} \uparrow g$ we see that

$$\int_\Omega f\, d\mu = \lim_{n\to\infty} \int_\Omega f_n\, d\mu \leqslant \lim_{n\to\infty} \int_\Omega \max\{f_n, g_n\}\, d\mu \leqslant \int_\Omega g\, d\mu.$$

The next theorem is the cornerstone of Integration Theory.

Theorem 1.15 (Monotone Convergence Theorem). *Let $0 \leqslant f_1 \leqslant f_2 \leqslant \dots$ be a sequence of non-negative measurable functions converging pointwise to a function f. Then,*

$$\lim_{n\to\infty} \int_\Omega f_n\, d\mu = \int_\Omega f\, d\mu.$$

Proof. First note that f is non-negative and measurable. For each $n \geqslant 1$ choose a sequence of simple functions $0 \leqslant f_{nk} \uparrow_k f_n$. Set

$$g_{nk} := \max\{f_{1k}, \dots, f_{nk}\}.$$

For $m \leqslant n$ we have $g_{mk} \leqslant g_{nk}$. Also, for $k \leqslant l$ we have $f_{mk} \leqslant f_{ml}$, $m = 1, \dots, n$, and therefore $g_{nk} \leqslant g_{nl}$. We conclude that the functions g_{nk} are monotone in both indices.

From $f_{mk} \leqslant f_m \leqslant f_n$, $1 \leqslant m \leqslant n$, we see that $f_{nk} \leqslant g_{nk} \leqslant f_n$, and we conclude that $0 \leqslant g_{nk} \uparrow_k f_n$. From

$$f_n = \lim_{k\to\infty} g_{nk} \leqslant \lim_{k\to\infty} g_{kk} \leqslant f$$

we deduce that $0 \leqslant g_{kk} \uparrow f$. Recalling that $g_{kk} \leqslant f_k$ it follows that

$$\int_\Omega f\, d\mu = \lim_{k\to\infty} \int_\Omega g_{kk}\, d\mu \leqslant \lim_{k\to\infty} \int_\Omega f_k\, d\mu \leqslant \int_\Omega f\, d\mu. \qquad \square$$

Example 1.16. In the situation of Example 1.7 we have the following *substitution formula*. For any measurable $f : \Omega_1 \to \Omega_2$ and non-negative measurable $\phi : \Omega_2 \to \mathbb{R}$,

$$\int_{\Omega_1} \phi \circ f \, d\mu = \int_{\Omega_2} \phi \, df_*(\mu).$$

To prove this, note the this is trivial for simple functions $\phi = 1_A$ with $A \in \mathscr{F}_2$. By linearity, the identity extends to non-negative simple functions ϕ, and by monotone convergence (using Proposition 1.13) to non-negative measurable functions ϕ.

From the monotone convergence theorem we deduce the following useful corollary.

Theorem 1.17 (Fatou Lemma). *Let $(f_n)_{n=1}^{\infty}$ be a sequence of non-negative measurable functions on $(\Omega, \mathscr{F}, \mu)$. Then*

$$\int_{\Omega} \liminf_{n\to\infty} f_n \, d\mu \leqslant \liminf_{n\to\infty} \int_{\Omega} f_n \, d\mu.$$

Proof. From $\inf_{k\geqslant n} f_k \leqslant f_m$, $m \geqslant n$, we infer

$$\int_{\Omega} \inf_{k\geqslant n} f_k \, d\mu \leqslant \inf_{m\geqslant n} \int_{\Omega} f_m \, d\mu.$$

Hence, by the monotone convergence theorem,

$$\begin{aligned}\int_{\Omega} \liminf_{n\to\infty} f_n \, d\mu &= \int_{\Omega} \lim_{n\to\infty} \inf_{k\geqslant n} f_k \, d\mu \\ &= \lim_{n\to\infty} \int_{\Omega} \inf_{k\geqslant n} f_k \, d\mu \leqslant \lim_{n\to\infty} \inf_{m\geqslant n} \int_{\Omega} f_m \, d\mu = \liminf_{n\to\infty} \int_{\Omega} f_n \, d\mu.\end{aligned}$$

□

1.4.3 Integration of real-valued functions

A measurable function $f : \Omega \to [-\infty, \infty]$ is called *integrable* if

$$\int_{\Omega} |f| \, d\mu < \infty.$$

Clearly, if f and g are measurable and $|g| \leqslant |f|$ pointwise, then g is integrable if f is integrable. In particular, if f is integrable, then the non-negative functions f^+ and f^- are integrable, and we define

$$\int_{\Omega} f \, d\mu := \int_{\Omega} f^+ \, d\mu - \int_{\Omega} f^- \, d\mu.$$

For a set $A \in \mathscr{F}$ we write

$$\int_A f \, d\mu := \int_{\Omega} 1_A f \, d\mu,$$

noting that $1_A f$ is integrable. The monotonicity and additivity properties of this integral carry over to this more general situation, provided we assume that the functions we integrate are integrable.

