PROBABILITY AND MEASURE

J. R. NORRIS

Contents

 Please send corrections to j.r.norris@statslab.cam.ac.uk.

Schedule

Measure spaces, σ-algebras, π-systems and uniqueness of extension, statement of Carathéodory's extension theorem. *Construction of Lebesgue measure on $\mathbb{R}$* Borel σ-algebra of $\mathbb{R}$, Lebesgue–Stieltjes measures and probability distribution functions. Independence of events, independence of σ-algebras. Borel–Cantelli lemmas. Kolmogorov's zero–one law.

Measurable functions, random variables, independence of random variables. Construction of the integral, expectation. Convergence in measure and convergence almost everywhere. Fatou's lemma, monotone and dominated convergence, uniform integrability, differentiation under the integral sign. Discussionof product measure and statement of Fubini's theorem.

Chebychev's inequality, tail estimates. Jensen's inequality. Completeness of L^p for $1 \leq p < \infty$. Hölder's and Minkowski's inequalities.

L^2 as a Hilbert space. Orthogonal projection, relation with elementary conditional probability. Variance and covariance. Gaussian random variables, the multivariate normal distribution.

The strong law of large numbers, proof for independent random variables with bounded fourth moments. Measure preserving transformations, Bernoulli shifts. Maximal ergodic theorem and Birkhoff's almost everywhere ergodic theorem, proof of the strong law.

The Fourier transform of a finite measure, characteristic functions, uniqueness and inversion. Weak convergence, statement of Lévy's continuity theorem for characteristic functions. The central limit theorem.

Appropriate books

P.Billingsley *Probability and Measure.* Wiley 1995 (£65.00 hardback).
R.M. Dudley *Real Analysis and Probability.* Chapman and Hall 1994 (£55.00 hardback).
D. Williams *Probability with Martingales.* Cambridge University Press 1991 (£15.95 paperback).

1. Measures

1.1. Let E be a set. A *σ-algebra* $\mathcal{E}$ is a set of subsets of E satisfying, for all $A \in \mathcal{E}$ and all sequences $(A_n : n \in \mathbb{N})$ in $\mathcal{E}$,

$$\emptyset \in \mathcal{E}, \quad A^c \in \mathcal{E}, \quad \cup_n A_n \in \mathcal{E}.$$

The pair $(E, \mathcal{E})$ is called a *measurable space*. Given $(E, \mathcal{E})$, each $A \in \mathcal{E}$ is called a *measurable set*.

It will be convenient, especially in setting up the basic theory, to have names for some other sorts of sets of subsets $\mathcal{A}$ of E, more general than σ-algebras. Say that $\mathcal{A}$ is a *π-system* if, for all $A, B \in \mathcal{A}$,

$$\emptyset \in \mathcal{A}, \quad A \cap B \in \mathcal{A}.$$

Say that $\mathcal{A}$ is a *d-system* if, for all $A, B \in \mathcal{A}$ with $A \subseteq B$ and all increasing sequences $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$,

$$E \in \mathcal{A}, \quad B \setminus A \in \mathcal{A}, \quad \cup_n A_n \in \mathcal{A}.$$

Say that $\mathcal{A}$ is a *ring* if, for all $A, B \in \mathcal{A}$,

$$\emptyset \in \mathcal{A}, \quad B \setminus A \in \mathcal{A}, \quad A \cup B \in \mathcal{A}.$$

Say that $\mathcal{A}$ is an *algebra* if, for all $A, B \in \mathcal{A}$,

$$\emptyset \in \mathcal{A}, \quad A^c \in \mathcal{A}, \quad A \cup B \in \mathcal{A}.$$

Note that $\mathcal{A}$ is a σ-algebra if and only if $\mathcal{A}$ is both a π-system and a d-system. In the case $E = \mathbb{R}$, the set of intervals of the form $(a, b]$ for $a < b$ is a π-system $\mathcal{I}$. The set of finite unions of disjoint elements of $\mathcal{I}$ is a ring but not an algebra.

1.2. The set of all subsets $\mathcal{P}(E)$ is obviously a σ-algebra. The intersection of any collection of σ-algebras is also a σ-algebra. Thus for any set of subsets $\mathcal{A}$, the intersection of all the σ-algebras containing $\mathcal{A}$ is itself a σ-algebra, called the σ-algebra *generated by* $\mathcal{A}$ and denoted $\sigma(\mathcal{A})$. If E is a topological space, we can do this to the set of open sets $\mathcal{T}$. The σ-algebra $\sigma(\mathcal{T})$ so obtained is called the *Borel* σ-algebra of E and is denoted $\mathcal{B}(E)$. The Borel σ-algebra of $\mathbb{R}$ is denoted simply by $\mathcal{B}$.

1.3. Let $\mathcal{A}$ be a ring of subsets of E. A *set function* is any function $\mu : \mathcal{A} \to [0, \infty]$ with $\mu(\emptyset) = 0$. Let μ be a set function. Say that μ is *increasing* if, for all $A, B \in \mathcal{A}$ with $A \subseteq B$,

$$\mu(A) \leq \mu(B).$$

Say that μ is *additive* if, for all disjoint sets $A, B \in \mathcal{A}$,

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

Say that μ is *countably additive* if, for all sequences of disjoint sets $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$ with $\cup_n A_n \in \mathcal{A}$,

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

Say that μ is *countably subadditive* if, for all sequences $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$ with $\cup_n A_n \in \mathcal{A}$,

$$\mu(\cup_n A_n) \leq \sum_n \mu(A_n).$$

On a ring $\mathcal{A}$, any additive set function is increasing.

1.4. Let $(E, \mathcal{E})$ be a measurable space. A countably additive set function $\mu : \mathcal{E} \to [0, \infty]$ is called a *measure*. The triple $(E, \mathcal{E}, \mu)$ is then called a *measure space*. If $\mu(E) = 1$ then μ is a *probability measure* and $(E, \mathcal{E}, \mu)$ is a *probability space*. The notation $(\Omega, \mathcal{F}, \mathbb{P})$ is often used to denote a probability space. If $\mu(E) < \infty$, then μ is a *finite* measure. If there exist sets $E_n \in \mathcal{E}, n \in \mathbb{N}$, with $\mu(E_n) < \infty$ for all n and $\cup_n E_n = E$, then μ is a σ-*finite* measure. If E is a topological space and if $\mathcal{E} = \mathcal{B}(E)$, then μ is a *Borel* measure on E; if moreover $\mu(K) < \infty$ for all compact sets K, then μ is a *Radon* measure.

1.5. **Discrete measure theory.** Let E be countable set and let $\mathcal{E} = \mathcal{P}(E)$. A *mass function* is any function $m : E \to [0, \infty]$. If μ is a measure on $(E, \mathcal{E})$, then, by countable additivity,

$$\mu(A) = \sum_{x \in A} \mu(\{x\}), \quad A \subseteq E.$$

So there is a one-to-one correspondence between measures and mass functions, given by

$$m(x) = \mu(\{x\}), \quad \mu(A) = \sum_{x \in A} m(x).$$

This sort of measure space provides a 'toy' version of the general theory, where each of the results we prove for general measure spaces reduces to some straightforward fact about the convergence of series. This is all one needs to do elementary discrete probability and discrete-time Markov chains, so these topics are usually introduced without discussing measure theory. Moreover measures associated with mass functions are essentially the only examples one can define explicitly: we turn now to the construction of more general measures.

1.6. **Construction and characterization of measures.** Most σ-algebras are so large that one cannot define measures directly. Instead one specifies the values to be taken on some smaller set of subsets, which generates the σ-algebra. This gives rise to two problems: first to know that there is a measure extending the given set function, second to know that there is not more than one. The first problem, which is one of construction, is dealt with by the following result. The proof is given towards the end of this section – however, once we have constructed our measures, the details of their construction prove to be largely irrelevant.

Theorem 1.6.1 (Carathéodory's extension theorem). *Let $\mathcal{A}$ be a ring of subsets of E and let $\mu : \mathcal{A} \to [0, \infty]$ be a countably additive set function. Then μ extends to a measure on the σ-algebra generated by $\mathcal{A}$.*

The key to the second problem, that of uniqueness, is provided by the following result. The proof is a piece of abstract nonsense, given at the end of this section.

Lemma 1.6.2 (Dynkin's π-system lemma). *Let $\mathcal{A}$ be a π-system. Then any d-system containing $\mathcal{A}$ contains also the σ-algebra generated by $\mathcal{A}$.*

Theorem 1.6.3 (Uniqueness of extension). *Let $\mathcal{A}$ be a π-system of subsets of E and let $\mathcal{E}$ be the σ-algebra generated by $\mathcal{A}$. Suppose that*

$$\mu_1 : \mathcal{E} \to [0, \infty], \quad \mu_2 : \mathcal{E} \to [0, \infty]$$

are measures with $\mu_1(E) = \mu_2(E) < \infty$. If $\mu_1 = \mu_2$ on $\mathcal{A}$, then $\mu_1 = \mu_2$ on $\mathcal{E}$.

Proof. Consider $\mathcal{D} = \{A \in \mathcal{E} : \mu_1(A) = \mu_2(A)\}$. By hypothesis, $E \in \mathcal{D}$; for $A, B \in \mathcal{E}$ with $A \subseteq B$, we have

$$\mu_1(A) + \mu_1(B \setminus A) = \mu_1(B) < \infty, \quad \mu_2(A) + \mu_2(B \setminus A) = \mu_2(B) < \infty$$

so, if $A, B \in \mathcal{D}$, then also $B \setminus A \in \mathcal{D}$; if $A_n \in \mathcal{D}, n \in \mathbb{N}$, with $A_n \uparrow A$, then

$$\mu_1(A) = \lim_n \mu_1(A_n) = \lim_n \mu_2(A_n) = \mu_2(A)$$

so $A \in \mathcal{D}$. Thus $\mathcal{D}$ is a d-system containing the π-system $\mathcal{A}$, so $\mathcal{D} = \mathcal{E}$ by Dynkin's lemma. □

Theorems 1.6.1 and 1.6.3 provide general tools for the construction and characterization of measures. We now apply them in a specific context.

Theorem 1.6.4. *There exists a unique Borel measure μ on $\mathbb{R}$ such that, for all $a, b \in \mathbb{R}$ with $a < b$,*

$$\mu((a, b]) = b - a.$$

The measure μ is called *Lebesgue measure* on $\mathbb{R}$.

Proof. (*Existence.*) Consider the ring $\mathcal{A}$ of finite unions of disjoint intervals of the form

$$A = (a_1, b_1] \cup \cdots \cup (a_n, b_n]$$

and define for $A \in \mathcal{A}$

$$\mu(A) = \sum_{i=1}^{n} (b_i - a_i).$$

Note that the presentation of A is not unique, as $(a, b] \cup (b, c] = (a, c]$ whenever $a < b < c$. Nevertheless, it is easy to check that μ is well-defined and additive. We aim to show that μ is countably additive on $\mathcal{A}$, which then proves existence by Carathéodory's extension theorem.

By additivity, it suffices to show that, if $A \in \mathcal{A}$ and if $(A_n : n \in \mathbb{N})$ is an increasing sequence in $\mathcal{A}$ with $A_n \uparrow A$, then $\mu(A_n) \to \mu(A)$. Set $B_n = A \setminus A_n$ then $B_n \in \mathcal{A}$ and $B_n \downarrow \emptyset$. By additivity again, it suffices to show that $\mu(B_n) \to 0$. Suppose, in fact, that for some $\varepsilon > 0$, we have $\mu(B_n) \geq 2\varepsilon$ for all n. For each n we can find $C_n \in \mathcal{A}$ with $\bar{C}_n \subseteq B_n$ and $\mu(B_n \setminus C_n) \leq \varepsilon 2^{-n}$. Then

$$\mu(B_n \setminus (C_1 \cap \cdots \cap C_n)) \leq \mu((B_1 \setminus C_1) \cup \cdots \cup (B_n \setminus C_n)) \leq \sum_{n \in \mathbb{N}} \varepsilon 2^{-n} = \varepsilon.$$

Since $\mu(B_n) \geq 2\varepsilon$, we must have $\mu(C_1 \cap \cdots \cap C_n) \geq \varepsilon$, so $C_1 \cap \cdots \cap C_n \neq \emptyset$, and so $K_n = \bar{C}_1 \cap \cdots \cap \bar{C}_n \neq \emptyset$. Now $(K_n : n \in \mathbb{N})$ is a sequence of bounded non-empty closed sets in $\mathbb{R}$, so $\emptyset \neq \cap_n K_n \subseteq \cap_n B_n$, which is a contradiction.
(*Uniqueness.*) Let μ be a measure on $\mathcal{B}$ with $\mu((a, b]) = b - a$ for all $a < b$. Fix n and consider

$$\mu_n(A) = \mu((n, n+1] \cap A).$$

Then μ_n is a probability measure on $\mathcal{B}$ so, by Theorem 1.6.3, μ_n is uniquely determined by its values on the π-system $\mathcal{I}$ generating $\mathcal{B}$. Since

$$\mu(A) = \sum_n \mu_n(A),$$

it follows that μ is also uniquely determined. □

Proof of Carathéodory's extension theorem. For any $B \subseteq E$, define the *outer measure*

$$\mu^*(B) = \inf \sum_n \mu(A_n)$$

where the infimum is taken over all sequences $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$ such that $B \subseteq \cup_n A_n$ and is taken to be ∞ if there is no such sequence. Note that μ^* is increasing and $\mu^*(\emptyset) = 0$. Let us say that $A \subseteq E$ is *μ^*-measurable* if, for all $B \subseteq E$,

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c).$$

Write $\mathcal{M}$ for the set of all μ^*-measurable sets. *We shall show that $\mathcal{M}$ is a σ-algebra containing $\mathcal{A}$ and that μ^* is a measure on $\mathcal{M}$, extending μ.* This will prove the theorem.

Step I. We show that μ^* is countably subadditive. Suppose that $B \subseteq \cup_n B_n$. If $\mu^*(B_n) < \infty$ for all n, then, given $\varepsilon > 0$, there exist sequences $(A_{nm} : m \in \mathbb{N})$ in $\mathcal{A}$, with

$$B_n \subseteq \cup_m A_{nm}, \quad \mu^*(B_n) + \varepsilon/2^n \geq \sum_m \mu(A_{nm}).$$

Then

$$B \subseteq \cup_n \cup_m A_{nm}$$

so

$$\mu^*(B) \leq \sum_n \sum_m \mu(A_{nm}) \leq \sum_n \mu^*(B_n) + \varepsilon.$$

Hence, in any case,

$$\mu^*(B) \leq \sum_n \mu^*(B_n).$$

Step II. We show that μ^* *extends* μ. Since $\mathcal{A}$ is a ring and μ is countably additive, μ is countably subadditive. Hence, for $A \in \mathcal{A}$ and any sequence $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$ with $A \subseteq \cup_n A_n$,

$$\mu(A) \leq \sum_n \mu(A_n).$$

On taking the infimum over all such sequences, we see that $\mu(A) \leq \mu^*(A)$. On the other hand, it is obvious that $\mu^*(A) \leq \mu(A)$ for $A \in \mathcal{A}$.
Step III. We show that $\mathcal{M}$ *contains* $\mathcal{A}$. Let $A \in \mathcal{A}$ and $B \subseteq E$. We have to show that

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c).$$

By subadditivity of μ^*, it is enough to show that

$$\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c).$$

If $\mu^*(B) = \infty$, this is trivial, so let us assume that $\mu^*(B) < \infty$. Then, given $\varepsilon > 0$, we can find a sequence $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$ such that

$$B \subseteq \cup_n A_n, \quad \mu^*(B) + \varepsilon \geq \sum_n \mu(A_n).$$

Then

$$B \cap A \subseteq \cup_n (A_n \cap A), \quad B \cap A^c \subseteq \cup_n (A_n \cap A^c)$$

so

$$\mu^*(B \cap A) + \mu^*(B \cap A^c) \leq \sum_n \mu(A_n \cap A) + \sum_n \mu(A_n \cap A^c) = \sum_n \mu(A_n) \leq \mu^*(B) + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we are done.
Step IV. We show that $\mathcal{M}$ *is an algebra.* Clearly $E \in \mathcal{M}$ and $A^c \in \mathcal{A}$ whenever $A \in \mathcal{A}$. Suppose that $A_1, A_2 \in \mathcal{M}$ and $B \subseteq E$. Then

$$\begin{aligned}
\mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\
&= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c) \\
&= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1^c) \\
&= \mu^*(B \cap (A_1 \cap A_2)) + \mu^*(B \cap (A_1 \cap A_2)^c).
\end{aligned}$$

