7 Approximation

Measurable functions are approximately nice, even if they look terrible.

7a A terrible integrable function

The indicator function of the set of all rational numbers is terrible for Riemann integration, but trivial for Lebesgue integration: it is equivalent to zero. An equivalence class of terrible integrable functions is shown below.

Recall the binary digits $\beta_1, \beta_2, \dots : [0,1) \to \{0,1\}$ (treated in Sect. 3d4), and consider the series

$$f = \sum_{k=1}^{\infty} \frac{1}{k}(2\beta_k - 1)\,.$$

Probabilistically, this is the random series $\sum \pm \frac{1}{k}$ with independent fair random signs.

In the Hilbert space $L_2[0,1]$ (treated in Sect. 5f) vectors $2\beta_k - 1$ are orthonormal (since the joint distribution of $2\beta_k - 1$ and $2\beta_l - 1$ is uniform on $\{-1,+1\} \times \{-1,+1\}$). Thus, partial sums $S_n = \sum_{k=1}^{n} \frac{1}{k}(2\beta_k - 1)$ satisfy $\|S_n\|_2^2 = \sum_{k=1}^{n} \frac{1}{k^2} \uparrow \frac{1}{6}\pi^2 < \infty$, and $\|S_{n+k} - S_n\|_2^2 \le \sum_{k=n+1}^{\infty} \frac{1}{k^2} \downarrow 0$, that is, $(S_n)_n$ is a Cauchy sequence; by 5f3, it converges to some f in $L_2[0,1]$ (in fact, it converges almost everywhere; never mind).

7a1 Proposition. $m\{x \in (a,b) : f(x) \in (c,d)\} > 0$ whenever $0 < a < b < 1$ and $-\infty < c < d < \infty$.

We consider binary intervals (of rank n) $I = \big[2^{-n}k, 2^{-n}(k+1)\big) \subset \big[0,1\big)$, note that $\beta_1, \dots, \beta_n$ are constant on such I, and denote by m_I the uniform distribution on I; that is,

$$m_I(A) = \frac{m(A \cap I)}{m(I)} = 2^n m(A \cap I)\,;$$

probabilistically, $f_* m_I$ is the conditional distribution of f given $\beta_1, \dots, \beta_n$. Here are the conditional expectation and conditional variance.

7a2 Exercise. For every binary interval I of rank n,

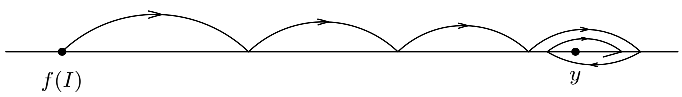
$$\int f \, \mathrm{d}m_I = \sum_{k=1}^n \frac{1}{k}(2\beta_k(I) - 1) \, ; \qquad \int \Big(f - \int f \, \mathrm{d}m_I \Big)^2 \mathrm{d}m_I = \sum_{k=n+1}^\infty \frac{1}{k^2} \, .$$

Prove it.

7a3 Exercise. For arbitrary binary interval I and arbitrary $y \in \mathbb{R}$, $\varepsilon > 0$ there exists a binary interval $J \subset I$ such that

$$\Big| \int f \, \mathrm{d}m_J - y \Big| \le \varepsilon \, , \quad \int \Big(f - \int f \, \mathrm{d}m_J \Big)^2 \mathrm{d}m_J \le \varepsilon^2 \, .$$

Prove it. Hint:



***Proof** of Prop. 7a1.* We take a binary interval $I \subset (a,b)$, $y = (c+d)/2$ and $\varepsilon = (d-c)/6$. By 7a3 there exists a binary interval $J \subset I$ such that $\big| \int f \, \mathrm{d}m_J - y \big| \le \varepsilon$ and $\int \big(f - \int f \, \mathrm{d}m_J \big)^2 \mathrm{d}m_J \le \varepsilon^2$. We have

$$\begin{aligned} m_J\{ x : f(x) \notin (c,d) \} &\le m_J \{ x : |f(x) - y| \ge 3\varepsilon \} \le \\ &\le m_J \{ x : |f(x) - \textstyle\int f \, \mathrm{d}m_J| \ge 2\varepsilon \} \le \tfrac{\varepsilon^2}{(2\varepsilon)^2} = \tfrac14 < 1 \, , \end{aligned}$$

thus,

$$\begin{aligned} m\{ x \in (a,b) : f(x) \in (c,d) \} &\ge m \{ x \in J : f(x) \in (c,d) \} = \\ &= 2^{-n} m_J \{ x : f(x) \in (c,d) \} > 0 \, , \end{aligned}$$