The next result is among the most useful in analysis.

Theorem 1.18 (Dominated Convergence Theorem). *Let $(f_n)_{n=1}^{\infty}$ be a sequence of integrable functions such that $\lim_{n\to\infty} f_n = f$ pointwise. If there exists an integrable function g such that $|f_n| \leqslant |g|$ for all $n \geqslant 1$, then*

$$\lim_{n\to\infty} \int_{\Omega} |f_n - f| \, d\mu = 0.$$

In particular,

$$\lim_{n\to\infty} \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu.$$

Proof. We make the preliminary observation that if $(h_n)_{n=1}^{\infty}$ is a sequence of non-negative measurable functions such that $\lim_{n\to\infty} h_n = 0$ pointwise and h is a non-negative integrable function such that $h_n \leqslant h$ for all $n \geqslant 1$, then by the Fatou lemma

$$\begin{aligned}\int_{\Omega} h \, d\mu &= \int_{\Omega} \liminf_{n\to\infty} (h - h_n) \, d\mu \\ &\leqslant \liminf_{n\to\infty} \int_{\Omega} h - h_n \, d\mu = \int_{\Omega} h \, d\mu - \limsup_{n\to\infty} \int_{\Omega} h_n \, d\mu.\end{aligned}$$

Since $\int_{\Omega} h \, d\mu$ is finite, it follows that $0 \leqslant \limsup_{n\to\infty} \int_{\Omega} h_n \, d\mu \leqslant 0$ and therefore

$$\lim_{n\to\infty} \int_{\Omega} h_n \, d\mu = 0.$$

The theorem follows by applying this to $h_n = |f_n - f|$ and $h = 2|g|$. □

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. The integral of an integrable random variable $f : \Omega \to \mathbb{R}$ is called the *expectation* of f, notation

$$\mathbb{E}f = \int_{\Omega} f \, d\mathbb{P}.$$

Proposition 1.19. *If $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ are independent integrable random variables, then $fg : \Omega \to \mathbb{R}$ is integrable and*

$$\mathbb{E}fg = \mathbb{E}f \, \mathbb{E}g.$$

Proof. To prove this we first observe that for any non-negative Borel measurable $\phi : \mathbb{R} \to \mathbb{R}$, the random variables $\phi \circ f$ and $\phi \circ g$ are independent. Taking $\phi(x) = x^+$ and $\phi(x) = x^-$, it follows that we may consider the positive and negative parts of f and g separately. In doing so, we may assume that f and g are non-negative.

Denoting the distributions of f, g, and (f, g) by ν_f, ν_g, and $\nu_{(f,g)}$ respectively, by the substitution formula (with $\phi(x, y) = xy1_{[0,\infty)^2}(x, y)$) and the fact that $\nu_{(f,g)} = \nu_f \times \nu_g$ we obtain

$$\mathbb{E}(fg) = \int_{[0,\infty)^2} xy\, d\nu_{(f,g)}(x,y) = \int_{[0,\infty)^2} xy\, d(\nu_f \times d\nu_g)(x,y)$$
$$= \int_{[0,\infty)} \int_{[0,\infty)} xy\, d\nu_f(x) d\nu_g(y) = \mathbb{E}f\, \mathbb{E}g.$$

This proves the integrability of fg along with the desired identity. □

1.4.4 The role of null sets

We begin with a simple observation.

Proposition 1.20. *Let f be a non-negative measurable function.*

1. *If $\int_\Omega f\, d\mu < \infty$, then $\mu\{f = \infty\} = 0$.*
2. *If $\int_\Omega f\, d\mu = 0$, then $\mu\{f \neq 0\} = 0$.*

Proof. For all $c > 0$ we have $0 \leqslant c1_{\{f=\infty\}} \leqslant f$ and therefore

$$0 \leqslant c\mu\{f = \infty\} \leqslant \int_\Omega f\, d\mu.$$

The first result follows from this by letting $c \to \infty$. For the second, note that for all $n \geqslant 1$ we have $\frac{1}{n}1_{\{f \geqslant \frac{1}{n}\}} \leqslant f$ and therefore

$$\frac{1}{n}\mu\{f \geqslant \frac{1}{n}\} \leqslant \int_\Omega f\, d\mu = 0.$$

It follows that $\mu\{f \geqslant \frac{1}{n}\} = 0$. Now note that $\{f > 0\} = \bigcup_{n=1}^\infty \{f \geqslant \frac{1}{n}\}$. □

Proposition 1.21. *If f is integrable and $\mu\{f \neq 0\} = 0$, then $\int_\Omega f\, d\mu = 0$.*

Proof. By considering f^+ and f^- separately we may assume f is non-negative. Choose simple functions $0 \leqslant f_n \uparrow f$. Then $\mu\{f_n > 0\} \leqslant \mu\{f > 0\} = 0$ and therefore $\int_\Omega f_n\, d\mu = 0$ for all $n \geqslant 1$. The result follows from this. □

What these results show is that in everything that has been said so far, we may replace pointwise convergence by pointwise convergence μ-almost everywhere, where the latter means that we allow an exceptional set of μ-measure zero. This applies, in particular, to the monotonicity assumption in the monotone convergence theorem and the convergence and domination assumptions in the dominated convergence theorem.