Hence $A_1 \cap A_2 \in \mathcal{M}$.
Step V. We show that $\mathcal{M}$ *is a* σ*-algebra and that* μ^* *is a measure on* $\mathcal{M}$. We already know that $\mathcal{M}$ is an algebra, so it suffices to show that, for any sequence of disjoint sets $(A_n : n \in \mathbb{N})$ in $\mathcal{M}$, for $A = \cup_n A_n$ we have

$$A \in \mathcal{M}, \quad \mu^*(A) = \sum_n \mu^*(A_n).$$

So, take any $B \subseteq E$, then

$$\begin{aligned}\mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\ &= \mu^*(B \cap A_1) + \mu^*(B \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c) \\ &= \cdots = \sum_{i=1}^{n} \mu^*(B \cap A_i) + \mu^*(B \cap A_1^c \cap \cdots \cap A_n^c).\end{aligned}$$

Note that $\mu^*(B \cap A_1^c \cap \cdots \cap A_n^c) \geq \mu^*(B \cap A^c)$ for all n. Hence, on letting $n \to \infty$ and using countable subadditivity, we get

$$\mu^*(B) \geq \sum_{n=1}^{\infty} \mu^*(B \cap A_n) + \mu^*(B \cap A^c) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c).$$

The reverse inequality holds by subadditivity, so we have equality. Hence $A \in \mathcal{M}$ and, setting $B = A$, we get

$$\mu^*(A) = \sum_{n=1}^{\infty} \mu^*(A_n).$$

□

Proof of Dynkin's π-system lemma. Denote by $\mathcal{D}$ the intersection of all d-systems containing $\mathcal{A}$. Then $\mathcal{D}$ is itself a d-system. We shall show that $\mathcal{D}$ is also a π-system and hence a σ-algebra, thus proving the lemma. Consider

$$\mathcal{D}' = \{B \in \mathcal{D} : B \cap A \in \mathcal{D} \text{ for all } A \in \mathcal{A}\}.$$

Then $\mathcal{A} \subseteq \mathcal{D}'$ because $\mathcal{A}$ is a π-system. Let us check that $\mathcal{D}'$ is a d-system: clearly $E \in \mathcal{D}'$; next, suppose $B_1, B_2 \in \mathcal{D}'$ with $B_1 \subseteq B_2$, then for $A \in \mathcal{A}$ we have

$$(B_2 \setminus B_1) \cap A = (B_2 \cap A) \setminus (B_1 \cap A) \in \mathcal{D}$$

because $\mathcal{D}$ is a d-system, so $B_2 \setminus B_1 \in \mathcal{D}'$; finally, if $B_n \in \mathcal{D}', n \in \mathbb{N}$, and $B_n \uparrow B$, then for $A \in \mathcal{A}$ we have

$$B_n \cap A \uparrow B \cap A$$

so $B \cap A \in \mathcal{D}$ and $B \in \mathcal{D}'$. Hence $\mathcal{D} = \mathcal{D}'$.

Now consider

$$\mathcal{D}'' = \{B \in \mathcal{D} : B \cap A \in \mathcal{D} \text{ for all } A \in \mathcal{D}\}.$$

Then $\mathcal{A} \subseteq \mathcal{D}''$ because $\mathcal{D} = \mathcal{D}'$. We can check that $\mathcal{D}''$ is a d-system, just as we did for $\mathcal{D}'$. Hence $\mathcal{D}'' = \mathcal{D}$ which shows that $\mathcal{D}$ is a π-system as promised. □

1.7. **Independence.** A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ provides a model for an experiment whose outcome is subject to chance, according to the following interpretation:

- Ω is the set of possible outcomes
- $\mathcal{F}$ is the set of observable sets of outcomes, or *events*
- $\mathbb{P}(A)$ is the probability of the event A.

Relative to measure theory, probability theory is enriched by the significance attached to the notion of independence. Let I be a countable set. Say that events $A_i, i \in I$, are *independent* if, for all finite subsets $J \subseteq I$,

$$\mathbb{P}\left(\cap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i).$$

Say that σ-algebras $\mathcal{A}_i \subseteq \mathcal{F}, i \in I$, are *independent* if $A_i, i \in I$, are independent whenever $A_i \in \mathcal{A}_i$ for all i. Here is a useful way to establish the independence of two σ-algebras.

Theorem 1.7.1. *Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be π-systems contained in $\mathcal{F}$ and suppose that*

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$$

whenever $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. Then $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2)$ are independent.

Proof. Fix $A_1 \in \mathcal{A}_1$ and define for $A \in \mathcal{F}$

$$\mu(A) = \mathbb{P}(A_1 \cap A), \quad \nu(A) = \mathbb{P}(A_1)\mathbb{P}(A).$$

Then μ and ν are measures which agree on the π-system $\mathcal{A}_2$, with $\mu(\Omega) = \nu(\Omega) = \mathbb{P}(A_1) < \infty$. So, by uniqueness of extension, for all $A_2 \in \sigma(\mathcal{A}_2)$,

$$\mathbb{P}(A_1 \cap A_2) = \mu(A_2) = \nu(A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2).$$

Now fix $A_2 \in \sigma(\mathcal{A}_2)$ and repeat the argument with

$$\mu'(A) = \mathbb{P}(A \cap A_2), \quad \nu'(A) = \mathbb{P}(A)\mathbb{P}(A_2)$$

to show that, for all $A_1 \in \sigma(\mathcal{A}_1)$,

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2).$$

□

1.8. **Borel-Cantelli lemmas.** Given events $A_n, n \in \mathbb{N}$, we may ask for the probability that infinitely many occur. Set

$$\limsup A_n = \cap_n \cup_{m \geq n} A_m, \quad \liminf A_n = \cup_n \cap_{m \geq n} A_m.$$

Then

$$\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}(\limsup A_n), \quad \mathbb{P}(A_n \text{ ev.}) = \mathbb{P}(\liminf A_n).$$

Here 'i.o.' stands for 'infinitely often' and 'ev.' for 'eventually.'

Lemma 1.8.1 (First Borel–Cantelli lemma)**.** *If $\sum_n \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 0$.*

Proof. As $n \to \infty$ we have

$$\mathbb{P}(A_n \text{ i.o.}) \leq \mathbb{P}(\cup_{m \geq n} A_m) \leq \sum_{m \geq n} \mathbb{P}(A_m) \to 0.$$

□

We note that this argument is valid whether or not $\mathbb{P}$ is a probability measure.

Lemma 1.8.2 (Second Borel–Cantelli lemma)**.** *Assume that $A_n, n \in \mathbb{N}$, are independent events. If $\sum_n \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 1$.*

Proof. We use the inequality $1 - a \leq e^{-a}$. Set $a_n = \mathbb{P}(A_n)$. Then, for all n we have

$$\mathbb{P}(\cap_{m \geq n} A_m^c) = \prod_{m \geq n} (1 - a_m) \leq \exp\{-\sum_{m \geq n} a_m\} = 0.$$

Hence $\mathbb{P}(A_n \text{ i.o.}) = 1 - \mathbb{P}(\cup_n \cap_{m \geq n} A_m^c) = 1$. □

2. Measurable functions and random variables

2.1. Let $(E, \mathcal{E})$ and $(G, \mathcal{G})$ be measurable spaces. A function $f : E \to G$ is *measurable* if $f^{-1}(A) \in \mathcal{E}$ whenever $A \in \mathcal{G}$. Here $f^{-1}(A)$ denotes the *inverse image* of A by f

$$f^{-1}(A) = \{x \in E : f(x) \in A\}.$$

Usually $G = \mathbb{R}$ or $G = [-\infty, \infty]$, in which case $\mathcal{G}$ is always taken to be the Borel σ-algebra. If E is a topological space and $\mathcal{E} = \mathcal{B}(E)$, then a measurable function on E is called a *Borel* function.

The inverse image preserves set operations:

$$f^{-1}(\cup_i A_i) = \cup_i f^{-1}(A_i), \quad f^{-1}(A^c) = f^{-1}(A)^c.$$

Hence, for any function $f : E \to G$, $\{f^{-1}(A) : A \in \mathcal{G}\}$ is a σ-algebra on E and $\{A : f^{-1}(A) \in \mathcal{E}\}$ is a σ-algebra on G. In particular, if $\mathcal{G} = \sigma(\mathcal{A})$ and $f^{-1}(A) \in \mathcal{E}$ whenever $A \in \mathcal{A}$, then $\{A : f^{-1}(A) \in \mathcal{E}\}$ is a σ-algebra containing $\mathcal{A}$ and hence $\mathcal{G}$, so f is measurable.

In the case $G = \mathbb{R}$, the Borel σ-algebra is generated by intervals of the form $(-\infty, y], y \in \mathbb{R}$, so, to show that $f : E \to \mathbb{R}$ is measurable, it suffices to show that $\{x \in E : f(x) \leq y\} \in \mathcal{E}$ for all y.

If E is any topological space and $f : E \to \mathbb{R}$ is continuous, then $f^{-1}(U)$ is open in E and hence measurable, whenever U is open in $\mathbb{R}$; the open sets U generate $\mathcal{B}$, so *any continuous function is measurable.*

Note that *the indicator function of any measurable set is a measurable function.* Also, *the composition of measurable functions is measurable.*

Given any family of functions $f_i : E \to G, i \in I$, we can make them all measurable by taking

$$\mathcal{E} = \sigma(f_i^{-1}(A) : A \in \mathcal{G}, i \in I).$$

Then $\mathcal{E}$ is the *σ-algebra generated by* $(f_i : i \in I)$.

Proposition 2.1.1. *Let $f_n : E \to \mathbb{R}, n \in \mathbb{N}$, be measurable functions. Then so are $f_1 + f_2, f_1 f_2$ and each of the following:*

$$\inf_n f_n, \quad \sup_n f_n, \quad \liminf_n f_n, \quad \limsup_n f_n.$$

Theorem 2.1.2 (Monotone class theorem)**.** *Let $(E, \mathcal{E})$ be a measurable space and let $\mathcal{A}$ be a π-system generating $\mathcal{E}$. Let $\mathcal{V}$ be a vector space of bounded functions $f : E \to \mathbb{R}$ such that:*

(i) *$1 \in \mathcal{V}$ and $1_A \in \mathcal{V}$ for all $A \in \mathcal{A}$;*
(ii) *if $f_n \in \mathcal{V}$ for all n and f is bounded with $0 \leq f_n \uparrow f$, then $f \in \mathcal{V}$.*

Then $\mathcal{V}$ contains every bounded $\mathcal{E}$-measurable function.

Proof. Consider $\mathcal{D} = \{A \in \mathcal{E} : 1_A \in \mathcal{V}\}$. Then $\mathcal{D}$ is a d-system containing $\mathcal{A}$, so $\mathcal{D} = \mathcal{E}$. Since $\mathcal{V}$ is a vector space, it thus contains all finite linear combinations of indicator functions of measurable sets. Hence, for $f \geq 0$, $\mathcal{E}$-measurable and bounded, the functions $f_n = 2^{-n}[2^n f], n \in \mathbb{N}$, belong to $\mathcal{V}$ and $0 \leq f_n \uparrow f$, hence $f \in \mathcal{V}$. Finally, any bounded $\mathcal{E}$-measurable function is the difference of two non-negative such functions, hence in $\mathcal{V}$. □

2.2. **Image measures.** Let $(E, \mathcal{E})$ and $(G, \mathcal{G})$ be measurable spaces and let μ be a measure on $\mathcal{E}$. Then any measurable function $f : E \to G$ induces an *image measure* $\nu = \mu \circ f^{-1}$ on $\mathcal{G}$, given by

$$\nu(A) = \mu(f^{-1}(A)).$$

We shall construct some new measures from Lebesgue measure in this way.

Lemma 2.2.1. *Let $g : \mathbb{R} \to \mathbb{R}$ be a right-continuous increasing function. Set $I = (g(-\infty), g(\infty))$ and define $f : I \to \mathbb{R}$ by*

$$f(x) = \inf\{y : x \leq g(y)\}.$$

Then f is left-continuous and increasing, and

$$f(x) \leq y \quad \text{if and only if} \quad x \leq g(y).$$

Proof. The set $\{y : x \leq g(y)\}$ is decreasing in x. It is closed by our assumptions on g. Hence f is increasing and $g(f(x)) \geq x$. Also

$$f(g(y)) = \inf\{y' : g(y) \leq g(y')\} \leq y.$$

Now $f(x) \leq y$ implies $g(f(x)) \leq g(y)$, so $x \leq g(y)$. Similarly, $x \leq g(y)$ implies $f(x) \leq y$, so the two conditions are equivalent. In particular, it follows that $\{x \in I : f(x) \leq y\}$ is closed, so f is left-continuous. □

Theorem 2.2.2. *Let $g : \mathbb{R} \to \mathbb{R}$ be right-continuous and increasing. Then there exists a unique Radon measure dg on $\mathbb{R}$ such that, for all $a, b \in \mathbb{R}$ with $a < b$,*

$$dg((a, b]) = g(b) - g(a).$$

Moreover, we obtain in this way all Radon measures on $\mathbb{R}$.

The measure dg is called the *Lebesgue-Stieltjes measure* associated with g.

Proof. (*Existence.*) Set $I = (g(-\infty), g(\infty))$. Let μ denote Lebesgue measure on I and let $f : I \to \mathbb{R}$ be the left-continuous increasing function constructed in the lemma. Then f is Borel measurable and the induced measure $\nu = \mu \circ f^{-1}$ on $\mathbb{R}$ satisfies

$$\nu((a,b]) = \mu(\{x : f(x) > a \text{ and } f(x) \leq b\}) = \mu((g(a), g(b)]) = g(b) - g(a).$$

(*Uniqueness.*) The argument used for uniqueness of Lebesgue measure applies.

Finally, if ν is any Radon measure on $\mathbb{R}$, we can define $g : \mathbb{R} \to \mathbb{R}$, right-continuous and increasing, by

$$g(y) = \begin{cases} \nu((0,y]), & \text{if } y \geq 0, \\ -\nu((y,0]), & \text{if } y < 0. \end{cases}$$

Then $\nu((a,b]) = g(b) - g(a)$ whenever $a < b$, so $\nu = dg$ by uniqueness. □

2.3. **Random variables.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(E, \mathcal{E})$ be a measurable space. A measurable function $X : \Omega \to E$ is called a *random variable in* E. It has the interpretation of a quantity, or state, determined by chance. Where no space E is mentioned, it is assumed that X takes values in $\mathbb{R}$. The image measure $\mu_X = \mathbb{P} \circ X^{-1}$ is called the *law* or *distribution* of X. For real-valued random variables, μ_X is uniquely determined by its values on the π-system of intervals $(-\infty, x], x \in \mathbb{R}$, given by

$$F_X(x) = \mu_X((-\infty, x]) = \mathbb{P}(X \leq x).$$

The function F_X is called the *distribution function of* X.

Note that $F = F_X$ is increasing and right-continuous, with

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1.$$

Let us call any function $F : \mathbb{R} \to [0,1]$ satisfying these conditions a *distribution function*.

Set $\Omega = (0,1]$ and $\mathcal{F} = \mathcal{B}((0,1])$. Let $\mathbb{P}$ denote the restriction of Lebesgue measure to $\mathcal{F}$. Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. Let F be any distribution function. Define $X : \Omega \to \mathbb{R}$ by

$$X(\omega) = \inf\{x : \omega \leq F(x)\}.$$

Then, by Lemma 2.2.1, X is a random variable and $X(\omega) \leq x$ if and only if $\omega \leq F(x)$. So

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}((0, F(x)]) = F(x).$$

Thus every distribution function is the distribution function of a random variable.

We say that random variables $X_n, n \in \mathbb{N}$, are *independent* if the σ-algebras $\sigma(X_n)$ are independent. For real valued random variables, this is equivalent to the condition

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \dots \mathbb{P}(X_n \leq x_n)$$

for all $x_1, \dots, x_n \in \mathbb{R}$ and all n. A sequence of random variables $(X_n : n \geq 0)$ is often regarded as a *process* evolving in time. The σ-algebra generated by $X_0, \dots, X_n$

$$\mathcal{F}_n = \sigma(X_0, \dots, X_n)$$

contains those events depending (measurably) on $X_0, \dots, X_n$ and represents what is known about the process by time n.