where n is the rank of J. □

We see what can happen to an integrable function. But can it, somehow, behave like β_∞? We have $\int_I \beta_k \, \mathrm{d}m = \frac12 m(I)$ whenever I is a binary interval of rank $n < k$. It should be $\int_I \beta_\infty \, \mathrm{d}m = \frac12 m(I)$ for all binary intervals I. But then $\beta_\infty \cdot m = \frac12 m$ by 6b12, which implies $\beta_\infty = \frac12$ a.e. by Lemma 7a4 below. You see, infinite frequency (like $t \mapsto \sin \infty t$) is impossible.[1]

7a4 Lemma. Let (X, S, μ) be a σ-finite measure space and $f, g : X \to [0, \infty]$ measurable functions. If $f \cdot \mu = g \cdot \mu$, then $f = g$ a.e.

[1] Probably one can invent a new generalization of the notion "function" that makes it possible, but surely outside Lebesgue's theory.

Proof. It is sufficient to prove that $f \le g$ a.e. Assume the contrary: $\mu\{x : g(x) < f(x)\} > 0$. By σ-finiteness there exists $A \in S$ such that $0 < \mu(A) < \infty$ and $g < f$ on A. WLOG, g is bounded on A and $f-g \ge \varepsilon > 0$ on A (think, why). Thus, $(f \cdot \mu)(A) = \int_A f \,\mathrm{d}\mu \ge \int_A g \,\mathrm{d}\mu + \varepsilon\mu(A) > \int_A g \,\mathrm{d}\mu = (g \cdot \mu)(A)$; a contradiction. $\square$

7b Approximation of sets

Convergence of sets may be treated as convergence of their indicator functions:

$$A_n \to A \quad \Longleftrightarrow \quad 1\!\!1_{A_n} \to 1\!\!1_A \,.$$

Thus, each mode of convergence for functions leads to the corresponding mode of convergence for sets. And quite often, different modes for functions lead to the same mode for sets.

Note that $|1\!\!1_A - 1\!\!1_B| = 1\!\!1_{A\triangle B}$, where $A\triangle B = (A \setminus B) \uplus (B \setminus A)$ is the symmetric difference. Each convergence mode for functions satisfies $f_n \to f \iff f_n - f \to 0 \iff |f_n - f| \to 0$; thus, each convergence mode for sets satisfies

$$A_n \to A \quad \Longleftrightarrow \quad A_n \triangle A \to \emptyset \,.$$

Clearly, $\|1\!\!1_A\|_p = \big(\mu(A)\big)^{1/p}$ for $1 \le p < \infty$; thus, for sets the relation "$A_n \to \emptyset$ in L_p" does not depend on p. Also, $\mu\{x : 1\!\!1_A(x) \ge \varepsilon\} = \mu(A)$ for all $\varepsilon \in (0,1)$; thus, for sets, $A_n \to \emptyset$ globally in measure if and only if $A_n \to \emptyset$ in L_1. In contrast, convergence a.e. is different (recall 5c3), and the local convergence in measure (defined in σ-finite spaces) is different (recall $1\!\!1_{[n,\infty)}$ before 5b3).

So, for sets we have two modes of convergence in general, and three modes in σ-finite spaces.

7b1 Exercise. (a) If $\sum_n \mu(A_n) < \infty$, then $A_n \to \emptyset$ a.e.;

(b) if $\sum_n \mu(A_n \triangle A_{n+1}) < \infty$, then $\exists A \;\; A_n \to A$ a.e.

Prove it.[1]

In σ-finite spaces convergence a.e. implies local convergence in measure by 5c1. And if μ is finite, then convergence a.e. implies L_1-convergence (for sets, of course).

7b2 Proposition. Let (X, S, μ) be a σ-finite measure space, and $\mathcal{E} \subset S$ a generating algebra of sets.[2] Then for every $A \in S$ there exist $E_1, E_2, \dots \in \mathcal{E}$ such that $E_n \to A$ a.e.

---

[1] Hint: simpler than the proof of 5c5.

[2] That is, the σ-algebra generated by $\mathcal{E}$ contains each set of S up to a null set.