1.5 L^p-spaces

Let μ be a measure on $(\Omega, \mathscr{F})$ and fix $1 \leqslant p < \infty$. We define $\mathscr{L}^p(\mu)$ as the set of all measurable functions $f : \Omega \to \mathbb{R}$ such that

$$\int_\Omega |f|^p \, d\mu < \infty.$$

For such functions we define

$$\|f\|_p := \Big(\int_\Omega |f|^p \, d\mu \Big)^{\frac{1}{p}}.$$

The next result shows that $\mathscr{L}^p(\mu)$ is a vector space:

Proposition 1.22 (Minkowski's inequality). *For all $f, g \in \mathscr{L}^p(\mu)$ we have $f + g \in \mathscr{L}^p(\mu)$ and*

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

Proof. By elementary calculus it is checked that for all non-negative real numbers a and b one has

$$(a + b)^p = \inf_{t \in (0,1)} t^{1-p} a^p + (1 - t)^{1-p} b^p.$$

Applying this identity to $|f(x)|$ and $|g(x)|$ and integrating with respect to μ, for all fixed $t \in (0, 1)$ we obtain

$$\begin{aligned} \int_\Omega |f(x) + g(x)|^p \, d\mu(x) &\leqslant \int_\Omega (|f(x)| + |g(x)|)^p \, d\mu(x) \\ &\leqslant t^{1-p} \int_\Omega |f(x)|^p \, d\mu(x) + (1 - t)^{1-p} \int_\Omega |g(x)|^p \, d\mu(x). \end{aligned}$$

Stated differently, this says that

$$\|f + g\|_p^p \leqslant t^{1-p} \|f\|_p^p + (1 - t)^{1-p} \|g\|_p^p.$$

Taking the infimum over all $t \in (0, 1)$ gives the result. □

In spite of this result, $\| \cdot \|_p$ is not a norm on $\mathscr{L}^p(\mu)$, because $\|f\|_p = 0$ only implies that $f = 0$ μ-almost everywhere. In order to get around this imperfection, we define an equivalence relation $\sim$ on $\mathscr{L}^p(\mu)$ by

$$f \sim g \Leftrightarrow f = g \ \mu\text{-almost everywhere.}$$

The equivalence class of a function f modulo $\sim$ is denoted by $[f]$. On the quotient space

$$L^p(\mu) := \mathscr{L}^p(\mu) / \sim$$

we define a scalar multiplication and addition in the natural way:

$$c[f] := [cf], \quad [f] + [g] := [f + g].$$

We leave it as an exercise to check that both operators are well defined. With these operations, $L^p(\mu)$ is a normed vector space with respect to the norm

$$\|[f]\|_p := \|f\|_p.$$

Following common practice we shall make no distinction between a function in $\mathscr{L}^p(\mu)$ and its equivalence class in $L^p(\mu)$.

The next result states that $L^p(\mu)$ is a Banach space (which, by definition, is a complete normed vector space):

Theorem 1.23. *The normed space $L^p(\mu)$ is complete.*

Proof. Let $(f_n)_{n=1}^\infty$ be a Cauchy sequence with respect to the norm $\|\cdot\|_p$ of $L^p(\mu)$. By passing to a subsequence we may assume that

$$\|f_{n+1} - f_n\|_p \leqslant \frac{1}{2^n}, \quad n = 1, 2, \ldots$$

Define the non-negative measurable functions

$$g_N := \sum_{n=0}^{N-1} |f_{n+1} - f_n|, \quad g := \sum_{n=0}^{\infty} |f_{n+1} - f_n|,$$

with the convention that $f_0 = 0$. By the monotone convergence theorem,

$$\int_\Omega g^p \, d\mu = \lim_{N\to\infty} \int_\Omega g_N^p \, d\mu.$$

Taking p-th roots and using Minkowski's inequality we obtain

$$\|g\|_p = \lim_{N\to\infty} \|g_N\|_p \leqslant \lim_{N\to\infty} \sum_{n=0}^{N-1} \|f_{n+1} - f_n\|_p = \sum_{n=0}^{\infty} \|f_{n+1} - f_n\|_p \leqslant 1 + \|f_1\|_p.$$