2.4. **Rademacher functions.** We continue with the particular choice of probability space $(\Omega, \mathcal{F}, \mathbb{P})$ made in the preceding section. Provided that we forbid infinite sequences of 0's, each $\omega \in \Omega$ has a unique binary expansion

$$\omega = 0.\omega_1\omega_2\omega_3 \dots .$$

Define random variables $X_n : \Omega \to \{0, 1\}$ by $X_n(\omega) = \omega_n$. Then

$$X_1 = 1_{(\frac{1}{2},1]}, \quad X_2 = 1_{(\frac{1}{4},\frac{1}{2}]} + 1_{(\frac{3}{4},1]}, \quad X_3 = 1_{(\frac{1}{8},\frac{1}{4}]} + 1_{(\frac{3}{8},\frac{1}{2}]} + 1_{(\frac{5}{8},\frac{3}{4}]} + 1_{(\frac{7}{8},1]}.$$

These are called the *Rademacher functions*. The random variables $X_1, X_2, \dots$ are independent and *Bernoulli*, that is to say

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = 1/2.$$

The strong law of large numbers (proved in §10) applies here to show that

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} \to \frac{1}{2}\right) = 1.$$

This is called Borel's normal number theorem: *almost every point in* $(0, 1]$ *is* normal, *that is, has 'equal' proportions of* 0*'s and* 1*'s in its binary expansion.*

We now show how to construct not just one but an infinite sequence of independent random variables $X_1, X_2, \dots$, on $\Omega = (0, 1]$, having given distribution functions $F_1, F_2, \dots$. Choose a bijection $m : \mathbb{N}^2 \to \mathbb{N}$ and set $Y_{k,n} = X_{m(k,n)}$, where X_m is the mth Rademacher function. Set

$$Y_n = \sum_{k=1}^{\infty} 2^{-k} Y_{k,n}.$$

Then $Y_1, Y_2, \dots$ are independent and, for all n,

$$\mathbb{P}(i2^{-k} < Y_n \leq (i + 1)2^{-k}) = 2^{-k}$$

so $\mathbb{P}(Y_n \leq x) = x$ for all $x \in (0, 1]$. Set

$$G_n(y) = \inf\{x : y \leq F_n(x)\}$$

then, by Lemma 2.2.1, G_n is Borel and $G_n(y) \leq x$ if and only if $y \leq F_n(x)$. So, if we set $X_n = G_n(Y_n)$, then $X_1, X_2, \dots$ are independent random variables on Ω and

$$\mathbb{P}(X_n \leq x) = \mathbb{P}(G_n(Y_n) \leq x) = \mathbb{P}(Y_n \leq F_n(x)) = F_n(x).$$

2.5. **Tail events.** Let $(X_n : n \in \mathbb{N})$ be a sequence of random variables. Define

$$\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \dots), \quad \mathcal{T} = \cap_n \mathcal{T}_n.$$

Then $\mathcal{T}$ is a σ-algebra, called the *tail σ-algebra* of $(X_n : n \in \mathbb{N})$. It contains the events which depend only on the limiting behaviour of the sequence.

Theorem 2.5.1 (Kolmogorov's zero-one law)**.** *Suppose that $(X_n : n \in \mathbb{N})$ is a sequence of independent random variables. Then the tail σ-algebra $\mathcal{T}$ of $(X_n : n \in \mathbb{N})$ contains only events of probability* 0 *or* 1*. Moreover, any $\mathcal{T}$-measurable random variable is almost surely constant.*

Proof. Set $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Then $\mathcal{F}_n$ is generated by the π-system of events

$$A = \{X_1 \le x_1, \dots, X_n \le x_n\}$$

whereas $\mathcal{T}_n$ is generated by the π-system of events

$$B = \{X_{n+1} \le x_{n+1}, \dots, X_{n+k} \le x_{n+k}\}, \quad k \in \mathbb{N}.$$

We have $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all such A and B, by independence. Hence $\mathcal{F}_n$ and $\mathcal{T}_n$ are independent, by Theorem 1.7.1. It follows that $\mathcal{F}_n$ and $\mathcal{T}$ are independent. Now $\cup_n \mathcal{F}_n$ is a π-system which generates the σ-algebra $\mathcal{F}_\infty = \sigma(X_n : n \in \mathbb{N})$. So by Theorem 1.7.1 again, $\mathcal{F}_\infty$ and $\mathcal{T}$ are independent. But $\mathcal{T} \subseteq \mathcal{F}_\infty$. So, if $A \in \mathcal{T}$,

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A)$$

so $\mathbb{P}(A) \in \{0, 1\}$.

Finally, if Y is any $\mathcal{T}$-measurable random variable, then $F_Y(y) = \mathbb{P}(Y \le y)$ takes values in $\{0, 1\}$, so $\mathbb{P}(Y = c) = 1$, where $c = \inf\{y : F_Y(y) = 1\}$. □

2.6. **Convergence in measure and convergence almost everywhere.**
Let $(E, \mathcal{E}, \mu)$ be a measure space. A set $A \in \mathcal{E}$ is sometimes defined by a property shared by its elements. If $\mu(A^c) = 0$, then we say that property holds *almost everywhere* (or *a.e.*). The alternative *almost surely* (or *a.s.*) is often used in a probabilistic context. Thus, for a sequence of measurable functions $(f_n : n \in \mathbb{N})$, we say f_n *converges to f a.e.* to mean that

$$\mu(\{x : f_n(x) \not\to f(x)\}) = 0.$$

If, on the other hand, we have that

$$\mu(\{|f_n - f| > \varepsilon\}) \to 0, \quad \text{for all } \varepsilon > 0,$$

then we say f_n *converges to f in measure* or, in a probabilistic context, *in probability.*

Theorem 2.6.1. *Let $(f_n : n \in \mathbb{N})$ be a sequence of measurable functions.*

(a) *Assume that $\mu(E) < \infty$. If $f_n \to 0$ a.e. then $f_n \to 0$ in measure.*
(b) *If $f_n \to 0$ in measure then $f_{n_k} \to 0$ a.e. for some subsequence (n_k).*

Proof. (a) Suppose $f_n \to 0$ a.e.. For each $\varepsilon > 0$,

$$\mu(|f_n| \le \varepsilon) \ge \mu\left(\cap_{m\ge n}\{|f_m| \le \varepsilon\}\right) \uparrow \mu(|f_n| \le \varepsilon \text{ ev.}) \ge \mu(f_n \to 0) = \mu(E).$$

Hence $\mu(|f_n| > \varepsilon) \to 0$ and $f_n \to 0$ in measure. (b) Suppose $f_n \to 0$ in measure, then we can find a subsequence (n_k) such that

$$\sum_k \mu(|f_{n_k}| > 1/k) < \infty.$$

So, by the first Borel–Cantelli lemma,

$$\mu(|f_{n_k}| > 1/k \text{ i.o.}) = 0$$

so $f_{n_k} \to 0$ a.e.. □

3. Integration

3.1. **Definition and basic properties.** Let $(E, \mathcal{E}, \mu)$ be a measure space. We shall define, where possible, for a measurable function $f : E \to [-\infty, \infty]$, the *integral* of f, to be denoted

$$\mu(f) = \int_E f d\mu = \int_E f(x)\mu(dx).$$

For a random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the integral is usually called instead the *expectation* of X and written $\mathbb{E}(X)$.

A *simple* function is one of the form

$$f = \sum_{k=1}^m a_k 1_{A_k}$$

where $0 \le a_k \le \infty$ and $A_k \in \mathcal{E}$ for all k, and where $m \in \mathbb{N}$. For simple functions f, we define

$$\mu(f) = \sum_{k=1}^m a_k \mu(A_k),$$

where we adopt the conventions $\infty.0 = 0.\infty = 0$. Although the representation of f is not unique, it is straightforward to check that $\mu(f)$ is well defined and, for simple functions f, g and constants $\alpha, \beta \ge 0$, we have

(a) $\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g)$,
(b) $f \le g$ implies $\mu(f) \le \mu(g)$,
(c) $f = 0$ a.e. if and only if $\mu(f) = 0$.

In particular, for simple functions f, we have

$$\mu(f) = \sup\{\mu(g) : g \text{ simple}, \ g \le f\}.$$

We define the integral $\mu(f)$ of a non-negative measurable function f by

$$\mu(f) = \sup\{\mu(g) : g \text{ simple}, \ g \le f\}.$$

We have seen that this is consistent with our definition for simple functions. Note that, for all non-negative measurable functions f, g with $f \leq g$, we have $\mu(f) \leq \mu(g)$. For any measurable function f, we set

$$f^+ = f \vee 0, \quad f^- = (-f) \vee 0.$$

Then $f = f^+ - f^-$ and $|f| = f^+ + f^-$. If $\mu(|f|) < \infty$, then we say that f is *integrable* and define

$$\mu(f) = \mu(f^+) - \mu(f^-).$$

Note that $|\mu(f)| \leq \mu(|f|)$ for all integrable functions f. Note also that the integral $\mu(f)$ remains well defined, though possibly taking the value ∞, under the weaker condition that $\mu(f^-) < \infty$. For $A \in \mathcal{E}$, we write

$$\mu_A(f|_A) = \int_A f d\mu$$

where μ_A denotes the restriction of μ to A. It is easy to check that, whenever the integrals are defined,

$$\int_A f d\mu = \int_E f 1_A d\mu.$$

Here is the key result for the theory of integration.

Theorem 3.1.1 (Monotone convergence). *Let $(f_n : n \in \mathbb{N})$ be a sequence of non-negative measurable functions. Then*

$$f_n \uparrow f \quad \textit{implies} \quad \mu(f_n) \uparrow \mu(f).$$

Proof. Case 1: $f_n = 1_{A_n}, f = 1_A$.
The result is a simple consequence of countable additivity.
Case 2: *f_n simple, $f = 1_A$.*
Fix $\varepsilon > 0$ and set $A_n = \{f_n > 1 - \varepsilon\}$. Then $A_n \uparrow A$ and

$$(1 - \varepsilon) 1_{A_n} \leq f_n \leq 1_A$$

so

$$(1 - \varepsilon)\mu(A_n) \leq \mu(f_n) \leq \mu(A).$$

But $\mu(A_n) \uparrow \mu(A)$ by Case 1 and $\varepsilon > 0$ was arbitrary, so the result follows.

Case 3: *f_n simple, f simple.*
We can write f in the form

$$f = \sum_{k=1}^{m} a_k 1_{A_k}$$

with $a_k > 0$ for all k and the sets A_k disjoint. Then $f_n \uparrow f$ implies

$$a_k^{-1} 1_{A_k} f_n \uparrow 1_{A_k}$$

so, by Case 2,

$$\mu(f_n) = \sum_k \mu(1_{A_k} f_n) \uparrow \sum_k a_k \mu(A_k) = \mu(f).$$

Case 4: *f_n simple, $f \geq 0$ measurable.*
Let g be simple with $g \leq f$. Then $f_n \uparrow f$ implies $f_n \wedge g \uparrow g$ so, by Case 3, given $\varepsilon > 0$, for n sufficiently large

$$\mu(f_n) \geq \mu(f_n \wedge g) \geq \mu(g) - \varepsilon.$$

Since $\varepsilon > 0$ and g were arbitrary, the result follows.
Case 5: *$f_n \geq 0$ measurable, $f \geq 0$ measurable.*
Set

$$g_n = (2^{-n}[2^n f_n]) \wedge n$$

then g_n is simple and $g_n \leq f_n \leq f$, so

$$\mu(g_n) \leq \mu(f_n) \leq \mu(f).$$

But $f_n \uparrow f$ forces $g_n \uparrow f$, so $\mu(g_n) \uparrow \mu(f)$, by Case 4. □

Theorem 3.1.2. *For all non-negative measurable functions f, g and all constants $\alpha, \beta \geq 0$,*
(a) $\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g)$,
(b) $f \leq g$ *implies* $\mu(f) \leq \mu(g)$,
(c) $f = 0$ *a.e. if and only if* $\mu(f) = 0$.

Proof. Define simple functions f_n, g_n by

$$f_n = (2^{-n}[2^n f]) \wedge n, \quad g_n = (2^{-n}[2^n g]) \wedge n.$$

Then $f_n \uparrow f$ and $g_n \uparrow g$, so $\alpha f_n + \beta g_n \uparrow \alpha f + \beta g$. Hence, by monotone convergence,

$$\mu(f_n) \uparrow \mu(f), \quad \mu(g_n) \uparrow \mu(g), \quad \mu(\alpha f_n + \beta g_n) \uparrow \mu(\alpha f + \beta g).$$

We know that $\mu(\alpha f_n + \beta g_n) = \alpha\mu(f_n) + \beta\mu(g_n)$, so we obtain (a) on letting $n \to \infty$. As we noted above, (b) is obvious from the definition of the integral. If $f = 0$ a.e., then $f_n = 0$ a.e., for all n, so $\mu(f_n) = 0$ and $\mu(f) = 0$. On the other hand, if $\mu(f) = 0$, then $\mu(f_n) = 0$ for all n, so $f_n = 0$ a.e. and $f = 0$ a.e.. □

Theorem 3.1.3. *For all integrable functions f, g and all constants $\alpha, \beta \in \mathbb{R}$,*
(a) $\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g)$,
(b) $f \leq g$ *implies* $\mu(f) \leq \mu(g)$,
(c) $f = 0$ *a.e. implies* $\mu(f) = 0$.

Proof. We note that $\mu(-f) = -\mu(f)$. For $\alpha \geq 0$, we have

$$\mu(\alpha f) = \mu(\alpha f^+) - \mu(\alpha f^-) = \alpha\mu(f^+) - \alpha\mu(f^-) = \alpha\mu(f).$$

If $h = f + g$ then $h^+ + f^- + g^- = h^- + f^+ + g^+$, so

$$\mu(h^+) + \mu(f^-) + \mu(g^-) = \mu(h^-) + \mu(f^+) + \mu(g^+)$$

and so $\mu(h) = \mu(f)+\mu(g)$. That proves (a). If $f \leq g$ then $\mu(g)-\mu(f) = \mu(g-f) \geq 0$, by (a). Finally, if $f = 0$ a.e., then $f^{\pm} = 0$ a.e., so $\mu(f^{\pm}) = 0$ and so $\mu(f) = 0$. □

Note that in Theorem 3.1.3(c) we lose the reverse implication. The following result is sometimes useful:

Proposition 3.1.4. *Let $\mathcal{A}$ be a π-system containing E and generating $\mathcal{E}$. Then, for any integrable function f,*

$$\mu(f1_A) = 0 \text{ for all } A \in \mathcal{A} \quad \text{implies} \quad f = 0 \text{ a.e..}$$

Here are some minor variants on the monotone convergence theorem.

Proposition 3.1.5. *Let $(f_n : n \in \mathbb{N})$ be a sequence of measurable functions, with $f_n \geq 0$ a.e.. Then*

$$f_n \uparrow f \text{ a.e.} \quad \Longrightarrow \quad \mu(f_n) \uparrow \mu(f).$$

Thus the pointwise hypotheses of non-negativity and monotone convergence can be relaxed to hold almost everywhere.

Proposition 3.1.6. *Let $(g_n : n \in \mathbb{N})$ be a sequnce of non-negative measurable functions. Then*

$$\sum_{n=1}^{\infty} \mu(g_n) = \mu\left(\sum_{n=1}^{\infty} g_n\right).$$

This reformulation of monotone convergence makes it clear that it is the counterpart for the integration of functions of the countable additivity property of the measure on sets.

3.2. **Integrals and limits.** In the monotone convergence theorem, the hypothesis that the given sequence of functions is increasing is essential. In this section we obtain some results on the integrals of limits of functions without such a hypothesis.

Lemma 3.2.1 (Fatou's lemma). *Let $(f_n : n \in \mathbb{N})$ be a sequence of non-negative measurable functions. Then*

$$\mu(\liminf f_n) \leq \liminf \mu(f_n).$$

Proof. For $k \geq n$, we have

$$\inf_{m \geq n} f_m \leq f_k$$

so

$$\mu(\inf_{m \geq n} f_m) \leq \inf_{k \geq n} \mu(f_k) \leq \liminf \mu(f_n).$$

But, as $n \to \infty$,

$$\inf_{m \geq n} f_m \uparrow \sup_n \left(\inf_{m \geq n} f_m\right) = \liminf f_n$$

so, by monotone convergence,

$$\mu(\inf_{m\geq n} f_m) \uparrow \mu(\liminf f_n).$$

□

Theorem 3.2.2 (Dominated convergence). *Let $(f_n : n \in \mathbb{N})$ be a sequence of integrable functions with $f_n \to f$ pointwise as $n \to \infty$. Suppose that, for some integrable function g*

$$|f_n| \leq g, \quad \textit{for all } n.$$

Then f is integrable and $\mu(f_n) \to \mu(f)$ as $n \to \infty$.