Proof. By 5b8, WLOG, $\mu(X) < \infty$. We endow S with the metric $\rho(A, B) = \mu(A \triangle B)$ (on equivalence classes of sets, of course), and consider the closure $\overline{\mathcal{E}}$ of $\mathcal{E}$ in S; that is, for $A \in S$

$$A \in \overline{\mathcal{E}} \iff \forall \varepsilon > 0 \, \exists E \in \mathcal{E} \;\; \rho(A, E) < \varepsilon \iff \\ \iff \exists E_1, E_2, \dots \in \mathcal{E} \;\; \rho(A, E_n) \to 0 \, ;$$

clearly, $\overline{\mathcal{E}}$ is closed (that is, equal to its closure). Therefore $\overline{\mathcal{E}}$ is a monotone class; by the Monotone class theorem 6b11, $\overline{\mathcal{E}} = S$.

Given $A \in S$, we take $E_n \in \mathcal{E}$ such that $\rho(A, E_n) \to 0$, choose a subsequence $(E_{n_k})_k$ such that $\sum_k \rho(A, E_{n_k}) < \infty$, and get $A \triangle E_{n_k} \to \emptyset$ a.e. by 7b1(a). $\square$

7b3 Remark. Instead of using the Monotone class theorem we may prove that $\overline{\mathcal{E}}$ is an algebra of sets (think, why) and therefore a σ-algebra by 6b10.

We say that μ is $\mathcal{E}$-σ-finite if there exist $E_1, E_2, \dots \in \mathcal{E}$ such that $\mu(E_k) < \infty$ and $\cup_k E_k = X$.

7b4 Exercise. Let $\mathcal{E} \subset S$ be a generating algebra of sets, and μ be $\mathcal{E}$-σ-finite. Then for every $A \in S$ such that $\mu(A) < \infty$ and every $\varepsilon > 0$ there exists $E \in \mathcal{E}$ such that $\mu(A \triangle E) < \varepsilon$.

Prove it.[1]

A useful algebra $\mathcal{E}$ on $\mathbb{R}$ is generated by intervals. Dealing with Lebesgue measure (or another nonatomic measure) we do not need to bother whether intervals are open, closed or neither; but unbounded intervals are allowed. Dealing with an arbitrary measure (possibly, with atoms) we define intervals as connected subsets of $\mathbb{R}$ (including, among others, $[a, b)$, $[a, a] = \{a\}$, $\emptyset$, $\mathbb{R}$). Taking into account that intersection of two intervals is an interval, and the complement of an interval is the union of (at most) two intervals, we conclude that $\mathcal{E}$ consists of all unions of finitely many (disjoint) intervals.

7b5 Corollary (of 7b4)**.** (a) For every Lebesgue measurable $A \subset \mathbb{R}$ of finite measure and every $\varepsilon > 0$ there exists $E \in \mathcal{E}$ such that $m(A \triangle E) < \varepsilon$;

(b) the same holds for arbitrary locally finite measure on $\mathbb{R}^d$.[2]

Every (measurable) set is nearly a finite sum of intervals. (Littlewood)[3]

[1] Hint: WLOG, $\mu(X) < \infty$; use 7b2.

[2] As before, by a measure on $\mathbb{R}^d$ we mean a (completed) measure on the Borel σ-algebra.

[3] See Tao, Sect. 1.3.5; Stein & Shakarchi Sect. 4.3; Wikipedia, "Littlewood's three principles of real analysis".

7b6 Example. The claim 7b5(b) can fail for a σ-finite measure. There exists a σ-finite measure μ on $[0,1]$ such that $\mu\big((a,b)\big) = \infty$ whenever $0 \le a < b \le 1$ and $\mu\big(\{a\}\big) = 0$ for all $a \in [0,1]$. We take $\mu = f^2 \cdot m$ with integrable $f : [0,1] \to [0,\infty)$ that is nowhere square integrable. For example, $f(x) = \sum_k c_k |x - x_k|^{-1/2}$ where $\sum_k |c_k| < \infty$ and $x_k \in [0,1]$ are a dense sequence.

7c Approximation of functions

Let (X, S, μ) be a σ-finite measure space, and $\mathcal{E} \subset S$ a generating algebra of sets. A function $f : X \to \mathbb{R}$ will be called $\mathcal{E}$-simple, if $f(X) \subset \mathbb{R}$ is finite, and $f^{-1}(y) \in \mathcal{E}$ for each $y \in f(X)$. Note that f is $\mathcal{E}$-simple if and only if f is a linear combination of indicators $\mathbb{1}_E$ for $E \in \mathcal{E}$.[1]

7c1 Proposition. For every $f \in L_0(\mu)$ there exist $\mathcal{E}$-simple functions $f_1, f_2, \dots$ such that $f_n \to f$ locally in measure.