If follows that g is finitely-valued μ-almost everywhere, which means that the sum defining g converges absolutely μ-almost everywhere. As a result, the sum

$$f := \sum_{n=0}^{\infty} (f_{n+1} - f_n)$$

converges on the set $\{g < \infty\}$. On this set we have

$$f = \lim_{N\to\infty} \sum_{n=0}^{N-1} (f_{n+1} - f_n) = \lim_{N\to\infty} f_N.$$

Defining f to be zero on the null set $\{g = \infty\}$, the resulting function f is measurable. From

$$|f_N|^p = \Big| \sum_{n=0}^{N-1} (f_{n+1} - f_n) \Big|^p \leqslant \Big(\sum_{n=0}^{N-1} |f_{n+1} - f_n| \Big)^p \leqslant |g|^p$$

and the dominated convergence theorem we conclude that

$$\lim_{N\to\infty} \|f - f_N\|_p = 0.$$

We have proved that a subsequence of the original Cauchy sequence converges to f in $L^p(\mu)$. As is easily verified, this implies that the original Cauchy sequence converges to f as well. □

In the course of the proof we obtained the following result:

Corollary 1.24. *Every convergence sequence in $L^p(\mu)$ has a μ-almost everywhere convergent subsequence.*

For $p = \infty$ the above definitions have to be modified a little. We define $\mathscr{L}^\infty(\mu)$ to be the vector space of all measurable functions $f : \Omega \to \mathbb{R}$ such that

$$\sup_{\omega\in\Omega} |f(\omega)| < \infty$$

and define, for such functions,

$$\|f\|_\infty := \sup_{\omega\in\Omega} |f(\omega)|.$$

The space

$$L^\infty(\mu) := \mathscr{L}^\infty(\mu)/\sim$$

is again a normed vector space in a natural way. The simple proof that $L^\infty(\mu)$ is a Banach space is left as an exercise.

We continue with an approximation result, the proof of which is an easy application of Proposition 1.13 together with the dominated convergence theorem:

Proposition 1.25. *For $1 \leqslant p \leqslant \infty$, the simple functions in $L^p(\mu)$ are dense in $L^p(\mu)$.*

We conclude with an important inequality, known as *Hölder's inequality.* For $p = q = 2$, it is known as the *Cauchy-Schwarz inequality.*

Theorem 1.26. *Let $1 \leqslant p, q \leqslant \infty$ satisfy $\frac{1}{p} + \frac{1}{q} = 1$. If $f \in L^p(\mu)$ and $g \in L^q(\mu)$, then $fg \in L^1(\mu)$ and*

$$\|fg\|_1 \leqslant \|f\|_p \|g\|_q.$$

Proof. For $p = 1$, $q = \infty$ and for $p = \infty$, $q = 1$, this follows by a direct estimate. Thus we may assume from now on that $1 < p, q < \infty$. The inequality is then proved in the same way as Minkowski's inequality, this time using the identity

$$ab = \inf_{t>0} \Big(\frac{t^p a^p}{p} + \frac{b^q}{q t^q} \Big).$$ □

1.6 Fubini's theorem

A measure space $(\Omega, \mathscr{F}, \mu)$ is called *σ-finite* if we can write $\Omega = \bigcup_{n=1}^{\infty} \Omega^{(n)}$ with sets $\Omega^{(n)} \in \mathscr{F}$ satisfying $\mu(\Omega^{(n)}) < \infty$.

Theorem 1.27 (Fubini). *Let $(\Omega_1, \mathscr{F}_1, \mu_1)$ and $(\Omega_1, \mathscr{F}_1, \mu_1)$ be σ-finite measure spaces. For any non-negative measurable function $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ the following assertions hold:*

(1) *The non-negative function $\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2)\, d\mu_2(\omega_2)$ is measurable;*
(2) *The non-negative function $\omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2)\, d\mu_1(\omega_1)$ is measurable;*
(3) *We have*

$$\int_{\Omega_1 \times \Omega_2} f\, d(\mu_1 \times \mu_2) = \int_{\Omega_2} \int_{\Omega_1} f\, d\mu_1\, d\mu_2 = \int_{\Omega_1} \int_{\Omega_2} f\, d\mu_2\, d\mu_1.$$

We shall not give a detailed proof, but sketch the main steps. First, for $f(\omega_1, \omega_2) = 1_{A_1}(\omega_1) 1_{A_2}(\omega_2)$ with $A_1 \in \mathscr{F}_1$, $A_2 \in \mathscr{F}_2$, the result is clear. Dynkin's lemma is used to extend the result to functions $f = 1_A$ with $A \in \mathscr{F} = \mathscr{F}_1 \times \mathscr{F}_2$. One has to assume first that μ_1 and μ_2 are finite; the σ-finite case then follows by approximation. By taking linear combinations, the result extends to simple functions. The result for arbitrary non-negative measurable functions then follows by monotone convergence.