Proof. The limit f is measurable and $|f| \leq g$, so $\mu(|f|) \leq \mu(g) < \infty$, so f is integrable. We have $0 \leq g \pm f_n \to g \pm f$ so certainly $\liminf(g \pm f_n) = g \pm f$. By Fatou's lemma,

$$\mu(g) + \mu(f) = \mu(\liminf(g + f_n)) \leq \liminf \mu(g + f_n) = \mu(g) + \liminf \mu(f_n),$$

$$\mu(g) - \mu(f) = \mu(\liminf(g - f_n)) \leq \liminf \mu(g - f_n) = \mu(g) - \limsup \mu(f_n).$$

Since $\mu(g) < \infty$, we can deduce that

$$\mu(f) \leq \liminf \mu(f_n) \leq \limsup \mu(f_n) \leq \mu(f).$$

This proves that $\mu(f_n) \to \mu(f)$ as $n \to \infty$. □

Theorem 3.2.3 (Differentiation under the integral sign). *Let $U \subseteq \mathbb{R}$ be open and suppose that $f : U \times E \to \mathbb{R}$ satisfies:*

(i) *$x \mapsto f(t, x)$ is integrable for all t,*
(ii) *$t \mapsto f(t, x)$ is differentiable for all x,*
(iii) *for some integrable function g, for all $x \in E$ and all $t \in U$,*

$$\left|\frac{\partial f}{\partial t}(t, x)\right| \leq g(x).$$

Then $x \mapsto (\partial f/\partial t)(t, x)$ is integrable for all t, the function $F : U \to \mathbb{R}$, defined by

$$F(t) = \int_E f(t, x)\mu(dx),$$

is differentiable and

$$\frac{d}{dt}F(t) = \int_E \frac{\partial f}{\partial t}(t, x)\mu(dx).$$

Proof. Take any sequence $h_n \to 0$ and set

$$g_n(x) = \frac{f(t + h_n, x) - f(t, x)}{h_n} - \frac{\partial f}{\partial t}(t, x).$$

Then $g_n(x) \to 0$ for all $x \in E$ and, by the mean value theorem, $|g_n| \le 2g$ for all n. In particular, $x \mapsto (\partial f/\partial t)(t,x)$ is measurable for all t and hence integrable, since $|(\partial f/\partial t)(t,x)| \le g(x)$ for all x. Then, by dominated convergence,

$$\frac{F(t+h_n)-F(t)}{h_n} - \int_E \frac{\partial f}{\partial t}(t,x)\mu(dx) = \int_E g_n(x)\mu(dx) \to 0.$$

□

3.3. **Integration and differentiation.** In this section we show that Lebesgue integration acts as an inverse to differentiation. For Lebesgue measure μ on $\mathbb{R}$, we write

$$\mu(f1_{(a,b]}) = \int_{(a,b]} f(x)dx = \int_a^b f(x)dx.$$

Theorem 3.3.1 (Fundamental theorem of calculus)**.**

(i) *Let $f : [a,b] \to \mathbb{R}$ be a continuous function and set*

$$F_a(t) = \int_a^t f(x)dx.$$

Then F_a is differentiable on $[a,b]$, with $F_a' = f$.

(ii) *Let $F : [a,b] \to \mathbb{R}$ be differentiable with continuous derivative f. Then*

$$\int_a^b f(x)dx = F(b) - F(a).$$

Proof. Fix $t \in [a,b)$. Given $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(x) - f(t)| \le \varepsilon$ whenever $|x - t| \le \delta$. So, for $0 < h \le \delta$,

$$\left|\frac{F_a(t+h)-F_a(t)}{h} - f(t)\right| = \frac{1}{h}\left|\int_t^{t+h}(f(x)-f(t))dx\right|$$
$$\le \frac{1}{h}\int_t^{t+h}|f(x)-f(t)|dx \le \frac{\varepsilon}{h}\int_t^{t+h}dx = \varepsilon.$$

Hence F_a is differentiable on the right at t with derivative $f(t)$. Similarly, for all $t \in (a,b]$, F_a is differentiable on the left at t with derivative $f(t)$. Finally, $(F - F_a)'(t) = 0$ for all $t \in (a,b)$ so, by the mean value theorem,

$$F(b) - F(a) = F_a(b) - F_a(a) = \int_a^b f(x)dx.$$

□

3.4. **Product measure and Fubini's theorem.** Let $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ be *finite* measure spaces. Then

$$\mathcal{A} = \{A_1 \times A_2 : A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2\}$$

is a π-system of subsets of $E = E_1 \times E_2$. Define the *product σ-algebra* $\mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2 = \sigma(\mathcal{A})$.

Lemma 3.4.1. *Let $f : E \to \mathbb{R}$ be $\mathcal{E}$-measurable. Then, for all $x_1 \in E_1$, the function $x_2 \mapsto f(x_1, x_2) : E_2 \to \mathbb{R}$ is $\mathcal{E}_2$-measurable.*

Proof. Denote by $\mathcal{V}$ the set of bounded $\mathcal{E}$-measurable functions for which the conclusion holds. Then $\mathcal{V}$ is a vector space, containing the indicator function 1_A of every set $A \in \mathcal{A}$. Moreover, if $f_n \in \mathcal{V}$ for all n and if f is bounded with $0 \leq f_n \uparrow f$, then also $f \in \mathcal{V}$. So, by the monotone class theorem, $\mathcal{V}$ contains all bounded $\mathcal{E}$-measurable functions. The rest is easy. □

Lemma 3.4.2. *For all bounded $\mathcal{E}$-measurable functions f, the function*

$$x_1 \mapsto f_1(x_1) = \int_{E_2} f(x_1, x_2)\mu_2(dx_2) : E_1 \to \mathbb{R}$$

is bounded and $\mathcal{E}_1$-measurable.

Proof. Apply the monotone class theorem, as in the preceding lemma. Note that finiteness of μ_1 and μ_2 is essential to the argument. □

Theorem 3.4.3 (Product measure)**.** *There exists a unique measure $\mu = \mu_1 \otimes \mu_2$ on $\mathcal{E}$ such that*

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

for all $A_1 \in \mathcal{E}_1$ and $A_2 \in \mathcal{E}_2$.

Proof. Uniqueness holds because $\mathcal{A}$ is a π-system generating $\mathcal{E}$. For existence, by the lemmas, we can define

$$\mu(A) = \int_{E_1} \left(\int_{E_2} 1_A(x_1, x_2)\mu_2(dx_2) \right) \mu_1(dx_1)$$

and use monotone convergence to see that μ is countably additive. □

Proposition 3.4.4. *Let $\hat{\mathcal{E}} = \mathcal{E}_2 \otimes \mathcal{E}_1$ and $\hat{\mu} = \mu_2 \otimes \mu_1$. For a function f on $E_1 \times E_2$, write $\hat{f}$ for the function on $E_2 \times E_1$ given by $\hat{f}(x_2, x_1) = f(x_1, x_2)$. Then $\hat{f}$ is $\hat{\mathcal{E}}$-measurable if and only if f is $\mathcal{E}$-measurable and, for f measurable and non-negative, $\hat{\mu}(\hat{f}) = \mu(f)$.*

Theorem 3.4.5 (Fubini's theorem)**.**

(a) *Let f be $\mathcal{E}$-measurable and non-negative. Then*

$$\mu(f) = \int_{E_1} \left(\int_{E_2} f(x_1, x_2)\mu_2(dx_2) \right) \mu_1(dx_1).$$

(b) *Let f be μ-integrable. Then*
  (i) $x_2 \mapsto f(x_1, x_2)$ *is μ_2-integrable for μ_1-almost all x_1,*
  (ii) $x_1 \mapsto \int_{E_2} f(x_1, x_2)\mu_2(dx_2)$ *is μ_1-integrable*
  and the formula for $\mu(f)$ in (a) *holds.*

Note that the *iterated integral* in (a) is well defined, for all bounded or non-negative measurable functions f, by Lemmas 3.4.1 and 3.4.2. Note also that, in combination wih Proposition 3.4.4, Fubini's theorem allows us to interchange the order of integration in multiple integrals,whenever the integrand is non-negative or μ-integrable.

Proof. Denote by $\mathcal{V}$ the set of all bounded $\mathcal{E}$-measurable functions f for which the formula holds. Then $\mathcal{V}$ contains the indicator function of every $\mathcal{E}$-measurable set so, by the monotone class theorem, $\mathcal{V}$ contains all bounded $\mathcal{E}$-measurable functions. Hence, for all $\mathcal{E}$-measurable functions f, we have

$$\mu(f_n) = \int_{E_1} \left(\int_{E_2} f_n(x_1, x_2)\mu_2(dx_2) \right) \mu_1(dx_1)$$

where $f_n = (-n) \vee f \wedge n$.

For f non-negative, we can pass to the limit as $n \to \infty$ by monotone convergence to extend the formula to f. That proves (a).

If f is μ-integrable, then, by (a)

$$\int_{E_1} \left(\int_{E_2} |f(x_1, x_2)|\mu_2(dx_2) \right) \mu_1(dx_1) = \mu(|f|) < \infty.$$

Hence we obtain (i) and (ii). Then, by dominated convergence, we can pass to the limit as $n \to \infty$ in the formula for $\mu(f_n)$ to obtain the desired formula for $\mu(f)$. □

The existence of product measure and Fubini's theorem extend easily to σ-finite measure spaces. The operation of taking the product of two measure spaces is associative, by a π-system uniqueness argument. So we can, by induction, take the product of a finite number, without specifying the order. The measure obtained by taking the n-fold product of Lebesgue measure on $\mathbb{R}$ is called *Lebesgue measure on* $\mathbb{R}^n$. The corresponding integral is written

$$\int_{\mathbb{R}^n} f(x)dx.$$

Proposition 3.4.6. *Let $(E, \mathcal{E}, \mu)$ be a σ-finite measure space and let f be a non-negative measurable function on E. Then*

$$\mu(f) = \int_0^\infty \mu(f \geq \lambda)d\lambda.$$

3.5. Transformations of integrals.

Proposition 3.5.1. *Let $(E, \mathcal{E}, \mu)$ and $(G, \mathcal{G}, \nu)$ be measure spaces. Suppose that $\nu = \mu \circ f^{-1}$ for some measurable function $f : E \to G$. Then $\nu(g) = \mu(g \circ f)$, for all non-negative measurable functions g on G.*

As a special case we get the formula

$$\mathbb{E}(g(X)) = \int_E g(x)\mu_X(dx)$$

for any random variable X in E and any non-negative measurable function g on E.

Proposition 3.5.2. *Let $(E, \mathcal{E}, \mu)$ be a measure space and let f be a non-negative measurable function with $\mu(f) < \infty$. Define $\nu(A) = \mu(f 1_A), A \in \mathcal{E}$. Then ν is a measure on E and $\nu(g) = \mu(fg)$ for all non-negative measurable functions g on E.*

A random variable X in $\mathbb{R}^n$ is said to have a *density function* f if, for all $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbb{P}(X \in A) = \int_A f(x)dx.$$

By Proposition 3.5.2, for a random variable X in $\mathbb{R}^n$ with density f_X,

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}^n} g(x) f_X(x)dx.$$

Proposition 3.5.3. *Let $\phi : [a, b] \to \mathbb{R}$ be continuously differentiable and strictly increasing. Then, for all non-negative measurable functions g on $[\phi(a), \phi(b)]$,*

$$\int_{\phi(a)}^{\phi(b)} g(y)dy = \int_a^b g(\phi(x))\phi'(x)dx.$$

It is a simple matter to extend each of the above propositions to all integrable functions g.

4. Norms and inequalities

4.1. L^p-norms.

Let $(E, \mathcal{E}, \mu)$ be a measure space. For $1 \leq p < \infty$, we denote by $L^p = L^p(E, \mathcal{E}, \mu)$ the set of measurable functions f with finite L^p-*norm*:

$$\|f\|_p = \left(\int_E |f|^p d\mu\right)^{1/p} < \infty.$$

We denote by $L^\infty = L^\infty(E, \mathcal{E}, \mu)$ the set of measurable functions f with finite L^∞-*norm*:

$$\|f\|_\infty = \inf\{\lambda : |f| \leq \lambda \text{ a.e.}\}.$$

Note that $\|f\|_p \leq \|f\|_\infty$ for all $1 \leq p < \infty$. For $1 \leq p \leq \infty$, we say that f_n *converges to f in L^p* if $\|f_n - f\|_p \to 0$.

4.2. **Chebyshev's inequality.** Let f be a non-negative measurable function and let $\lambda \geq 0$. We use the notation $\{f \geq \lambda\}$ for the set $\{x \in E : f(x) \geq \lambda\}$. Note that

$$\lambda 1_{\{f \geq \lambda\}} \leq f$$

so on integrating we obtain *Chebyshev's inequality*

$$\lambda \mu(f \geq \lambda) \leq \mu(f).$$

Now let g be any measurable function. We can deduce inequalities for g by choosing some non-negative measurable function ϕ and applying Chebyshev's inequality to $f = \phi \circ g$. For example, if $g \in L^p$ and $\lambda > 0$, then

$$\mu(|g| \geq \lambda) = \mu(|g|^p \geq \lambda^p) \leq \lambda^{-p} \mu(|g|^p) < \infty.$$

So we obtain the *tail estimate*

$$\mu(|g| \geq \lambda) = O(\lambda^{-p}), \quad \text{as } \lambda \to \infty.$$

4.3. **Jensen's inequality.** A function $c : \mathbb{R} \to \mathbb{R}$ is *convex* if, for all $x, y \in \mathbb{R}$ and $t \in [0, 1]$,

$$c(tx + (1-t)y) \leq tc(x) + (1-t)c(y).$$

Lemma 4.3.1. *Let $c : \mathbb{R} \to \mathbb{R}$ be convex and let $m \in \mathbb{R}$. Then there exist $a, b \in \mathbb{R}$ such $c(x) \geq ax + b$ for all x, with equality at $x = m$.*

Proof. By convexity, for $x < m < y$ we have

$$\frac{c(m) - c(x)}{m - x} \leq \frac{c(y) - c(m)}{y - m}.$$

So, fixing m, there exists $a \in \mathbb{R}$ such that, for all $x < m$ and all $y > m$

$$\frac{c(m) - c(x)}{m - x} \leq a \leq \frac{c(y) - c(m)}{y - m}.$$

Then $c(x) \geq a(x - m) + c(m)$, for all $x \in \mathbb{R}$. □

Theorem 4.3.2 (Jensen's inequality)**.** *Let X be an integrable random variable and let $c : \mathbb{R} \to \mathbb{R}$ be convex. Then $\mathbb{E}(c(X))$ is well defined and*

$$\mathbb{E}(c(X)) \geq c(\mathbb{E}(X)).$$

Proof. Set $m = \mathbb{E}(X)$ and choose $a, b \in \mathbb{R}$ as in the lemma. Then $c(X) \geq aX + b$. In particular $\mathbb{E}(c(X)^-) \leq |a|\mathbb{E}(|X|) + |b| < \infty$, so $\mathbb{E}(c(X))$ is well defined. Moreover

$$\mathbb{E}(c(X)) \geq a\mathbb{E}(X) + b = am + b = c(m) = c(\mathbb{E}(X)).$$

□

We deduce from Jensen's inequality *the monotonicity of L^p-norms with respect to a probability measure.* Let $1 \leq p < q < \infty$. Set $c(x) = |x|^{q/p}$, then c is convex. So, for any $X \in L^p(\mathbb{P})$,

$$\|X\|_p = (\mathbb{E}|X|^p)^{1/p} = (c(\mathbb{E}|X|^p))^{1/q} \leq (\mathbb{E}c(|X|^p))^{1/q} = (\mathbb{E}|X|^q)^{1/q} = \|X\|_q.$$

In particular, $L^p(\mathbb{P}) \supseteq L^q(\mathbb{P})$.

4.4. **Hölder's inequality and Minkowski's inequality.** Let $(E, \mathcal{E}, \mu)$ be a measure space. For $p, q \in [1, \infty]$, we say that p and q are *conjugate indices* if

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Theorem 4.4.1 (Hölder's inequality). *Let $p, q \in (1, \infty)$ be conjugate indices. Then, for all measurable functions f and g, we have*

$$\mu(|fg|) \leq \|f\|_p \|g\|_q.$$

Proof. The cases where $\|f\|_p = 0$ or $\|f\|_p = \infty$ are trivial so we can exclude them. Then, by multiplying f by an appropriate constant, we are reduced to the case where $\|f\|_p = 1$. So we can define a probability measure $\mathbb{P}$ on $\mathcal{E}$ by

$$\mathbb{P}(A) = \int_A |f|^p d\mu.$$

For measurable functions $X \geq 0$, we have

$$\mathbb{E}(X) = \mu(X|f|^p), \quad \mathbb{E}(X) \leq \mathbb{E}(X^q)^{1/q}.$$

Note that $q(p-1) = p$. Then we have

$$\begin{aligned}\mu(|fg|) &= \mu\left(\frac{|g|}{|f|^{p-1}}|f|^p\right) = \mathbb{E}\left(\frac{|g|}{|f|^{p-1}}\right) \\ &\leq \mathbb{E}\left(\frac{|g|^q}{|f|^{q(p-1)}}\right)^{1/q} = \mu(|g|^q)^{1/q} = \|f\|_p\|g\|_q.\end{aligned}$$

□

Theorem 4.4.2 (Minkowski's inequality). *For $p \in [1, \infty)$ and measurable functions f and g, we have*

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

Proof. The case $p = 1$ is easy. The cases where $\|f\|_p = \infty$ or $\|g\|_p = \infty$ are trivial, so we can assume that both norms are finite. Then, since $|f + g|^p \leq 2^p(|f|^p + |g|^p)$, we have

$$\mu(|f + g|^p) \leq 2^p\{\mu(|f|^p) + \mu(|g|^p)\} < \infty.$$

The case where $\|f + g\|_p = 0$ is trivial, so let us assume $\|f + g\|_p > 0$. Observe that

$$\||f + g|^{p-1}\|_q = \mu(|f + g|^{(p-1)q})^{1/q} = \mu(|f + g|^p)^{1-1/p}.$$

So, by Hölder's inequality,

$$\mu(|f+g|^p) \leq \mu(|f||f+g|^{p-1}) + \mu(|g||f+g|^{p-1})$$
$$\leq (\|f\|_p + \|g\|_p)\||f+g|^{p-1}\|_q.$$

The result follows on dividing both sides by $\||f+g|^{p-1}\|_q$. □

Minkowski's inequality shows that L^p-norms are continuous: if $f_n \to f$ in L^p, then

$$\|f\|_p - \|f_n - f\|_p \leq \|f_n\|_p \leq \|f\|_p + \|f_n - f\|_p$$

so $\|f_n\|_p \to \|f\|_p$.