Proof. The closure, in $L_0(\mu)$, of the vector space of all $\mathcal{E}$-simple functions is a closed vector subspace of $L_0(\mu)$. By 7b2, this subspace contains $\mathbb{1}_A$ for all $A \in S$. Therefore it contains all simple (not just $\mathcal{E}$-simple!) functions, and their limits; these are the whole $L_0(\mu)$ (recall the paragraph before 4c15). $\square$

7c2 Proposition. If μ is $\mathcal{E}$-σ-finite then, for every $p \in [1,\infty)$, $\mathcal{E}$-simple functions of $L_p(\mu)$ are dense in $L_p(\mu)$.

Proof. Approximating a given $f \in L_p(\mu)$ by $f\mathbb{1}_E$ where $E \in \mathcal{E}$ and $\mu(E) < \infty$, we may assume WLOG that $\mu(X) < \infty$. Similarly to the proof of 7c1, the closure in $L_p(\mu)$ of $\mathcal{E}$-simple functions contains all simple functions (of $L_p(\mu)$). Given $f \in L_p(\mu)$, we take simple $f_1, f_2, \dots \in L_p(\mu)$ such that $f_n \to f$ a.e., and $|f_n| \le |f|$ a.e. (think, how); then $f_n \to f$ in $L_p(\mu)$ by the Dominated Convergence Theorem 5d1, since $|f_n - f|^p \le (2|f|)^p$. $\square$

When $\mathcal{E}$ is the algebra (on $\mathbb{R}$) generated by intervals, $\mathcal{E}$-simple functions with bounded support are step functions.

7c3 Corollary. For every $p \in [1,\infty)$,

(a) step functions are dense in $L_p(\mathbb{R})$;

(b) the same holds for $\mathbb{R}^d$ with arbitrary locally finite measure.

Every function (of class L_p) is nearly a step function.

[1] Recall the finite subalgebras treated in the proof of 6b7.

7c4 Exercise. (a) $L_p(\mathbb{R}^d, \mu)$ is separable[1] for arbitrary $p \in [1, \infty)$ and σ-finite μ;

(b) this claim can fail if μ is not σ-finite;

(c) $L_\infty[0, 1]$ is not separable.[2]

Prove it.[3]

7c5 Corollary. The Hilbert space $L_2(\mathbb{R}^d, \mu)$ has a (finite or) countable orthonormal basis (and therefore, is either finite-dimensional or isomorphic to l_2), provided that μ is σ-finite.

7c6 Exercise. For every $p \in [1, \infty)$,

(a) compactly supported continuous functions are dense in $L_p(\mathbb{R})$;

(b) the same holds for $\mathbb{R}$ with arbitrary locally finite measure;

(c) the same holds for $\mathbb{R}^d$ with arbitrary locally finite measure;

(d) this claim can fail for a σ-finite measure.

Prove it.[4]

Thus, in 7c5 one can find a basis that consists of compactly supported continuous functions, provided that μ is locally finite.

Every function (of class L_p) is nearly continuous.	(Littlewood)

7d Introduction to convolution

7d1 Definition. The *convolution* $\mu * \nu$ of two σ-finite measures μ, ν on $\mathbb{R}^d$ is the pushforward measure of the product measure $\mu \times \nu$ under the map $\mathbb{R}^d \times \mathbb{R}^d \ni (x, y) \mapsto x + y \in \mathbb{R}^d$.

That is,

$$(\mu * \nu)(B) = (\mu \times \nu)\big(\{(x, y) : x + y \in B\}\big)$$

for all Borel sets $B \subset \mathbb{R}^d$.

Clearly, the convolution is positively bilinear: $\mu * (\nu_1 + \nu_2) = \mu * \nu_1 + \mu * \nu_2$ etc. Also, $(\mu * \nu)(\mathbb{R}^d) = \mu(\mathbb{R}^d)\nu(\mathbb{R}^d)$. The convolution of finite measures is a finite measure. But the convolution of σ-finite measures need not be σ-finite

[1]That is, contains some (finite or) countable dense subset.

[2]And what do you think about a necessary and sufficient condition for separability of $L_\infty(\mathbb{R}^d, \mu)$?

[3]Hint: (a) WLOG μ is finite; (b) try the counting measure; (c) think about $\|\mathbb{1}_{[0,a]} - \mathbb{1}_{[0,b]}\|_\infty$.