A variation of the Fubini theorem holds if we replace 'non-negative measurable' by 'integrable'. In that case, for μ_1-almost all $\omega_1 \in \Omega_1$ the function $\omega_2 \mapsto f(\omega_1, \omega_2)$ is integrable and for μ_2-almost all $\omega_2 \in \Omega_2$ the function $\omega_1 \mapsto f(\omega_1, \omega_2)$ is integrable. Moreover, the functions $\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2)\, d\mu_2(\omega_2)$ and $\omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2)\, d\mu_1(\omega_1)$ are integrable and the above identities hold.

1.7 Notes

The proof of the monotone convergence theorem is taken from Kallenberg [1]. The proofs of the Minkowski inequality and Hölder inequality, although not the ones that are usually presented in introductory courses, belong to the folklore of the subject.

2

Conditional expectations

In this lecture we introduce conditional expectations and study some of their basic properties.

We begin by recalling some terminology. A *random variable* is a measurable function on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. In this context, Ω is often called the *sample space* and the sets in $\mathscr{F}$ the *events*. For an integrable random variable $f : \Omega \to \mathbb{R}$ we write

$$\mathbb{E}f := \int_\Omega f \, d\mathbb{P}$$

for its *expectation*.

Below we fix a sub-σ-algebra $\mathscr{G}$ of $\mathscr{F}$. For $1 \leqslant p \leqslant \infty$ we denote by $L^p(\Omega, \mathscr{G})$ the subspace of all $f \in L^p(\Omega)$ having a $\mathscr{G}$-measurable representative. With this notation, $L^p(\Omega) = L^p(\Omega, \mathscr{F})$. We will be interested in the problem of finding good approximations of a random variable $f \in L^p(\Omega)$ by random variables in $L^p(\Omega; \mathscr{G})$.

Lemma 2.1. *$L^p(\Omega, \mathscr{G})$ is a closed subspace of $L^p(\Omega)$.*

Proof. Suppose that $(\xi_n)_{n=1}^\infty$ is a sequence in $L^p(\Omega, \mathscr{G})$ such that $\lim_{n\to\infty} f_n = f$ in $L^p(\Omega)$. We may assume that the f_n are pointwise defined and $\mathscr{G}$-measurable. After passing to a subsequence (when $1 \leqslant p < \infty$) we may furthermore assume that $\lim_{n\to\infty} f_n = f$ almost surely. The set C of all $\omega \in \Omega$ where the sequence $(f_n(\omega))_{n=1}^\infty$ converges is $\mathscr{G}$-measurable. Put $\widetilde{f} := \lim_{n\to\infty} 1_C f_n$, where the limit exists pointwise. The random variable $\widetilde{f}$ is $\mathscr{G}$-measurable and agrees almost surely with f. This shows that f defines an element of $L^p(\Omega, \mathscr{G})$. $\square$

Our aim is to show that $L^p(\Omega, \mathscr{G})$ is the range of a contractive projection in $L^p(\Omega)$. For $p = 2$ this is clear: we have the orthogonal decomposition

$$L^2(\Omega) = L^2(\Omega, \mathscr{G}) \oplus L^2(\Omega, \mathscr{G})^\perp$$

and the projection we have in mind is the orthogonal projection, denoted by $P_{\mathscr{G}}$, onto $L^2(\Omega, \mathscr{G})$ along this decomposition. Following common usage we write

$$\mathbb{E}(f|\mathscr{G}) := P_{\mathscr{G}}(f), \quad f \in L^2(\Omega),$$

and call $\mathbb{E}(f|\mathscr{G})$ the *conditional expectation* of f with respect to $\mathscr{G}$. Let us emphasise that $\mathbb{E}(f|\mathscr{G})$ is defined as an element of $L^2(\Omega, \mathscr{G})$, that is, as an equivalence class of random variables.

Lemma 2.2. *For all $f \in L^2(\Omega)$ and $G \in \mathscr{G}$ we have*

$$\int_G \mathbb{E}(f|\mathscr{G})\, d\mathbb{P} = \int_G f\, d\mathbb{P}.$$

As a consequence, if $f \geqslant 0$ almost surely, then $\mathbb{E}(f|\mathscr{G}) \geqslant 0$ almost surely.