5. Completeness of L^p and orthogonal projection

5.1. $\mathcal{L}^p$ as a Banach space. Let V be a vector space. A map $v \mapsto \|v\| : V \to [0,\infty)$ is a *norm* if

(i) $\|u+v\| \leq \|u\| + \|v\|$ for all $u, v \in V$,
(ii) $\|\alpha v\| = |\alpha|\|v\|$ for all $v \in V$ and $\alpha \in \mathbb{R}$,
(iii) $\|v\| = 0$ implies $v = 0$.

We note that, for any norm, if $\|v_n - v\| \to 0$ then $\|v_n\| \to \|v\|$.

A symmetric bilinear map $(u,v) \mapsto \langle u, v\rangle : V \times V \to \mathbb{R}$ is an *inner product* if $\langle v, v\rangle \geq 0$, with equality only if $v = 0$. For any inner product, $\langle .,.\rangle$, the map $v \mapsto \sqrt{\langle v, v\rangle}$ is a norm, by the Cauchy–Schwarz inequality.

Minkowski's inequality shows that each L^p space is a vector space and that the L^p-norms satisfy condition (i) above. Condition (ii) also holds. Condition (iii) fails, because $\|f\|_p = 0$ does not imply that $f = 0$, only that $f = 0$ a.e.. However, it is possible to make the L^p-norms into true norms by quotienting out by the subspace of measurable functions vanishing a.e.. This quotient will be denoted $\mathcal{L}^p$. Note that, for $f \in L^2$, we have $\|f\|_2^2 = \langle f, f\rangle$, where $\langle .,.\rangle$ is the symmetric bilinear form on L^2 given by

$$\langle f, g\rangle = \int_E fg d\mu.$$

Thus $\mathcal{L}^2$ is an inner product space. The notion of convergence in L^p defined in §4.1 is the usual notion of convergence in a normed space.

A normed vector space V is *complete* if every Cauchy sequence in V converges, that is to say, given any sequence $(v_n : n \in \mathbb{N})$ in V such that $\|v_n - v_m\| \to 0$ as $n, m \to \infty$, there exists $v \in V$ such that $\|v_n - v\| \to 0$ as $n \to \infty$. A complete normed vector space is called a *Banach space*. A complete inner product space is called a *Hilbert space*. Such spaces have many useful properties, which makes the following result important.

Theorem 5.1.1 (Completeness of L^p)**.** *Let $p \in [1,\infty]$. Let $(f_n : n \in \mathbb{N})$ be a sequence in L^p such that*

$$\|f_n - f_m\|_p \to 0 \quad \text{as } n, m \to \infty.$$

Then there exists $f \in L^p$ such that

$$\|f_n - f\|_p \to 0 \quad \textit{as } n \to \infty.$$

Proof. The case $p = \infty$ is left as an exercise. We assume from now on that $p < \infty$. Choose a subsequence (n_k) such that

$$S = \sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p < \infty.$$

By Minkowski's inequality, for any $K \in \mathbb{N}$,

$$\|\sum_{k=1}^{K} |f_{n_{k+1}} - f_{n_k}|\|_p \leq S.$$

By monotone convergence this bound holds also for $K = \infty$, so

$$\sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| < \infty \quad \text{a.e..}$$

Hence, by completeness of $\mathbb{R}$, f_{n_k} converges a.e.. We define

$$f(x) = \begin{cases} \lim f_{n_k}(x) & \text{if the limit exists,} \\ 0 & \text{otherwise.} \end{cases}$$

Given $\varepsilon > 0$, we can find N so that $n \geq N$ implies

$$\mu(|f_n - f_m|^p) \leq \varepsilon, \quad \text{for all } m \geq n,$$

in particular $\mu(|f_n - f_{n_k}|^p) \leq \varepsilon$ for all sufficiently large k. Hence, by Fatou's lemma, for $n \geq N$,

$$\mu(|f_n - f|^p) = \mu(\liminf_k |f_n - f_{n_k}|^p) \leq \liminf_k \mu(|f_n - f_{n_k}|^p) \leq \varepsilon.$$

Hence $f \in L^p$ and, since $\varepsilon > 0$ was arbitrary, $\|f_n - f\|_p \to 0$. □

Corollary 5.1.2. *We have*

(a) $\mathfrak{L}^p$ *is a Banach space, for all* $1 \leq p \leq \infty$,
(b) $\mathfrak{L}^2$ *is a Hilbert space.*

5.2. **$\mathfrak{L}^2$ as a Hilbert space.** We shall apply some general Hilbert space arguments to L^2. First, we note *Pythagoras' rule*

$$\|f + g\|_2^2 = \|f\|_2^2 + 2\langle f, g\rangle + \|g\|_2^2$$

and the *parallelogram law*

$$\|f + g\|_2^2 + \|f - g\|_2^2 = 2(\|f\|_2^2 + \|g\|_2^2).$$

If $\langle f, g\rangle = 0$, then we say that f and g are *orthogonal*. For any subset $V \subseteq L^2$, we define

$$V^\perp = \{f \in L^2 : \langle f, v\rangle = 0 \text{ for all } v \in V\}.$$

A subset $V \subseteq L^2$ is *closed* if, for every sequence $(f_n : n \in \mathbb{N})$ in V, with $f_n \to f$ in L^2, we have $f = v$ a.e., for some $v \in V$.

Theorem 5.2.1 (Orthogonal projection). *Let V be a closed subspace of L^2. Then each $f \in L^2$ has a decomposition $f = v + u$, with $v \in V$ and $u \in V^\perp$. Moreover, $\|f - v\|_2 \leq \|f - g\|_2$ for all $g \in V$, with equality only if $g = v$ a.e..*

The function v is called (*a version of*) the *orthogonal projection of f on V*.

Proof. Choose a sequence $g_n \in V$ such that

$$\|f - g_n\|_2 \to d(f, V) = \inf\{\|f - g\|_2 : g \in V\}.$$

By the parallelogram law,

$$\|2(f - (g_n + g_m)/2)\|_2^2 + \|g_n - g_m\|_2^2 = 2(\|f - g_n\|_2^2 + \|f - g_m\|_2^2).$$

But $\|2(f-(g_n+g_m)/2)\|_2^2 \geq 4d(f, V)^2$, so we must have $\|g_n - g_m\|_2 \to 0$ as $n, m \to \infty$. By completeness, $\|g_n - g\|_2 \to 0$, for some $g \in L^2$. By closure, $g = v$ a.e., for some $v \in V$. Hence

$$\|f - v\|_2 = \lim_n \|f - g_n\|_2 = d(f, V).$$

Now, for any $h \in V$ and $t \in \mathbb{R}$, we have

$$d(f, V)^2 \leq \|f - (v + th)\|_2^2 = d(f, V)^2 - 2t\langle f - v, h\rangle + t^2\|h\|_2^2.$$

So we must have $\langle f - v, h\rangle = 0$. Hence $u = f - v \in V^\perp$, as required. □

5.3. **Variance, covariance and conditional expectation.** In this section we look at some L^2 notions relevant to probability. For $X, Y \in L^2(\mathbb{P})$, with means $m_X = \mathbb{E}(X), m_Y = \mathbb{E}(Y)$, we define *variance*, *covariance* and *correlation* by

$$\mathrm{var}(X) = \mathbb{E}[(X - m_X)^2],$$
$$\mathrm{cov}(X, Y) = \mathbb{E}[(X - m_X)(Y - m_y)],$$
$$\mathrm{corr}(X, Y) = \mathrm{cov}(X, Y)/\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}.$$

Note that $\mathrm{var}(X) = 0$ if and only if $X = m_X$ a.s.. Note also that, if X and Y are independent, then $\mathrm{cov}(X, Y) = 0$. The converse is generally false. For a random variable $X = (X_1, \ldots, X_n)$ in $\mathbb{R}^n$, we define its *covariance matrix*

$$\mathrm{var}(X) = (\mathrm{cov}(X_i, X_j))_{i,j=1}^n.$$

Proposition 5.3.1. *Every covariance matrix is non-negative definite.*

Suppose now we are given a countable family of disjoint events $(G_i : i \in I)$, whose union is Ω. Set $\mathcal{G} = \sigma(G_i : i \in I)$. Let X be an integrable random variable. The *conditional expectation* of X given $\mathcal{G}$ is given by

$$Y = \sum_i \mathbb{E}(X|G_i)1_{G_i},$$

where we set $\mathbb{E}(X|G_i) = \mathbb{E}(X1_{G_i})/\mathbb{P}(G_i)$ when $\mathbb{P}(G_i) > 0$ and define $\mathbb{E}(X|G_i)$ in some arbitrary way when $\mathbb{P}(G_i) = 0$. Set $V = L^2(\mathcal{G}, \mathbb{P})$ and note that $Y \in V$. Then V is a subspace of $L^2(\mathcal{F}, \mathbb{P})$, and V is complete and therefore closed.

Proposition 5.3.2. *If $X \in L^2$, then Y is a version of the orthogonal projection of X on V.*

6. Convergence in $L^1(\mathbb{P})$

6.1. **Bounded convergence.** We begin with a basic, but easy to use, condition for convergence in $L^1(\mathbb{P})$.

Theorem 6.1.1 (Bounded convergence). *Let $(X_n : n \in \mathbb{N})$ be a sequence of random variables, with $X_n \to X$ in probability and $|X_n| \leq C$ for all n, for some constant $C < \infty$. Then $X_n \to X$ in L^1.*

Proof. By Theorem 2.6.1, X is the almost sure limit of a subsequence, so $|X| \leq C$ a.s.. For $\varepsilon > 0$, there exists N such that $n \geq N$ implies

$$\mathbb{P}(|X_n - X| > \varepsilon/2) \leq \varepsilon/(4C).$$

Then

$$\mathbb{E}|X_n - X| = \mathbb{E}(|X_n - X|1_{|X_n - X|>\varepsilon/2}) + \mathbb{E}(|X_n - X|1_{|X_n - X|\leq\varepsilon/2}) \leq 2C(\varepsilon/4C) + \varepsilon/2 = \varepsilon.$$

□

6.2. **Uniform integrability.**

Lemma 6.2.1. *Let X be an integrable random variable and set*

$$I_X(\delta) = \sup\{\mathbb{E}(|X|1_A) : A \in \mathcal{F}, \mathbb{P}(A) \leq \delta\}.$$

Then $I_X(\delta) \downarrow 0$ as $\delta \downarrow 0$.

Proof. Suppose not. Then, for some $\varepsilon > 0$, there exist $A_n \in \mathcal{F}$, with $\mathbb{P}(A_n) \leq 2^{-n}$ and $\mathbb{E}(|X|1_{A_n}) \geq \varepsilon$ for all n. By the first Borel–Cantelli lemma, $\mathbb{P}(A_n \text{ i.o.}) = 0$. But then, by dominated convergence,

$$\varepsilon \leq \mathbb{E}(|X|1_{\cup_{m \geq n} A_m}) \to \mathbb{E}(|X|1_{\{A_n \text{ i.o.}\}}) = 0$$

which is a contradiction. □

Let $\mathcal{X}$ be a family of random variables. For $1 \leq p \leq \infty$, we say that $\mathcal{X}$ is *bounded in L^p* if

$$\sup_{X \in \mathcal{X}} \|X\|_p < \infty.$$

Let us define

$$I_{\mathcal{X}}(\delta) = \sup\{\mathbb{E}(|X|1_A) : X \in \mathcal{X}, A \in \mathcal{F}, \mathbb{P}(A) \leq \delta\}.$$

Obviously, $\mathcal{X}$ is bounded in L^1 if and only if $I_{\mathcal{X}}(1) < \infty$. We say that $\mathcal{X}$ is *uniformly integrable* or *UI* if $\mathcal{X}$ is bounded in L^1 and

$$I_{\mathcal{X}}(\delta) \downarrow 0, \quad \text{as } \delta \downarrow 0.$$

Note that, by Hölder's inequality, for conjugate indices $p, q \in (1, \infty)$,

$$\mathbb{E}(|X|1_A) \leq \|X\|_p (\mathbb{P}(A))^{1/q}.$$

Hence, if $\mathcal{X}$ is bounded in L^p, for some $p \in (1, \infty)$, then $\mathcal{X}$ is UI. The sequence $X_n = n1_{(0,1/n)}$ is bounded in L^1 for Lebesgue measure on $(0, 1]$, but not uniformly integrable.

Lemma 6.2.1 shows that any single integrable random variable is uniformly integrable. This extends easily to any finite collection of integrable random variables. Moreover, for any integrable random variable Y, the set

$$\mathcal{X} = \{X : X \text{ a random variable}, |X| \leq Y\}$$

is uniformly integrable, because

$$\mathbb{E}(|X|1_A) \leq \mathbb{E}(Y1_A), \quad \text{for all } A.$$

The following result provides an alternative characterization of uniform integrability.

Proposition 6.2.2. *Let $\mathcal{X}$ be a family of random variables. Then $\mathcal{X}$ is UI if and only if*

$$\sup\{\mathbb{E}(|X|1_{|X|\geq K}) : X \in \mathcal{X}\} \to 0, \quad \text{as } K \to \infty.$$

Here is the definitive result on L^1-convergence of random variables.

Theorem 6.2.3. *Let $X_n, n \in \mathbb{N}$, and X be random variables. The following are equivalent:*

(a) *$X_n \in L^1$ for all n, $X \in L^1$ and $X_n \to X$ in L^1,*
(b) *$\{X_n : n \in \mathbb{N}\}$ is UI and $X_n \to X$ in probability.*

Proof. Suppose (a) holds. By Chebyshev's inequality, for $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon^{-1}\mathbb{E}(|X_n - X|) \to 0$$

so $X_n \to X$ in probability. Moreover, given $\varepsilon > 0$, there exists N such that $\mathbb{E}(|X_n - X|) < \varepsilon/2$ whenever $n \geq N$. Then we can find $\delta > 0$ so that $\mathbb{P}(A) \leq \delta$ implies

$$\mathbb{E}(|X|1_A) \leq \varepsilon/2, \quad \mathbb{E}(|X_n|1_A) \leq \varepsilon, \quad n = 1, \ldots, N.$$

Then, for $n \geq N$ and $\mathbb{P}(A) \leq \delta$,

$$\mathbb{E}(|X_n|1_A) \leq \mathbb{E}(|X_n - X|) + \mathbb{E}(|X|1_A) \leq \varepsilon.$$

Hence $\{X_n : n \in \mathbb{N}\}$ is UI. We have shown that (a) implies (b).

Suppose, on the other hand, that (b) holds. Then there is a subsequence (n_k) such that $X_{n_k} \to X$ a.s.. So, by Fatou's lemma, $\mathbb{E}(|X|) \le \liminf_k \mathbb{E}(|X_{n_k}|) < \infty$. Now, given $\varepsilon > 0$, there exists $K < \infty$ such that, for all n,

$$\mathbb{E}(|X_n|1_{|X_n|\ge K}) < \varepsilon/3, \quad \mathbb{E}(|X|1_{|X|\ge K}) < \varepsilon/3.$$

Consider the uniformly bounded sequence $X_n^K = (-K) \vee X_n \wedge K$ and set $X^K = (-K) \vee X \wedge K$. Then $X_n^K \to X^K$ in probability, so, by bounded convergence, there exists N such that, for all $n \ge N$,

$$\mathbb{E}|X_n^K - X^K| < \varepsilon/3.$$

But then, for all $n \ge N$,

$$\mathbb{E}|X_n - X| \le \mathbb{E}(|X_n|1_{|X_n|\ge K}) + \mathbb{E}|X_n^K - X^K| + \mathbb{E}(|X|1_{|X|\ge K}) < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we have shown that (b) implies (a). □

7. Characteristic functions

7.1. **Definitions.** For a finite Borel measure μ on $\mathbb{R}^n$, we define the *Fourier transform*

$$\hat{\mu}(u) = \int_{\mathbb{R}^n} e^{i\langle u,x\rangle}\mu(dx), \quad u \in \mathbb{R}^n.$$

Here, $\langle .,. \rangle$ denotes the usual inner product on $\mathbb{R}^n$. For a random variable X in $\mathbb{R}^n$, we define the *characteristic function*

$$\phi_X(u) = \mathbb{E}(e^{i\langle u,X\rangle}), \quad u \in \mathbb{R}^n.$$

Thus $\phi_X = \hat{\mu}_X$, where μ_X is the law of X.