[4]Hint: (a) approximate $\mathbb{1}_{[a,b]}$; (b) and (c) mind the endpoints; (d) recall 3f2.

(and indeed, $m_d * m_d$ is not).[1] Rather, it is the sum of countably many finite measures.

Probabilistically, for probability measures on $\mathbb{R}$,

$$P_{X+Y} = P_X * P_Y \quad \text{for independent random variables } X, Y \,,$$

and the same holds for d-dimensional random vectors. In this case commutativity and associativity suggest themselves:

$$\begin{gathered} P_X * P_Y = P_{X+Y} = P_{Y+X} = P_Y * P_X \,; \\ (P_X * P_Y) * P_Z = P_{(X+Y)+Z} = P_{X+(Y+Z)} = P_X * (P_Y * P_Z) \,. \end{gathered}$$

Commutativity holds in general:

$$\begin{aligned} (\mu * \nu)(B) = (\mu \times \nu)\{(x,y) : x+y \in B\} = (\mu \times \nu)\{(y,x) : x+y \in B\} = \\ = (\nu \times \mu)\{(x,y) : x+y \in B\} = (\nu * \mu)(B) \,. \end{aligned}$$

Associativity is more problematic, since $\mu * \nu$ need not be σ-finite.

7d2 Exercise. (a) Multiplication of measure spaces is associative. That is, if (X, S_1, μ), (Y, S_2, ν), (Z, S_3, ξ) are σ-finite measure spaces, then $\big((X, S_1, \mu) \times (Y, S_2, \nu)\big) \times (Z, S_3, \xi) = (X, S_1, \mu) \times \big((Y, S_2, \nu) \times (Z, S_3, \xi)\big)$. (As usual, we treat $((x, y), z)$ and $(x, (y, z))$ as just (x, y, z).)

(b) Let (X, S_1, μ), (Y, S_2, ν), (Z, S_3, ξ) be σ-finite measure spaces, and $\varphi : X \to Y$ a measure preserving map (that is, measurable, and $\varphi_* \mu = \nu$). Then $\varphi \times \mathrm{id} : X \times Z \to Y \times Z$ is a measure preserving map from $(X, S_1, \mu) \times (Z, S_3, \xi)$ to $(Y, S_2, \nu) \times (Z, S_3, \xi)$. (Here $(\varphi \times \mathrm{id})(x, z) = (\varphi(x), z)$, of course.)

(c) If σ-finite measures μ, ν, ξ on $\mathbb{R}^d$ are such that $\mu * \nu$ and $\nu * \xi$ are σ-finite, then $(\mu * \nu) * \xi = \mu * (\nu * \xi)$.
Prove it.[2]

Each measurable function $f : \mathbb{R}^d \to [0, \infty]$ leads to a measure $f \cdot m$; this measure is σ-finite if and only if f is finite a.e. (think, why); and the map $[f] \mapsto f \cdot m$ is one-to-one by 7a4 (but not onto, of course). In this sense we may treat functions as (special) measures and write, say, "$\mu * f$" instead of "$\mu * (f \cdot m)$". If this $\mu * f$ appears to be some $g \cdot m$, then we may write "$\mu * f = g$"; such g is unique, but does it exist? Yes, always!

[1]Another example, with μ finite and ν locally finite on $\mathbb{R}$: $\mu = \big(x \mapsto \frac{1}{1+x^2}\big) \cdot m_1$ and $\nu = \big(x \mapsto x^2\big) \cdot m_1$.

[2]Hint: (a) it is sufficient to check product sets in $Y \times Z$; (c) apply (b) to $((x, y), z) \mapsto (x + y, z)$.

7d3 Proposition. For every measurable function $f : \mathbb{R}^d \to [0,\infty)$ and every σ-finite measure μ on $\mathbb{R}^d$,

$$\mu * (f \cdot m) = g \cdot m \quad \text{where } g : y \mapsto \int f(y - \cdot) \, \mathrm{d}\mu \in [0,\infty] \, .$$

Proof. We prove, for arbitrary Borel $h : \mathbb{R}^d \to \{0,1\}$, that $\int h \, \mathrm{d}\big(\mu * (f \cdot m)\big) = \int h \, \mathrm{d}(g \cdot m)$, that is,

$$\int \big((x,y) \mapsto h(x+y) \big) \, \mathrm{d}\big(\mu \times (f \cdot m)\big) = \int gh \, \mathrm{d}m \, .$$