Proof. By definition we have $f - \mathbb{E}(f|\mathscr{G}) \perp L^2(\Omega, \mathscr{G})$. If $G \in \mathscr{G}$, then $1_G \in L^2(\Omega, \mathscr{G})$ and therefore

$$\int_\Omega 1_G(f - \mathbb{E}(f|\mathscr{G}))\, d\mathbb{P} = 0.$$

This gives the desired identity. For the second assertion, choose a $\mathscr{G}$-measurable representative of $g := \mathbb{E}(f|\mathscr{G})$ and apply the identity to the $\mathscr{G}$-measurable set $\{g < 0\}$. □

Taking $G = \Omega$ we obtain the identity $\mathbb{E}(\mathbb{E}(f|\mathscr{G})) = \mathbb{E}f$. This will be used in the lemma, which asserts that the mapping $f \mapsto \mathbb{E}(f|\mathscr{G})$ is L^1-bounded.

Lemma 2.3. *For all $f \in L^2(\Omega)$ we have $\mathbb{E}|\mathbb{E}(f|\mathscr{G})| \leqslant \mathbb{E}|f|$.*

Proof. It suffices to check that $|\mathbb{E}(f|\mathscr{G})| \leqslant \mathbb{E}(|f|\,|\mathscr{G})$, since then the lemma follows from $\mathbb{E}|\mathbb{E}(f|\mathscr{G})| \leqslant \mathbb{E}\mathbb{E}(|f|\,|\mathscr{G}) = \mathbb{E}|f|$. Splitting f into positive and negative parts, almost surely we have

$$\begin{aligned} |\mathbb{E}(f|\mathscr{G})| &= |\mathbb{E}(f^+|\mathscr{G}) - \mathbb{E}(f^-|\mathscr{G})| \\ &\leqslant |\mathbb{E}(f^+|\mathscr{G})| + |\mathbb{E}(f^-|\mathscr{G})| = \mathbb{E}(f^+|\mathscr{G}) + \mathbb{E}(f^-|\mathscr{G}) = \mathbb{E}(|f|\,|\mathscr{G}). \end{aligned}$$ □

Since $L^2(\Omega)$ is dense in $L^1(\Omega)$ this lemma shows that the condition expectation operator has a unique extension to a contractive linear operator on $L^1(\Omega)$, which we also denote by $\mathbb{E}(\cdot|\mathscr{G})$. This operator is a projection (this follows by noting that it is a projection in $L^2(\Omega)$ and then using that $L^2(\Omega)$ is dense in $L^1(\Omega)$) and it is positive in the sense that it maps positive random variables to positive random variables; this follows from Lemma 2.2 by approximation.

Lemma 2.4 (Conditional Jensen inequality). *If $\phi : \mathbb{R} \to \mathbb{R}$ is convex, then for all $f \in L^1(\Omega)$ such that $\phi \circ f \in L^1(\Omega)$ we have, almost surely,*

$$\phi \circ \mathbb{E}(f|\mathscr{G}) \leqslant \mathbb{E}(\phi \circ f|\mathscr{G}).$$

Proof. If $a, b \in \mathbb{R}$ are such that $at + b \leqslant \phi(t)$ for all $t \in \mathbb{R}$, then the positivity of the conditional expectation operator gives

$$a\mathbb{E}(f|\mathscr{G}) + b = \mathbb{E}(af + b|\mathscr{G}) \leqslant \mathbb{E}(\phi \circ f|\mathscr{G})$$

almost surely. Since ϕ is convex we can find real sequences $(a_n)_{n=1}^\infty$ and $(b_n)_{n=1}^\infty$ such that $\phi(t) = \sup_{n\geqslant 1}(a_n t + b_n)$ for all $t \in \mathbb{R}$; we leave the proof of this fact as an exercise. Hence almost surely,

$$\phi \circ \mathbb{E}(f|\mathscr{G}) = \sup_{n\geqslant 1} a_n \mathbb{E}(f|\mathscr{G}) + b_n \leqslant \mathbb{E}(\phi \circ f|\mathscr{G}). \qquad \square$$

Theorem 2.5 (L^p-contractivity). *For all $1 \leqslant p \leqslant \infty$ the conditional expectation operator extends to a contractive positive projection on $L^p(\Omega)$ with range $L^p(\Omega, \mathscr{G})$. For $f \in L^p(\Omega)$, the random variable $\mathbb{E}(f|\mathscr{G})$ is the unique element of $L^p(\Omega, \mathscr{G})$ with the property that for all $G \in \mathscr{G}$,*

$$\int_G \mathbb{E}(f|\mathscr{G})\, d\mathbb{P} = \int_G f\, d\mathbb{P}. \tag{2.1}$$

Proof. For $1 \leqslant p < \infty$ the L^p-contractivity follows from Lemma 2.4 applied to the convex function $\phi(t) = |t|^p$. For $p = \infty$ we argue as follows. If $f \in L^\infty(\Omega)$, then $0 \leqslant |f| \leqslant \|f\|_\infty 1_\Omega$ and therefore $0 \leqslant \mathbb{E}(|f|\,|\mathscr{G}) \leqslant \|f\|_\infty 1_\Omega$ almost surely. Hence, $\mathbb{E}(|f|\,|\mathscr{G}) \in L^\infty(\Omega)$ and $\|\mathbb{E}(|f|\,|\mathscr{G})\|_\infty \leqslant \|f\|_\infty$.