A random variable X in $\mathbb{R}^n$ is *standard Gaussian* if

$$\mathbb{P}(X \in A) = \int_A \frac{1}{(2\pi)^{n/2}} e^{-|x|^2/2} dx, \quad A \in \mathcal{B}.$$

Let us compute the characteristic function of a standard Gaussian random variable X in $\mathbb{R}$. We have

$$\phi_X(u) = \int_{\mathbb{R}} e^{iux} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{-u^2/2} I$$

where

$$I = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-(x-iu)^2/2} dx.$$

The integral I can be evaluated by considering the integral of the analytic function $e^{-z^2/2}$ around the rectangular contour with corners $R, R - iu, -R - iu, -R$: by Cauchy's theorem, the integral round the contour vanishes, as do, in the limit $R \to \infty$, the contributions from the vertical sides of the rectangle. We deduce that

$$I = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1.$$

Hence $\phi_X(u) = e^{-u^2/2}$.

7.2. **Uniqueness and inversion.** Our aim in this section is to show that a finite Borel measure is determined uniquely by its Fourier transform and to obtain, where possible an *inversion* formula by which to compute the measure from its transform. To this end, we define, for $t > 0$ and $x, y \in \mathbb{R}^n$, the *heat kernel*

$$p(t, x, y) = \frac{1}{(2\pi t)^{n/2}} e^{-|y-x|^2/2t}.$$

We note from the preceding section the identity

$$e^{-w^2/2} = \int_{\mathbb{R}} e^{iuw} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

By setting $w = (x - y)/\sqrt{t}$ and making a simple change of variable, we deduce, for $n = 1$,

$$p(t, x, y) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{iux} e^{-u^2 t/2} e^{-iuy} du.$$

Hence we obtain, for $n \geq 1$,

$$p(t, x, y) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\langle u,x\rangle} e^{-|u|^2 t/2} e^{-i\langle u,y\rangle} du.$$

Theorem 7.2.1. *Let X be a random variable in $\mathbb{R}^n$. The law μ_X of X is uniquely determined by its characteristic function ϕ_X. Moreover, if ϕ_X is integrable, then X has a density function $f_X(x)$, given by*

$$f_X(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \phi_X(u) e^{-i\langle u,x\rangle} du.$$

Proof. Let Y be a standard Gaussian random variable in $\mathbb{R}^n$, independent of X, and let f be a bounded Borel function on $\mathbb{R}^n$. Then, for $t > 0$, by a change of variable $y' = y - x/\sqrt{t}$ andFubini's theorem,

$$\begin{aligned}
\mathbb{E}(f(X + \sqrt{t}Y)) &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x + \sqrt{t}y)(2\pi)^{-n/2} e^{-|y|^2/2} dy \mu_X(dx) \\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} p(t, x, y) f(y) dy \mu_X(dx) \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\langle u,x\rangle} e^{-|u|^2 t/2} e^{-i\langle u,y\rangle} du \mu_X(dx) \right) f(y) dy \\
&= \int_{\mathbb{R}^n} \left(\frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \phi_X(u) e^{-|u|^2 t/2} e^{-i\langle u,y\rangle} du \right) f(y) dy.
\end{aligned}$$

By this formula, ϕ_X determines $\mathbb{E}(f(X+\sqrt{t}Y))$. For any bounded continuous function f, we have

$$\mathbb{E}(f(X+\sqrt{t}Y)) \to \mathbb{E}(f(X))$$

as $t \downarrow 0$, so ϕ_X determines $\mathbb{E}(f(X))$. Hence ϕ_X determines μ_X.

If ϕ_X is integrable and if f is continuous with compact support, then

$$|\phi_X(u)||f(y)| \in L^1(du \otimes dy).$$

So, by dominated convergence, as $t \downarrow 0$,

$$\int_{\mathbb{R}^n} \left(\frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \phi_X(u) e^{-|u|^2 t/2} e^{-i\langle u,y\rangle} du \right) f(y) dy$$
$$\to \int_{\mathbb{R}^n} \left(\frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \phi_X(u) e^{-i\langle u,y\rangle} du \right) f(y) dy.$$

Hence X has the claimed density function. □

7.3. Characteristic functions and independence.

Theorem 7.3.1. *Let $X = (X_1, \dots, X_n)$ be a random variable in $\mathbb{R}^n$. Then the following are equivalent:*

(a) *$X_1, \dots, X_n$ are independent,*
(b) $\mu_X = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$,
(c) $\mathbb{E}(\prod_k f_k(X_k)) = \prod_k \mathbb{E}(f_k(X_k))$, for all bounded Borel functions $f_1, \dots, f_n$,
(d) *$\phi_X(u) = \prod_k \phi_{X_k}(u_k)$, for all $u = (u_1, \dots, u_n) \in \mathbb{R}^n$.*

Proof. If (a) holds, then

$$\mu_X(A_1 \times \cdots \times A_n) = \prod_k \mu_{X_k}(A_k)$$

for all Borel sets $A_1, \dots, A_n$, so (b) holds, since this formula characterizes the product measure.

If (b) holds, then, for $f_1, \dots, f_n$ bounded Borel,

$$\mathbb{E}(\prod_k f_k(X_k)) = \int_{\mathbb{R}^n} \prod_k f_k(x_k) \mu_X(dx) = \prod_k \int_{\mathbb{R}} f_k(x_k) \mu_{X_k}(dx_k) = \prod_k \mathbb{E}(f_k(X_k)),$$

so (c) holds. Statement (d) is a special case of (c). Suppose, finally, that (d) holds and take independent random variables $\tilde{X}_1, \dots, \tilde{X}_n$ with $\mu_{\tilde{X}_k} = \mu_{X_k}$ for all k. We know that (a) implies (d), so

$$\phi_{\tilde{X}}(u) = \prod_k \phi_{\tilde{X}_k}(u_k) = \prod_k \phi_{X_k}(u_k) = \phi_X(u)$$

so $\mu_{\tilde{X}} = \mu_X$ by uniqueness of characteristic functions. Hence (a) holds. □

8. Gaussian random variables

8.1. **Gaussian random variables in $\mathbb{R}$.** A random variable X in $\mathbb{R}$ is *Gaussian* if, for some $\mu \in \mathbb{R}$ and some $\sigma^2 \in (0, \infty)$, X has density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}.$$

We also admit as Gaussian any random variable X with $X = \mu$ a.s., this degenerate case corresponding to taking $\sigma^2 = 0$. We write $X \sim N(\mu, \sigma^2)$.

Proposition 8.1.1. *Suppose $X \sim N(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$. Then*
(a) $\mathbb{E}(X) = \mu$,
(b) $\mathrm{var}(X) = \sigma^2$,
(c) $aX + b \sim N(a\mu + b, a^2\sigma^2)$,
(d) $\phi_X(u) = e^{iu\mu - u^2\sigma^2/2}$.

8.2. **Gaussian random variables in $\mathbb{R}^n$.** A random variable X in $\mathbb{R}^n$ is *Gaussian* if $\langle u, X\rangle$ is Gaussian, for all $u \in \mathbb{R}^n$. An example of such a random variable is provided by $X = (X_1, \dots, X_n)$, where $X_1, \dots, X_n$ are independent $N(0, 1)$ random variables. To see this, we note that

$$\mathbb{E} e^{i\langle u, X\rangle} = \mathbb{E} \prod_k e^{iu_k X_k} = e^{-|u|^2/2}$$

so $\langle u, X\rangle$ is $N(0, |u|^2)$ for all $u \in \mathbb{R}^n$.

Theorem 8.2.1. *Let X be a Gaussian random variable in $\mathbb{R}^n$. Then*
(a) *$AX + b$ is Gaussian, for all $n \times n$ matrices A and all $b \in \mathbb{R}^n$,*
(b) *$X \in L^2$ and its distribution is determined by its mean μ and its covariance matrix Σ,*
(c) $\phi_X(u) = e^{i\langle u, \mu\rangle - \langle u, \Sigma u\rangle/2}$,
(d) *if Σ is invertible, then X has a density function on $\mathbb{R}^n$, given by*

$$f_X(x) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\{-\langle x - \mu, \Sigma^{-1}(x - \mu)\rangle/2\},$$

(e) *if $X = (Y, Z)$, with Y in $\mathbb{R}^m$ and Z in $\mathbb{R}^p$, then*

$$\mathrm{cov}(Y, Z) = 0 \quad \textit{implies} \quad Y, Z \textit{ independent.}$$

Proof. For $u \in \mathbb{R}^n$, we have

$$\langle u, AX + b\rangle = \langle A^T u, X\rangle + \langle u, b\rangle$$

so $\langle u, AX + b\rangle$ is Gaussian, by Proposition 8.1.1. This proves (a).

Each component X_k is Gaussian, so $X \in L^2$. Set $\mu = \mathbb{E}(X)$ and $\Sigma = \mathrm{var}(X)$. For $u \in \mathbb{R}^n$ we have $\mathbb{E}(\langle u, Y\rangle) = \langle u, \mu\rangle$ and

$$\mathrm{var}(\langle u, X\rangle) = \mathrm{cov}(\langle u, X\rangle, \langle u, X\rangle) = \langle u, \Sigma u\rangle.$$

Since $\langle u, X\rangle$ is Gaussian, by Proposition 8.1.1, we must have $\langle u, X\rangle \sim N(\langle u, \mu\rangle, \langle u, \sigma u\rangle)$ and

$$\phi_X(u) = \mathbb{E}e^{i\langle u,X\rangle} = e^{i\langle u,\mu\rangle - \langle u,\Sigma u\rangle/2}.$$

This is (c) and (b) follows by uniqueness of characteristic functions.

Let $Y_1, \dots, Y_n$ be independent $N(0,1)$ random variables. Then $Y = (Y_1, \dots, Y_n)$ has density

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^n}} \exp\{-|y|^2/2\}.$$

Set $\tilde{X} = \Sigma^{1/2}Y + \mu$, then $\tilde{X}$ is Gaussian, with $\mathbb{E}(\tilde{X}) = \mu$ and $\text{var}(\tilde{X}) = \Sigma$, so $\tilde{X} \sim X$. If Σ is invertible, then $\tilde{X}$ and hence X has the density claimed in (d), by a linear change of variables in $\mathbb{R}^n$.

Finally, if $X = (Y, Z)$ with $\text{cov}(Y, Z) = 0$, then, for $v \in \mathbb{R}^m$ and $w \in \mathbb{R}^p$,

$$\langle (v, w), \Sigma(v, w)\rangle = \langle v, \Sigma_Y v\rangle + \langle w, \Sigma_Z w\rangle,$$

where $\Sigma_Y = \text{var}(Y)$ and $\Sigma_Z = \text{var}(Z)$. The joint characteristic function $\phi_{Y,Z}$ then splits as a product

$$\phi_{Y,Z}(v, w) = e^{i\langle v,\mu_Y\rangle - \langle v,\Sigma_Y v\rangle/2} e^{i\langle w,\mu_Z\rangle - \langle w,\Sigma_Z w\rangle/2}$$

so Y and Z are independent by Theorem 7.3.1. □

9. Ergodic theory

9.1. **Measure-preserving transformations.** Let $(E, \mathcal{E}, \mu)$ be a σ-finite measure space. A measurable function $\theta : E \to E$ is called a *measure-preserving transformation* if

$$\mu(\theta^{-1}(A)) = \mu(A), \quad \text{for all } A \in \mathcal{E}.$$

A set $A \in \mathcal{E}$ is *invariant* if $\theta^{-1}(A) = A$. A measurable function f is *invariant* if $f = f \circ \theta$. The class of all invariant sets forms a σ-algebra, which we denote by $\mathcal{E}_\theta$. Then f is invariant if and only if f is $\mathcal{E}_\theta$-measurable. We say that θ is *ergodic* if $\mathcal{E}_\theta$ contains only sets of measure zero and their complements.

Here are two simple examples of measure preserving transformations.

(i) *Translation map on the torus.* Take $E = (0,1]^n$ with Lebesgue measure and addition modulo 1 in each coordinate. For $a \in E$ set

$$\theta_a(x_1, \dots, x_n) = (x_1 + a_1, \dots, x_n + a_n).$$

(ii) *Bakers' map.* Take $E = (0,1]$ with Lebesgue measure. Set

$$\theta(x) = 2x - [2x].$$

Proposition 9.1.1. *If f is integrable and θ is measure-preserving, then $f \circ \theta$ is integrable and*

$$\int_E f d\mu = \int_E f \circ \theta \, d\mu.$$

Proposition 9.1.2. *If θ is ergodic and f is invariant, then $f = c$ a.e., for some constant c.*

9.2. **Bernoulli shifts.** Let m be a probability measure on $\mathbb{R}$. In §2.4, we constructed a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which there exists a sequence of independent random variables $(Y_n : n \in \mathbb{N})$, all having distribution m. Consider now the infinite product space

$$E = \mathbb{R}^{\mathbb{N}} = \{x = (x_n : n \in \mathbb{N}) : x_n \in \mathbb{R} \text{ for all } n\}$$

and the σ-algebra $\mathcal{E}$ on E generated by the coordinate maps $X_n(x) = x_n$

$$\mathcal{E} = \sigma(X_n : n \in \mathbb{N}).$$

Note that $\mathcal{E}$ is also generated by the π-system

$$\mathcal{A} = \{\prod_{n\in\mathbb{N}} A_n : A_n \in \mathcal{B} \text{ for all } n,\ A_n = \mathbb{R} \text{ for sufficiently large } n\}.$$

Define $Y : \Omega \to E$ by $Y(\omega) = (Y_n(\omega) : n \in \mathbb{N})$. Then Y is measurable and the image measure $\mu = \mathbb{P} \circ Y^{-1}$ satisfies, for $A = \prod_{n\in\mathbb{N}} A_n \in \mathcal{A}$,

$$\mu(A) = \prod_{n\in\mathbb{N}} m(A_n).$$

By uniqueness of extension, μ is the unique measure on $\mathcal{E}$ having this property. Note that, under the probability measure μ, the coordinate maps $(X_n : n \in \mathbb{N})$ are themselves a sequence of independent random variables with law m. The probability space $(E, \mathcal{E}, \mu)$ is called the *canonical model* for such sequences. Define the *shift map* $\theta : E \to E$ by

$$\theta(x_1, x_2, \dots) = (x_2, x_3, \dots).$$

Theorem 9.2.1. *The shift map is an ergodic measure-preserving transformation.*

Proof. The details of showing that θ is measurable and measure-preserving are left as an exercise. To see that θ is ergodic, we recall the definition of the tail σ-algebras

$$\mathcal{T}_n = \sigma(X_m : m \geq n+1), \quad \mathcal{T} = \cap_n \mathcal{T}_n.$$

For $A = \prod_{n\in\mathbb{N}} A_n \in \mathcal{A}$ we have

$$\theta^{-n}(A) = \{X_{n+k} \in A_k \text{ for all } k\} \in \mathcal{T}_n.$$

Since $\mathcal{T}_n$ is a σ-algebra, it follows that $\theta^{-n}(A) \in \mathcal{T}_n$ for all $A \in \mathcal{E}$, so $\mathcal{E}_\theta \subseteq \mathcal{T}$. Hence θ is ergodic by Kolmogorov's zero-one law. □

9.3. **Birkhoff's and von Neumann's ergodic theorems.** Throughout this section, $(E, \mathcal{E}, \mu)$ will denote a σ-finite measure space, on which is given a measure-preserving transformation θ. We fix an integrable function f, set $S_0 = 0$ and, for $n \geq 1$,

$$S_n = S_n(f) = f + f \circ \theta + \cdots + f \circ \theta^{n-1}.$$

Lemma 9.3.1 (Maximal ergodic lemma)**.** *Let $S^* = \sup_n S_n$. Then*

$$\int_{\{S^*>0\}} f d\mu \geq 0.$$

Proof. Set $S_n^* = \max_{0\leq m\leq n} S_m$ and $A_n = \{S_n^* > 0\}$. Then, for $m = 1, \dots, n$,

$$S_m = f + S_{m-1} \circ \theta \leq f + S_n^* \circ \theta.$$

On A_n we have $S_n^* = \max_{1\leq m\leq n} S_m$, so

$$S_n^* \leq f + S_n^* \circ \theta.$$

On A_n^c we have

$$S_n^* = 0 \leq S_n^* \circ \theta.$$

So, integrating and adding, we obtain

$$\int_E S_n^* d\mu \leq \int_{A_n} f d\mu + \int_E S_n^* \circ \theta d\mu.$$

But S_n^* is integrable, so

$$\int_E S_n^* \circ \theta d\mu = \int_E S_n^* d\mu < \infty$$

which forces

$$\int_{A_n} f d\mu \geq 0.$$

As $n \to \infty$, $A_n \uparrow \{S^* > 0\}$ so, by dominated convergence,

$$\int_{\{S^*>0\}} f d\mu \geq 0.$$

□

Theorem 9.3.2 (Birkhoff's almost everywhere ergodic theorem)**.** *There exists an invariant function $\bar{f}$, with $\mu(|\bar{f}|) \leq \mu(|f|)$, such that $S_n/n \to \bar{f}$ a.e. as $n \to \infty$.*

Proof. The functions $\liminf_n(S_n/n)$ and $\limsup_n(S_n/n)$ are invariant. Therefore, for $a < b$, so is the following set

$$D = D(a, b) = \{\liminf_n(S_n/n) < a < b < \limsup_n(S_n/n)\}.$$

We shall show that $\mu(D) = 0$. First, by invariance, we can restrict everything to D and thereby reduce to the case $D = E$. Note that either $b > 0$ or $a < 0$. We can interchange the two cases by replacing f by $-f$. Let us assume then that $b > 0$.