By Tonelli theorem 6b15 the left-hand side is $\int \big(x \mapsto \int h(x + \cdot) \, \mathrm{d}(f \cdot m) \big) \, \mathrm{d}\mu$. Taking into account that $\int h(x + \cdot) \, \mathrm{d}(f \cdot m) = \int f(\cdot) h(x + \cdot) \, \mathrm{d}m = \int f(\cdot - x) h(\cdot) \, \mathrm{d}m$ (due to the shift invariance of Lebesgue measure m) we get (by Tonelli again)

$$\begin{aligned} &\int \bigg(x \mapsto \int f(\cdot - x) h(\cdot) \, \mathrm{d}m \bigg) \mathrm{d}\mu = \\ &\qquad = \int \big((x,y) \mapsto f(y-x) h(y) \big) \, \mathrm{d}(\mu \times m) = \\ &\qquad\qquad = \int \bigg(y \mapsto h(y) \int f(y - \cdot) \, \mathrm{d}\mu \bigg) \mathrm{d}m = \int gh \, \mathrm{d}m \, . \end{aligned}$$

□

We define the convolution $\mu * f$ of a σ-finite measure μ on $\mathbb{R}^d$ and a measurable function $f : \mathbb{R}^d \to [0,\infty)$ by

$$(\mu * f)(x) = \int f(x - \cdot) \, \mathrm{d}\mu \in [0,\infty] \, , \tag{7d4}$$

and get $\mu * (f \cdot m) = (\mu * f) \cdot m$. The function $\mu * f$ is finite a.e. if and only if the measure $\mu * (f \cdot m)$ is σ-finite. The equality $(\mu * \nu)(\mathbb{R}^d) = \mu(\mathbb{R}^d)\nu(\mathbb{R}^d)$ gives

$$\int \mu * f \, \mathrm{d}m = \mu(\mathbb{R}^d) \int f \, \mathrm{d}m \, .$$

If $f = 0$ a.e., then $\mu * f = 0$ a.e. (since the integral vanishes). Thus, the equivalence class of $\mu * f$ is uniquely determined by (μ and) the equivalence class of f.[1]

[1] If $f_1 = f_2$ outside a null set Z, then $f_1 + \infty \mathbb{1}_Z = f_2 + \infty \mathbb{1}_Z$.

Probabilistically, if a random variable Y independent of X has a density p_Y (that is, $P_Y = p_Y \cdot m$), then $X+Y$ has the density

$$p_{X+Y} = P_X * p_Y \,.$$

Also,

$$p_{X+Y}(z) = \mathbb{E}\, p_Y(z-X) \,,$$

which is a special case of the equality

$$(P_X * f)(z) = \mathbb{E}\, f(z-X) \,.$$

Associativity 7d2(c) implies

(7d5) $$(\mu * \nu) * f = \mu * (\nu * f) \quad \text{a.e.}$$

whenever $\mu * \nu$ and $\nu * (f \cdot m)$ are σ-finite (that is, $\nu * f$ is finite a.e).

For measurable $f : \mathbb{R}^d \to \mathbb{R}$ and σ-finite μ we define $\mu * f : \mathbb{R}^d \to \mathbb{R}$ by

(7d6) $$(\mu * f)(x) = \int f(x - \cdot) \, \mathrm{d}\mu \in \mathbb{R}$$

provided that $f(x - \cdot)$ is μ-integrable for almost all x (otherwise $\mu * f$ is undefined). Clearly, (7d6) is consistent with (7d4), and now $\mu * f$ is linear in f (and still positively linear in μ); it means, if $\mu * f_1$ and $\mu * f_2$ are defined, then $\mu * (f_1 - f_2) = \mu * f_1 - \mu * f_2$, etc. Associativity (7d5) still holds, whenever both sides are defined (since it holds for f^+ and f^-).

If $\mu = g \cdot m$ for some $g : \mathbb{R}^d \to [0,\infty)$, then $\int f(x - \cdot) \, \mathrm{d}\mu = \int f(x - \cdot) g(\cdot) \, \mathrm{d}m$. We define the convolution $f * g$ of functions $f, g : \mathbb{R}^d \to \mathbb{R}$ by

$$(g * f)(x) = \int f(x - \cdot) g(\cdot) \, \mathrm{d}m$$

provided that $f(x - \cdot) g(\cdot)$ is m-integrable for almost all x, and get $(g \cdot m) * f = g * f$. Commutativity gives $f * g = g * f$ (take f^+, f^-, g^+, g^-). Associativity gives $(f * g) * h = f * (g * h)$ whenever both sides are defined.