For $2 \leqslant p \leqslant \infty$, (2.1) follows from Lemma 2.2. For $f \in L^p(\Omega)$ with $1 \leqslant p < 2$ we choose a sequence $(f_n)_{n=1}^\infty$ in $L^2(\Omega)$ such that $\lim_{n\to\infty} f_n = f$ in $L^p(\Omega)$. Then $\lim_{n\to\infty} \mathbb{E}(f_n|\mathscr{G}) = \mathbb{E}(f|\mathscr{G})$ in $L^p(\Omega)$ and therefore, for any $G \in \mathscr{G}$,

$$\int_G \mathbb{E}(f|\mathscr{G})\, d\mathbb{P} = \lim_{n\to\infty} \int_G \mathbb{E}(f_n|\mathscr{G})\, d\mathbb{P} = \lim_{n\to\infty} \int_G f_n\, d\mathbb{P} = \int_G f\, d\mathbb{P}.$$

If $g \in L^p(\Omega, \mathscr{G})$ satisfies $\int_G g\, d\mathbb{P} = \int_G f\, d\mathbb{P}$ for all $G \in \mathscr{G}$, then $\int_G g\, d\mathbb{P} = \int_G \mathbb{E}(f|\mathscr{G})\, d\mathbb{P}$ for all $G \in \mathscr{G}$. Since both g and $E(f|\mathscr{G})$ are $\mathscr{G}$-measurable, as in the proof of the second part of Lemma 2.2 this implies that $g = \mathbb{E}(f|\mathscr{G})$ almost surely.

In particular, $\mathbb{E}(\mathbb{E}(f|\mathscr{G})|\mathscr{G}) = \mathbb{E}(f|\mathscr{G})$ for all $f \in L^p(\Omega)$ and $\mathbb{E}(f|\mathscr{G}) = f$ for all $f \in L^p(\Omega, \mathscr{G})$. This shows that $\mathbb{E}(\cdot|\mathscr{G})$ is a projection onto $L^p(\Omega, \mathscr{G})$. $\square$

The next two results develop some properties of conditional expectations.

Proposition 2.6.

(1) *If $f \in L^1(\Omega)$ and $\mathscr{H}$ is a sub-σ-algebra of $\mathscr{G}$, then almost surely*

$$\mathbb{E}(\mathbb{E}(f|\mathscr{G})|\mathscr{H}) = \mathbb{E}(f|\mathscr{H}).$$

(2) *If $f \in L^1(\Omega)$ is independent of $\mathscr{G}$ (that is, f is independent of 1_G for all $G \in \mathscr{G}$), then almost surely*

$$\mathbb{E}(f|\mathscr{G}) = \mathbb{E}f.$$

(3) *If $f \in L^p(\Omega)$ and $g \in L^q(\Omega, \mathscr{G})$ with $1 \leqslant p, q \leqslant \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, then almost surely*

$$\mathbb{E}(gf|\mathscr{G}) = g\mathbb{E}(f|\mathscr{G}).$$

Proof. (1): For all $H \in \mathscr{H}$ we have $\int_H \mathbb{E}(\mathbb{E}(f|\mathscr{G})|\mathscr{H})\, d\mathbb{P} = \int_H \mathbb{E}(f|\mathscr{G})\, d\mathbb{P} = \int_H f\, d\mathbb{P}$ by Theorem 2.5, first applied to $\mathscr{H}$ and then to $\mathscr{G}$ (observe that $H \in \mathscr{G}$). Now the result follows from the uniqueness part of the theorem.

(2): For all $G \in \mathscr{G}$ we have $\int_G \mathbb{E}f\, d\mathbb{P} = \mathbb{E}1_G\mathbb{E}f = \mathbb{E}1_G f = \int_G f\, d\mathbb{P}$, and the result follows from the uniqueness part of Theorem 2.5.

(3): For all $G, G' \in \mathscr{G}$ we have $\int_G 1_{G'}\mathbb{E}(f|\mathscr{G})\, d\mathbb{P} = \int_{G\cap G'} \mathbb{E}(f|\mathscr{G})\, d\mathbb{P} = \int_{G\cap G'} f\, d\mathbb{P} = \int_G f 1_{G'}\, d\mathbb{P}$. Hence $\mathbb{E}(f1_{G'}|\mathscr{G}) = 1_{G'}\mathbb{E}(f|\mathscr{G})$ by the uniqueness part of Theorem 2.5. By linearity, this gives the result for simple functions g, and the general case follows by approximation. □

Example 2.7. Let $f : (-r, r) \to \mathbb{R}$ be integrable and let $\mathscr{G}$ be the σ-algebra of all symmetric Borel sets in $(-r, r)$ (we call a set B *symmetric* if $B = -B$). Then,

$$\mathbb{E}(f|\mathscr{G}) = \frac{1}{2}(f + \check{f}) \ \text{ almost surely,}$$

where $\check{f}(x) = f(-x)$. This is verified by checking the condition (2.1).