Let $B \in \mathcal{E}$ with $\mu(B) < \infty$, then $g = f - b1_B$ is integrable and, for each $x \in D$, for some n,

$$S_n(g)(x) \geq S_n(f)(x) - nb > 0.$$

Hence $S^*(g) > 0$ everywhere and, by the maximal ergodic lemma,

$$0 \leq \int_D (f - b1_B) d\mu = \int_D f d\mu - b\mu(B).$$

Since μ is σ-finite, we can let $B \uparrow D$ to obtain

$$b\mu(D) \leq \int_D f d\mu.$$

In particular, we see that $\mu(D) < \infty$. A similar argument applied to $-f$ and $-a$, this time with $B = D$, shows that

$$(-a)\mu(D) \leq \int_D (-f) d\mu.$$

Hence

$$b\mu(D) \leq \int_D f d\mu \leq a\mu(D).$$

Since $a < b$ and the integral is finite, this forces $\mu(D) = 0$. Set

$$\Delta = \{\liminf_n (S_n/n) < \limsup_n (S_n/n)\}$$

then Δ is invariant. Also, $\Delta = \cup_{a,b \in \mathbb{Q}, a<b} D(a,b)$, so $\mu(\Delta) = 0$. On the complement of Δ, S_n/n converges in $[-\infty, \infty]$, so we can define an invariant function $\bar{f}$ by

$$\bar{f} = \begin{cases} \lim_n (S_n/n) & \text{on } \Delta^c \\ 0 & \text{on } \Delta. \end{cases}$$

Finally, we have $\mu(|f \circ \theta^n|) = \mu(|f|)$, so $\mu(|S_n|) \leq n\mu(|f|)$ for all n. Hence, by Fatou's lemma,

$$\mu(|\bar{f}|) = \mu(\liminf_n |S_n/n|) \leq \liminf_n \mu(|S_n/n|) \leq \mu(|f|).$$

□

Theorem 9.3.3 (von Neumann's L^p ergodic theorem)**.** *Assume that $\mu(E) < \infty$. Let $p \in [1, \infty)$. Then, for $f \in L^p$, $S_n/n \to \bar{f}$ in L^p.*

Proof. We have

$$\|f \circ \theta^n\|_p = \left(\int_E |f|^p \circ \theta^n d\mu \right)^{1/p} = \|f\|_p.$$

So, by Minkowski's inequality,

$$\|S_n(f)/n\|_p \leq \|f\|_p.$$

Given $\varepsilon > 0$, choose $K < \infty$ so that $\|f - g\|_p < \varepsilon/3$, where $g = (-K) \vee f \wedge K$. By Birkhoff's theorem, $S_n(g)/n \to \bar{g}$ a.e.. We have $|S_n(g)/n| \leq K$ for all n so, by bounded convergence, there exists N such that, for $n \geq N$,

$$\|S_n(g)/n - \bar{g}\|_p < \varepsilon/3.$$

By Fatou's lemma,

$$\begin{aligned}\|\bar{f} - \bar{g}\|_p^p &= \int_E \liminf_n |S_n(f-g)/n|^p d\mu \\ &\leq \liminf_n \int_E |S_n(f-g)/n|^p d\mu \leq \|f-g\|_p^p.\end{aligned}$$

Hence, for $n \geq N$,

$$\begin{aligned}\|S_n(f)/n - \bar{f}\|_p &\leq \|S_n(f-g)/n\|_p + \|S_n(g)/n - \bar{g}\|_p + \|\bar{g} - \bar{f}\|_p \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.\end{aligned}$$

□

10. Sums of independent random variables

10.1. **Strong law of large numbers for finite fourth moment.** The result we obtain in this section will be largely superseded in the next. We include it because its proof is much more elementary than that needed for the definitive version of the strong law which follows.

Theorem 10.1.1. *Let $(X_n : n \in \mathbb{N})$ be a sequence of independent random variables such that, for some constants μ and M*

$$\mathbb{E}(X_n) = \mu, \quad \mathbb{E}(X_n^4) \leq M \quad \textit{for all } n.$$

Set $S_n = X_1 + \cdots + X_n$. Then

$$S_n/n \to \mu \quad \textit{a.s., as } n \to \infty.$$

Proof. Consider $Y_n = X_n - \mu$. Then

$$\mathbb{E}(Y_n^4) \leq 16(M + \mu^4)$$

and it suffices to show that $(Y_1 + \cdots + Y_n)/n \to 0$ a.s.. So we are reduced to the case where $\mu = 0$.

Note that X_n, X_n^2, X_n^3 are all integrable since X_n^4 is. Since $\mu = 0$, by independence,

$$\mathbb{E}(X_i X_j^3) = \mathbb{E}(X_i X_j X_k^2) = \mathbb{E}(X_i X_j X_k X_l) = 0$$

for distinct indices i, j, k, l. Hence

$$\mathbb{E}(S_n^4) = \mathbb{E}\Big(\sum_{1 \leq i \leq n} X_k^4 + 6 \sum_{1 \leq i < j \leq n} X_i^2 X_j^2\Big).$$

Now for $i < j$, by independence and the Cauchy–Schwarz inequality

$$\mathbb{E}(X_i^2 X_j^2) = \mathbb{E}(X_i^2)\mathbb{E}(X_j^2) \leq \mathbb{E}(X_i^4)^{1/2}\mathbb{E}(X_j^4)^{1/2} \leq M.$$

So we get the bound

$$\mathbb{E}(S_n^4) \leq nM + 3n(n-1)M \leq 3n^2 M.$$

Thus

$$\mathbb{E}\sum_n (S_n/n)^4 \leq 3M \sum_n 1/n^2 < \infty$$

which implies

$$\sum_n (S_n/n)^4 < \infty \quad \text{a.s.}$$

and hence $S_n/n \to \mu$ a.s.. □

10.2. **Strong law of large numbers.**

Theorem 10.2.1. *Let m be a probability measure on $\mathbb{R}$, with*

$$\int_{\mathbb{R}} |x| m(dx) < \infty, \qquad \int_{\mathbb{R}} x m(dx) = \nu.$$

Let $(E, \mathcal{E}, \mu)$ be the canonical model for a sequence of independent random variables with law m. Then

$$\mu(\{x : (x_1 + \cdots + x_n)/n \to \nu \text{ as } n \to \infty\}) = 1.$$

Proof. By Theorem 9.2.1, the shift map θ on E is measure-preserving and ergodic. The coordinate function $f = X_1$ is integrable and $S_n(f) = f + f \circ \theta + \cdots + f \circ \theta^{n-1} = X_1 + \cdots + X_n$. So $(X_1 + \cdots + X_n)/n \to \bar{f}$ a.e. and in L^1, for some invariant function $\bar{f}$, by Birkhoff's theorem. Since θ is ergodic, $\bar{f} = c$ a.e., for some constant c and then $c = \mu(\bar{f}) = \lim_n \mu(S_n/n) = \nu$. □

Theorem 10.2.2 (Strong law of large numbers)**.** *Let $(Y_n : n \in \mathbb{N})$ be a sequence of independent, identically distributed, integrable random variables with mean ν. Set $S_n = Y_1 + \cdots + Y_n$. Then*

$$S_n/n \to \nu \quad \textit{a.s., as } n \to \infty.$$

Proof. In the notation of Theorem 10.2.1, take m to be the law of the random variables Y_n. Then $\mu = \mathbb{P} \circ Y^{-1}$, where $Y : \Omega \to E$ is given by $Y(\omega) = (Y_n(\omega) : n \in \mathbb{N})$. Hence

$$\mathbb{P}(S_n/n \to \nu \text{ as } n \to \infty) = \mu(\{x : (x_1 + \cdots + x_n)/n \to \nu \text{ as } n \to \infty\}) = 1.$$

□

10.3. **Central limit theorem.** We will use the following special case of Lévy's continuity theorem for characteristic functions. *Let $X_n, n \in \mathbb{N}$, and X be random variables. Suppose that $\phi_{X_n}(u) \to \phi_X(u)$ for all u and that F_X is continuous. Then $F_{X_n}(x) \to F_X(x)$ for all $x \in \mathbb{R}$.*

Theorem 10.3.1 (Central limit theorem). *Let $(X_n : n \in \mathbb{N})$ be a sequence of independent, identically distributed, random variables with mean 0 and variance 1. Set $S_n = X_1 + \dots + X_n$. Then, for all $a < b$, as $n \to \infty$,*

$$\mathbb{P}(S_n/\sqrt{n} \in [a, b]) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

Proof. Set $\phi(u) = \mathbb{E}(e^{iuX_1})$. Since $\mathbb{E}(X_1^2) < \infty$, we can differentiate $\mathbb{E}(e^{iuX_1})$ twice under the expectation, to show that

$$\phi(0) = 1, \quad \phi'(0) = 0, \quad \phi''(0) = -1.$$

Hence, by Taylor's theorem, as $u \to 0$,

$$\phi(u) = 1 - u^2/2 + o(u^2).$$

So, for the characteristic function ϕ_n of $S_n/\sqrt{n}$,

$$\phi_n(u) = \mathbb{E}(e^{iu(X_1+\dots+X_n)/\sqrt{n}}) = \{\mathbb{E}(e^{i(u/\sqrt{n})X_1})\}^n = (1 - u^2/2n + o(u^2/n))^n.$$

The complex logarithm satisfies, as $z \to 0$,

$$\log(1 + z) = z + o(|z|)$$

so, for each $u \in \mathbb{R}$, as $n \to \infty$,

$$\log \phi_n(u) = n \log(1 - u^2/2n + o(u^2/n)) = -u^2/2 + o(1).$$

Hence $\phi_n(u) \to e^{-u^2/2}$ for all u. But $e^{-u^2/2}$ is the characteristic function of the $N(0, 1)$ distribution, so Lévy's continuity theorem now completes the proof. □

Here is an alternative argument, which does not rely on Lévy's continuity theorem. Take a random variable $Y \sim N(0, 1)$, independent of the sequence $(X_n : n \in \mathbb{N})$. Fix $a < b$ and $\delta > 0$ and consider the function f which interpolates linearly the points $(-\infty, 0), (a - \delta, 0), (a, 1), (b, 1), (b + \delta, 0), (\infty, 0)$. Note that $|f(x + y) - f(x)| \leq |y|/\delta$ for all x, y. So, given $\varepsilon > 0$, for $t = (\pi/2)(\varepsilon\delta/3)^2$ and any random variable X,

$$|\mathbb{E}(f(X + \sqrt{t}Y)) - \mathbb{E}(f(X))| \leq \mathbb{E}(\sqrt{t}|Y|)/\delta = \varepsilon/3.$$

In particular, this inequality holds for X an independent copy of Y, when we have $X + \sqrt{t}Y \sim \sqrt{1 + t}Y$, so

$$|\mathbb{E}(f(\sqrt{1 + t}Y)) - \mathbb{E}(f(Y))| \leq \varepsilon/3.$$

Recall from the proof of the Fourier inversion formula that

$$\mathbb{E}(f((S_n/\sqrt{n}) + \sqrt{t}Y)) = \int_{\mathbb{R}} \left(\frac{1}{2\pi} \int_{\mathbb{R}} \phi_n(u) e^{-u^2t/2} e^{-iuy} du \right) f(y) dy.$$

Now $e^{-u^2t/2}f(y) \in L^1(du \otimes dy)$ and ϕ_n is bounded, with $\phi_n(u) \to e^{-u^2/2}$ for all u as $n \to \infty$, so, by dominated convergence, for n sufficiently large,

$$|\mathbb{E}(f((S_n/\sqrt{n}) + \sqrt{t}Y)) - \mathbb{E}(f(\sqrt{1+t}Y))| \leq \varepsilon/3$$

and then

$$|\mathbb{E}(f(S_n/\sqrt{n})) - \mathbb{E}(f(Y))| \leq \varepsilon.$$

Hence $\mathbb{E}(f(S_n/\sqrt{n})) \to \mathbb{E}(f(Y))$ as $n \to \infty$. The same argument applies to the function g, defined like f, but with a, b replaced by $a + \delta, b - \delta$ respectively. Now $g \leq 1_{[a,b]} \leq f$, so

$$\mathbb{E}(g(S_n/\sqrt{n})) \leq \mathbb{P}(S_n/\sqrt{n} \in [a,b]) \leq \mathbb{E}(f(S_n/\sqrt{n})).$$

On the other hand, as $\delta \downarrow 0$,

$$\mathbb{E}(g(Y)) \uparrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \quad \mathbb{E}(f(Y)) \downarrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

so we must have, as $n \to \infty$,

$$\mathbb{P}(S_n/\sqrt{n} \in [a,b]) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

Exercises

Students should attempt Exercises 1.1–2.7 for their first supervision, then 3.1–3.13, 4.1–7.6 and 8.1–10.2 for later supervisions.

1.1 Show that a π-system which is also a d-system is a σ-algebra.

1.2 Show that the following sets of subsets of $\mathbb{R}$ all generate the same σ-algebra:
(a) $\{(a,b) : a < b\}$, (b) $\{(a,b] : a < b\}$, (c) $\{(-\infty, b] : b \in \mathbb{R}\}$.

1.3 Show that a countably additive set function on a ring is both increasing and countably subadditive.

1.4 Let μ be a finite-valued additive set function on a ring $\mathcal{A}$. Show that μ is countably additive if and only if

$$A_n \supseteq A_{n+1} \in \mathcal{A}, n \in \mathbb{N}, \quad \bigcap_n A_n = \emptyset \quad \Rightarrow \mu(A_n) \to 0.$$

1.5 Let $(E, \mathcal{E}, \mu)$ be a measure space. Show that, for any sequence of sets $(A_n : n \in \mathbb{N})$ in $\mathcal{E}$,

$$\mu(\liminf A_n) \leq \liminf \mu(A_n).$$

Show that, if μ is finite, then also

$$\mu(\limsup A_n) \geq \limsup \mu(A_n)$$

and give an example to show this inequality may fail if μ is not finite.

1.6 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A_n, n \in \mathbb{N}$, a sequence of events. Show that $A_n, n \in \mathbb{N}$, are independent if and only if the σ-algebras they generate

$$\sigma(A_n) = \{\emptyset, A_n, A_n^c, \Omega\}$$

are independent.

1.7 Show that, for every Borel set $B \subseteq \mathbb{R}$ of finite Lebesgue measure and every $\varepsilon > 0$, there exists a finite union of disjoint intervals $A = (a_1, b_1] \cup \cdots \cup (a_n, b_n]$ such that the Lebesgue measure of $A \triangle B$ $(= (A^c \cap B) \cup (A \cap B^c))$ is less than ε.

1.8 Let $(E, \mathcal{E}, \mu)$ be a measure space. Call a subset $N \subseteq E$ *null* if

$$N \subseteq B \text{ for some } B \in \mathcal{E} \text{ with } \mu(B) = 0.$$

Prove that the set of subsets

$$\mathcal{E}^\mu = \{A \cup N : A \in \mathcal{E}, N \text{ null}\}$$

is a σ-algebra and show that μ has a well-defined and countably additive extension to $\mathcal{E}^\mu$ given by

$$\mu(A \cup N) = \mu(A).$$

We call $\mathcal{E}^\mu$ the *completion of* $\mathcal{E}$ *with respect to* μ.

2.1 Prove Proposition 2.1.1 and deduce that, for any sequence $(f_n : n \in \mathbb{N})$ of measurable functions on $(E, \mathcal{E})$,

$$\{x \in E : f_n(x) \text{ converges as } n \to \infty\} \in \mathcal{E}.$$

2.2 Let X and Y be two random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and suppose that for all $x, y \in \mathbb{R}$

$$\mathbb{P}(X \le x, Y \le y) = \mathbb{P}(X \le x)\mathbb{P}(Y \le y).$$

Show that X and Y are independent.