If μ is purely atomic, $\mu(A) = \sum_{y_k \in A} p_k$, then $(\mu * f)(x) = \sum_k p_k f(x - y_k)$ is a linear combination of shifts of f.

It follows that $\|\mu * f\|_p \le \|f\|_p \mu(\mathbb{R})$ for arbitrary $p \in [1,\infty]$ (and moreover, the same holds for every shift-invariant norm) provided that μ is purely atomic. But also if it is not!

7d7 Proposition. For every finite measure μ on $\mathbb{R}$ and every $f \in L_p(m)$ the function $\mu * f$ is defined, and

$$\|\mu * f\|_p \le \|f\|_p \, \mu(\mathbb{R}) \,.$$

7d8 Lemma (Jensen inequality)**.** Let (X, S, μ) be a probability space, $\varphi : \mathbb{R} \to \mathbb{R}$ a convex function, and $f \in L_1(\mu)$; then

$$\varphi\Big(\int_X f \, \mathrm{d}\mu \Big) \le \int_X \varphi \circ f \, \mathrm{d}\mu \in (-\infty, +\infty] \, .$$

Proof. By convexity, there exist $a, b \in \mathbb{R}$ such that $\varphi(u) \ge au + b$ for all u, and $\varphi(u) = au + b$ for $u = \int_X f \, \mathrm{d}\mu$. Then $\varphi \circ f \ge af + b$, whence $\int (\varphi \circ f)^- \le \int (af + b)^- < \infty$, and $\varphi\big(\int_X f \, \mathrm{d}\mu\big) = a \int_X f \, \mathrm{d}\mu + b = \int_X (af + b) \, \mathrm{d}\mu \le \int_X \varphi \circ f \, \mathrm{d}\mu$. □

For $0 < \mu(X) < \infty$ (instead of $\mu(X) = 1$) we get

$$\varphi\Big(\frac{1}{\mu(X)} \int_X f \, \mathrm{d}\mu \Big) \le \frac{1}{\mu(X)} \int_X \varphi \circ f \, \mathrm{d}\mu \, .$$

***Proof** of Prop. 7d7.* WLOG, $\mu(\mathbb{R}) = 1$. We have $|f|^p \in L_1(m)$, therefore $\mu * |f|^p$ is defined, and $\int \mu * |f|^p \, \mathrm{d}m = \int |f|^p \, \mathrm{d}m = \|f\|_p^p$. For almost every x the function $|f(x - \cdot)|^p$ is μ-integrable, that is, $f(x - \cdot) \in L_p(\mu)$; by 5f4, $f(x - \cdot) \in L_1(\mu)$, which shows that $\mu * f$ is defined.

Applying 7d8 to $f(x - \cdot)$ (and $\varphi(\cdot) = |\cdot|^p$) we get $\big| \int f(x - \cdot) \, \mathrm{d}\mu \big|^p \le \int |f(x - \cdot)|^p \, \mathrm{d}\mu$; $|(\mu * f)(x)|^p \le (\mu * |f|^p)(x)$; $|\mu * f|^p \le \mu * |f|^p$ a.e.; thus,

$$\int |\mu * f|^p \, \mathrm{d}m \le \int \mu * |f|^p \, \mathrm{d}m = \int |f|^p \, \mathrm{d}m \, .$$

□

We see that $f * g \in L_p$ whenever $f \in L_1$ and $g \in L_p$.[1]

7e Approximation by convolution

7e1 Proposition. Let probability measures $\mu_1, \mu_2, \dots$ on $\mathbb{R}^d$ satisfy $\forall \varepsilon > 0$ $\mu_n\big(\{x : |x| < \varepsilon\}\big) \to 1$ as $n \to \infty$. Then for every $p \in [1, \infty)$ and every $f \in L_p(m_d)$

$$\|f - \mu_n * f\|_p \to 0 \quad \text{as } n \to \infty \, .$$

Proof. First, let f be continuous and compactly supported, therefore, uniformly continuous. Given ε we take δ such that $|x - y| \le \delta \implies |f(x) -$

[1] In fact, $f * g$ is defined whenever $f \in L_p$, $g \in L_q$ and $\frac{1}{p} + \frac{1}{q} \ge 1$; moreover, $f * g \in L_r$ whenever $f \in L_p$, $g \in L_q$ and $\frac{1}{p} + \frac{1}{q} - 1 = \frac{1}{r}$. See Jones, Chapter 12(B).