Example 2.8. Let $f : [0, 1) \to \mathbb{R}$ be integrable and let $\mathscr{G}$ be the σ-algebra generated by the sets $I_n := [\frac{n-1}{N}, \frac{n}{N})$, $n = 1, \ldots, N$. Then,

$$\mathbb{E}(f|\mathscr{G}) = \sum_{n=1}^{N} m_n\, 1_{[\frac{n-1}{N}, \frac{n}{N})} \ \text{ almost surely,}$$

where

$$m_n = \frac{1}{|I_n|} \int_{I_n} f(x)\, dx.$$

Again this is verified by checking the condition (2.1).

If f and g are random variables on $(\Omega, \mathscr{F}, \mathbb{P})$, we write $\mathbb{E}(g|f) := \mathbb{E}(g|\sigma(f))$, where $\sigma(f)$ is the σ-algebra generated by f.

Example 2.9. Let $f_1, \ldots, f_N$ be independent integrable random variables satisfying $\mathbb{E}(f_1) = \cdots = \mathbb{E}(f_N) = 0$, and put $S_N := f_1 + \cdots + f_N$. Then, for $1 \leqslant n \leqslant N$,

$$\mathbb{E}(S_N|f_n) = f_n \ \text{ almost surely.}$$

This follows from the following almost sure identities:

$$\mathbb{E}(S_N|f_n) = \mathbb{E}(f_n|f_n) + \sum_{m\neq n} \mathbb{E}(f_m|f_n) = f_n + \sum_{m\neq n} \mathbb{E}(f_m) = f_n.$$

Example 2.10. Let $f_1, \dots, f_N$ be independent and identically distributed integrable random variables and put $S_N := f_1 + \cdots + f_N$. Then

$$\mathbb{E}(f_1|S_N) = \cdots = \mathbb{E}(f_N|S_N) \quad \text{almost surely}$$

and therefore

$$\mathbb{E}(f_n|S_N) = S_N/N \quad \text{almost surely}, \quad n = 1, \dots, N.$$

It is clear that the second identity follows from the first. To prove the first, let μ denote the (common) distribution of the random variables f_n. By independence, the distribution of the $\mathbb{R}^N$-valued random variable $(f_1, \dots, f_N)$ equals the N-fold product measure $\mu^N := \mu \times \cdots \times \mu$. Then, for any Borel set $B \in \mathscr{B}(\mathbb{R})$ we have, using the substitution formula and Fubini's theorem,

$$\begin{aligned}
&\int_{\{S_N \in B\}} \mathbb{E}(f_n|S_N)\, d\mathbb{P} \\
&\quad = \int_{\{S_N \in B\}} f_n \, d\mathbb{P} \\
&\quad = \int_{\{x_1 + \cdots + x_N \in B\}} x_n \, d\mu^N(x) \\
&\quad = \int_{\mathbb{R}^{N-1}} \int_{\{y \in B - z_1 - \cdots - z_{N-1}\}} y \, d\mu(y)\, d\mu^{N-1}(z_1, \dots, z_{N-1}),
\end{aligned}$$

which is independent of n.

2.1 Notes

We recommend Williams [2] for an introduction to the subject. The bible for modern probabilists is Kallenberg [1].

The definition of conditional expectations starting from orthogonal projections in $L^2(\Omega)$ follows Kallenberg [1]. Many textbooks prefer the shorter (but less elementary and transparent) route via the Radon-Nikodým theorem. In this approach, one notes that if f is a random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ and $\mathscr{G}$ is a sub-σ-algebra of $\mathscr{F}$, then

$$\mathbb{Q}(G) := \int_G f \, d\mathbb{P}, \quad G \in \mathscr{G},$$

defines a probability measure $\mathbb{Q}$ on $(\Omega, \mathscr{G}, \mathbb{P}|_{\mathscr{G}})$ which is absolutely continuous with respect to $\mathbb{P}|_{\mathscr{G}}$. Hence by the Radon-Nikodým, $\mathbb{Q}$ has an integrable density with respect to $\mathbb{P}|_{\mathscr{G}}$. This density equals the conditional expectation $\mathbb{E}(f|\mathscr{G})$ almost surely.

References

1. O. Kallenberg, "Foundations of Modern Probability", second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002.
2. D. Williams, "Probability with Martingales", Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991.