2.3 Let $X_1, X_2, \ldots$ be random variables with

$$X_n = \begin{cases} n^2 - 1 & \text{with probability } 1/n^2 \\ -1 & \text{with probability } 1 - 1/n^2. \end{cases}$$

Show that

$$\mathbb{E}\left(\frac{X_1 + \cdots + X_n}{n}\right) = 0$$

but with probability one, as $n \to \infty$,

$$\frac{X_1 + \cdots + X_n}{n} \longrightarrow -1.$$

2.4 For $s > 1$ define the zeta function by
$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}.$$
Let X and Y be independent random variables with
$$\mathbb{P}(X = n) = \mathbb{P}(Y = n) = n^{-s}/\zeta(s).$$
Show that the events
$$\{p \text{ divides } X\},\ p \text{ prime}$$
are independent and deduce Euler's formula
$$\frac{1}{\zeta(s)} = \prod_p \left(1 - \frac{1}{p^s}\right).$$
Prove also that
$$\mathbb{P}(X \text{ is square-free}) = 1/\zeta(2s)$$
and
$$\mathbb{P}\big(\text{h.c.f.}(X, Y) = n\big) = n^{-2s}/\zeta(2s).$$

2.5 Let $X_1, X_2, \dots$ be independent random variables with distribution uniform on $[0, 1]$. Let A_n be the event that a record occurs at time n, that is,
$$X_n > X_m \quad \text{for all } m < n.$$
Find the probability of A_n and show that $A_1, A_2, \dots$ are independent. Deduce that, with probability one, infinitely many records occur.

2.6 Let $X_1, X_2, \dots$ be independent $N(0, 1)$ random variables. Prove that
$$\limsup_n \big(X_n/\sqrt{2\log n}\big) = 1 \qquad \text{a.s.}$$

2.7 Let C_n denote the nth approximation to the Cantor set C: thus $C_0 = [0, 1]$, $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$, $C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$, etc. and $C_n \downarrow C$ as $n \to \infty$. Denote by F_n the distribution function of a random variable uniformly distributed on C_n. Show

(i) $F(x) = \lim_{n\to\infty} F_n(x)$ exists for all $x \in [0, 1]$,
(ii) F is continuous, $F(0) = 0$ and $F(1) = 1$,
(iii) F is differentiable a.e. with $F' = 0$.

3.1 A simple function f has two representations:
$$f = \sum_{k=1}^{m} a_k 1_{A_k} = \sum_{j=1}^{n} b_k 1_{B_k}.$$

For $\varepsilon \in \{0,1\}^m$ define $A_\varepsilon = A_1^{\varepsilon_1} \cap \cdots \cap A_m^{\varepsilon_m}$ where $A_k^0 = A_k^c$, $A_k^1 = A_k$. For $\delta \in \{0,1\}^n$ define B_δ similarly. Then set

$$f_{\varepsilon,\delta} = \begin{cases} \sum_{k=1}^{m} \varepsilon_k a_k & \text{if } A_\varepsilon \cap B_\delta \neq \emptyset \\ \infty & \text{otherwise.} \end{cases}$$

Show that for any measure μ

$$\sum_{k=1}^{m} a_k \mu(A_k) = \sum_{\varepsilon,\delta} f_{\varepsilon,\delta} \mu(A_\varepsilon \cap B_\delta)$$

and deduce that

$$\sum_{k=1}^{m} a_k \mu(A_k) = \sum_{j=1}^{n} b_j \mu(B_j).$$

3.2 Show that any continuous function $f : \mathbb{R} \to \mathbb{R}$ is Lebesgue integrable over any finite interval.

3.3 Prove Propositions 3.1.4, 3.1.5 and 3.1.6.

3.4 Let X be a non-negative integer-valued random variable. Show that

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n).$$

Deduce that, if $\mathbb{E}(X) = \infty$ and $X_1, X_2, \ldots$ is a sequence of independent random variables with the same distribution as X, then

$$\limsup (X_n/n) \geq 1 \qquad \text{a.s.}$$

and indeed

$$\limsup (X_n/n) = \infty \qquad \text{a.s.}$$

Now suppose that $Y_1, Y_2, \ldots$ is any sequence of independent identically distributed random variables with $\mathbb{E}|Y_1| = \infty$. Show that

$$\limsup (|Y_n|/n) = \infty \qquad \text{a.s.}$$

and indeed

$$\limsup (|Y_1 + \cdots + Y_n|/n) = \infty \qquad \text{a.s.}$$

3.5 For $\alpha \in (0, \infty)$ and $p \in [1, \infty)$ and for

$$f_\alpha(x) = 1/x^\alpha, \quad x > 0,$$

show carefully that

$$f_\alpha \in L^p((0,1],dx) \;\Leftrightarrow\; \alpha p < 1,$$
$$f_\alpha \in L^p([1,\infty),dx) \;\Leftrightarrow\; \alpha p > 1.$$

3.6 Show that the function

$$f(x) = \frac{\sin x}{x}$$

is not Lebesgue integrable over $[1,\infty)$ but that the following limit does exist:

$$\lim_{N\to\infty} \int_1^N \frac{\sin x}{x}\,dx.$$

3.7 Show

(i): $\displaystyle\int_0^\infty \sin(e^x)/(1+nx^2)dx \to 0 \quad \text{as} \quad n\to\infty,$

(ii): $\displaystyle\int_0^1 (n\cos x)/(1+n^2x^{\frac{3}{2}})dx \to 0 \quad \text{as} \quad n\to\infty.$

3.8 Let u and v be differentiable functions on $[a,b]$ with continuous derivatives u' and v'. Show that for $a<b$

$$\int_a^b u(x)v'(x)dx = \{u(b)v(b) - u(a)v(a)\} - \int_a^b u'(x)v(x)dx.$$

3.9 Prove Propositions 3.4.4, 3.4.6, 3.5.1, 3.5.2 and 3.5.3.

3.10 The moment generating function ϕ of a real-valued random variable X is defined by

$$\phi(\tau) = \mathbb{E}(e^{\tau X}), \quad \tau \in \mathbb{R}.$$

Show that the set $I = \{\tau : \phi(\tau) < \infty\}$ is an interval and find examples where I is $\mathbb{R}$, $\{0\}$ and $(-\infty,1)$. Assume for simplicity that $X \geq 0$. Show that if I contains a neighbourhood of 0 then X has finite moments of all orders given by

$$\mathbb{E}(X^n) = \left(\frac{d}{d\tau}\right)^n\Big|_{\tau=0}\phi(\tau).$$

Find a necessary and sufficient condition on the sequence of moments $m_n = \mathbb{E}(X^n)$ for I to contain a neighbourhood of 0.

3.11 Let $X_1,\dots,X_n$ be random variables with density functions $f_1,\dots,f_n$ respectively. Suppose that the $\mathbb{R}^n$-valued random variable $X = (X_1,\dots,X_n)$ also has a density function f. Show that $X_1,\dots,X_n$ are independent if and only if

$$f(x_1,\dots,x_n) = f_1(x_1)\dots f_n(x_n) \qquad \text{a.e.}$$

3.12 Let $(f_n : n \in \mathbb{N})$ be a sequence of integrable functions and suppose that $f_n \to f$ a.e. for some integrable function f. Show that, if $\|f_n\|_1 \to \|f\|_1$, then $\|f_n - f\|_1 \to 0$.

3.13 Let μ and ν be probability measures on $(E, \mathcal{E})$ and suppose that, for some measurable function $f : E \to [0, R]$,

$$\nu(A) = \int_A f \, d\mu, \qquad A \in \mathcal{E}.$$

Let $(X_n : n \in \mathbb{N})$ be a sequence of independent random variables in E with law μ and let $(U_n : n \in \mathbb{N})$ be a sequence of independent $U[0,1]$ random variables. Set

$$T = \min\{n \in \mathbb{N} : RU_n \leq f(X_n)\}, \qquad Y = X_T.$$

Show that Y has law ν.

4.1 Let X be a random variable and let $1 \leq p < q < \infty$. Show that

$$\mathbb{E}(|X|^p) = \int_0^\infty p\lambda^{p-1}\mathbb{P}(|X| \geq \lambda) d\lambda$$

and deduce

$$X \in L^q(\mathbb{P}) \Rightarrow \mathbb{P}(|X| \geq \lambda) = O(\lambda^{-q}) \Rightarrow X \in L^p(\mathbb{P}).$$

4.2 Give a simple proof of Schwarz' inequality for measurable functions f and g:

$$\|fg\|_1 \leq \|f\|_2 \|g\|_2.$$

4.3 Show that for independent random variables X and Y

$$\|XY\|_1 = \|X\|_1 \|Y\|_1$$

and that if both X and Y are integrable then

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

4.4 A *stepfunction* $f : \mathbb{R} \to \mathbb{R}$ is any finite linear combination of indicator functions of finite intervals. Show that the set of stepfunctions $\mathcal{I}$ is dense in $L^p(\mathbb{R})$ for all $p \in [1, \infty)$: that is, for all $f \in L^p(\mathbb{R})$ and all $\varepsilon > 0$ there exists $g \in \mathcal{I}$ such that $\|f - g\|_p < \varepsilon$.

4.5 Let $(X_n : n \in \mathbb{N})$ be an identically distributed sequence in $L^2(\mathbb{P})$. Show that, for $\varepsilon > 0$,

(i): $n\mathbb{P}(|X_1| > \varepsilon\sqrt{n}) \to 0$ as $n \to \infty$,

(ii): $n^{-\frac{1}{2}} \max_{k \leq n} |X_k| \to 0$ in probability.

5.1 Let $(E, \mathcal{E}, \mu)$ be a measure space and let $V_1 \leq V_2 \leq \dots$ be an increasing sequence of closed subspaces of $L^2 = L^2(E, \mathcal{E}, \mu)$ for $f \in L^2$, denote by f_n the orthogonal projection of f on V_n. Show that f_n converges in L^2.

5.2 Prove Propositions 5.3.1 and 5.3.2.

6.1 Prove Proposition 6.2.2.

6.2 Find a uniformly integrable sequence of random variables $(X_n : n \in \mathbb{N})$ such that

$$X_n \to 0 \text{ a.s.} \quad \text{and} \quad \mathbb{E}\big(\sup_n |X_n|\big) = \infty.$$

6.3 Let $(X_n : n \in \mathbb{N})$ be an identically distributed sequence in $L^2(\mathbb{P})$. Show that

$$\mathbb{E}\big(\max_{k \leq n} |X_k|\big)/\sqrt{n} \to 0 \qquad \text{as} \quad n \to \infty.$$

7.1 Show that the Fourier transform of a finite Borel measure is a bounded continuous function.

7.2 Let μ be a Borel measure on $\mathbb{R}$ of finite total mass. Suppose the Fourier transform $\hat{\mu}$ is Lebesgue integrable. Show that μ has a continuous density function f with respect to Lebesgue measure:

$$\mu(A) = \int_A f(x)dx.$$

7.3 Show that there do not exist independent identically distributed random variables X, Y such that

$$X - Y \sim U[-1, 1].$$

7.4 The Cauchy distribution has density function

$$f(x) = \frac{1}{\pi(1 + x^2)}, \qquad x \in \mathbb{R}.$$

Show that the corresponding characteristic function is given by

$$\varphi(u) = e^{-|u|}.$$

Show also that, if $X_1, \dots, X_n$ are independent Cauchy random variables, then $(X_1 + \dots + X_n)/n$ is also Cauchy. Comment on this in the light of the strong law of large numbers and central limit theorem.

7.5 For a finite Borel measure μ on the line show that, if $\int |x|^k d\mu(x) < \infty$, then the Fourier transform $\hat{\mu}$ of μ has a kth continuous derivative, which at 0 is given by

$$\hat{\mu}^{(k)}(0) = i^k \int x^k d\mu(x).$$

7.6 (i) Show that for any real numbers a, b one has $\int_a^b e^{itx}dx \to 0$ as $|t| \to \infty$.

(ii) Suppose that μ is a finite Borel measure on $\mathbb{R}$ which has a density f with respect to Lebesgue measure. Show that its Fourier transform

$$\hat{\mu}(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx$$

tends to 0 as $|t| \to \infty$. This is the *Riemann–Lebesgue Lemma.*

(iii) Suppose that the density f of μ has an integrable and continuous derivative f'. Show that

$$\hat{\mu}(t) = o(t^{-1}), \qquad \text{i.e.,} \quad t\hat{\mu}(t) \to 0 \text{ as } |t| \to \infty.$$

Extend to higher derivatives.

8.1 Prove Proposition 8.1.1.

8.2 Suppose that $X_1, \ldots, X_n$ are jointly Gaussian random variables with

$$\mathbb{E}(X_i) = \mu_i, \quad \operatorname{cov}(X_i, X_j) = \Sigma_{ij}$$

and that the matrix $\Sigma = (\Sigma_{ij})$ is invertible. Set $Y = \Sigma^{-\frac{1}{2}}(X - \mu)$. Show that $Y_1, \ldots, Y_n$ are independent $N(0,1)$ random variables.

Show that we can write X_2 in the form $X_2 = aX_1 + Z$ where Z is independent of X_1 and determine the distribution of Z.

8.3 Let $X_1, \ldots, X_n$ be independent $N(0,1)$ random variables. Show that

$$\left(\overline{X}, \sum_{m=1}^{n} (X_m - \overline{X})^2\right) \quad \text{and} \quad \left(X_n/\sqrt{n}, \sum_{m=1}^{n-1} X_m^2\right)$$

have the same distribution, where $\overline{X} = (X_1 + \cdots + X_n)/n$.

9.1 Let $(E, \mathcal{E}, \mu)$ be a measure space and $\tau : E \to E$ a measure-preserving transformation. Show that

$$\mathcal{E}_\tau := \{A \in \mathcal{E} : \tau^{-1}(A) = A\}$$

is a σ-algebra, and that a measurable function f is $\mathcal{E}_\tau$-measurable if and only if it is *invariant*, that is $f \circ \tau = f$.

9.2 Prove Propositions 9.1.1 and 9.1.2.

9.3 For $E = [0,1), a \in E$ and $\mu(dx) = dx$, show that

$$\tau(x) = x + a \pmod 1$$

is measure-preserving. Determine for which values of a the transformation τ is ergodic.

Let f be an integrable function on $[0,1)$. Determine for each value of a the limit

$$\overline{f} = \lim_{n \to \infty} \frac{1}{n}(f + f \circ \tau + \cdots + f \circ \tau^{n-1}).$$

9.4 Show that

$$\tau(x) = 2x \pmod 1$$

is another measure-preserving transformation of Lebesgue measure on $[0, 1)$, and that τ is ergodic. Find $\overline{f}$ for each integrable function f.

9.5 Call a sequence of random variables $(X_n : n \in \mathbb{N})$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *stationary* if for each $n, k \in \mathbb{N}$ the random vectors $(X_1, \dots, X_n)$ and $(X_{k+1}, \dots, X_{k+n})$ have the same distribution: for $A_1, \dots, A_n \in \mathcal{B}$,

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_{k+1} \in A_1, \dots, X_{k+n} \in A_n).$$

Show that, if $(X_n : n \in \mathbb{N})$ is a stationary sequence and $X_1 \in L^p$, for some $p \in [1, \infty)$, then

$$\frac{1}{n} \sum_{i=1}^{n} X_i \to X \qquad \text{a.s. and in } L^p,$$

for some random variable $X \in L^p$ and find $\mathbb{E}[X]$.

10.1 Let f be a bounded continuous function on $(0, \infty)$, having Laplace transform

$$\hat{f}(\lambda) = \int_0^\infty e^{-\lambda x} f(x) dx, \quad \lambda \in (0, \infty).$$

Let $(X_n : n \in \mathbb{N})$ be a sequence of independent exponential random variables, of parameter λ. Show that $\hat{f}$ has derivatives of all orders on $(0, \infty)$ and that, for all $n \in \mathbb{N}$, for some $C(\lambda, n) \neq 0$ independent of f, we have

$$(d/d\lambda)^{n-1} \hat{f}(\lambda) = C(\lambda, n) \mathbb{E}\big(f(S_n)\big)$$

where $S_n = X_1 + \dots + X_n$. Deduce that if $\hat{f} \equiv 0$ then also $f \equiv 0$.

10.2 For each $n \in \mathbb{N}$, there is a unique probability measure μ_n on the unit sphere $S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$ such that $\mu_n(A) = \mu_n(UA)$ for all Borel sets A and all orthogonal $n \times n$ matrices U. Fix $k \in \mathbb{N}$ and, for $n \geq k$, let γ_n denote the probability measure on $\mathbb{R}^k$ which is the law of $\sqrt{n}(x^1, \dots, x^k)$ under μ_n. Show

(i) if $X \sim N(0, I_n)$ then $X/|X| \sim \mu_n$,

(ii) if $(X_n : n \in \mathbb{N})$ is a sequence of independent $N(0, 1)$ random variables and if $R_n = (X_1^2 + \dots + X_n^2)^{\frac{1}{2}}$ then $R_n/\sqrt{n} \to 1$ a.s.,

(iii) for all bounded continuous functions f on $\mathbb{R}^k$, $\gamma_n(f) \to \gamma(f)$, where γ is the standard Gaussian distribution on $\mathbb{R}^k$.

Index