$f(y)| \le \varepsilon$, and N such that $n \ge N \implies \mu_n\big(\{x : |x| < \delta\}\big) \ge 1 - \varepsilon$. Then

$$|f(x)-(\mu_n*f)(x)| = \left|f(x)-\int f(x-\cdot)\,\mathrm{d}\mu_n\right| = \left|\int \big(f(x)-f(x-\cdot)\big)\,\mathrm{d}\mu_n\right| \le \\ \le \int |f(x)-f(x-\cdot)|\,\mathrm{d}\mu_n = \int_{|\cdot|<\delta}(\dots) + \int_{|\cdot|\ge\delta}(\dots) \le \varepsilon + 2\varepsilon \max|f|$$

for all x and all $n \ge N$.

Second, for arbitrary $f \in L_p(m_d)$ and $\varepsilon > 0$, by 7c6 there exists a continuous compactly supported g such that $\|f-g\|_p \le \varepsilon$. For n large enough,

$$\|f - \mu_n * f\|_p = \|f-g\|_p + \|g - \mu_n * g\|_p + \|\mu_n * g - \mu_n * f\|_p \le \varepsilon + \varepsilon + \varepsilon$$

by the first part of the proof, since $\|\mu_n * (g-f)\|_p \le \|g-f\|_p$ by 7d7. $\square$

The convergence $\mu_n * f \to f$ is not uniform in f such that $\|f\|_p \le 1$. For example, on $\mathbb{R}$ the function $f : x \mapsto \sin nx$ turns into $(-f)$ being shifted by $\frac{\pi}{n}$.

In particular, we may take the uniform distribution on $[-\frac1n, \frac1n]$ as μ_n, then $\mu_n * f$ are continuous (recall (5d4)), and we get 7c6(a) again. But we can do more.

7e2 Lemma. If μ is a compactly supported finite measure on $\mathbb{R}^d$ and $f \in C(\mathbb{R}^d)$, then $\mu * f \in C(\mathbb{R}^d)$.

Proof. If $x_n \to x$, then $\int f(x_n - \cdot)\,\mathrm{d}\mu \to \int f(x-\cdot)\,\mathrm{d}\mu$ due to the uniform convergence of the integrand on the support of μ. $\square$

7e3 Lemma. If μ is a compactly supported finite measure on $\mathbb{R}^d$ and $f \in C^1(\mathbb{R}^d)$, then $\mu * f \in C^1(\mathbb{R}^d)$, and $\nabla(\mu * f) = \mu * \nabla f$.

Proof. For arbitrary x,

$$\int f(x+h-\cdot)\,\mathrm{d}\mu = \int \big(f(x-\cdot) + \langle \nabla f(x-\cdot), h\rangle + o(h)\big)\,\mathrm{d}\mu$$

where $\frac{o(h)}{|h|} \to 0$ uniformly (due to the uniform continuity of ∇f) on the support of μ; thus,

$$(\mu * f)(x+h) = (\mu * f)(x) + \big(\mu * \langle \nabla f, h\rangle\big)(x) + o(h)\,,$$

which shows that $\nabla(\mu * f) = \mu * \nabla f$, that is, $D_i(\mu * f) = \mu * D_i f$ for $i = 1, \dots, d$. $\square$

7e4 Corollary. By induction, $f \in C^n(\mathbb{R}^d)$ implies $\mu * f \in C^n(\mathbb{R}^d)$, and $D_{i_1} \dots D_{i_n}(\mu * f) = \mu * D_{i_1} \dots D_{i_n} f$ for all $i_1, \dots, i_n \in \{1, \dots, d\}$.

7e5 Corollary. If f is a polynomial of degree $\le n$, then $\mu * f$ is a polynomial of degree $\le n$ (since its $(n+1)$-th derivatives vanish).

(In 7e4 and 7e5 we still assume that μ is finite and compactly supported.)

It is easy to find functions $g_n \in C^n(\mathbb{R}^d)$ that vanish outside the ball $\{x : |x| < 1\}$ and satisfy $g_n \ge 0$,

$$\int_{\mathbb{R}^d} g_n \, \mathrm{d}m = 1 \,, \quad \forall \varepsilon > 0 \quad \int_{|\cdot| < \varepsilon} g_n \, \mathrm{d}m \to 1 \,;$$

for example,

$$g_n(x) = \begin{cases} \mathrm{const}_{n,d} (1 - |x|^2)^{n+1} & \text{for } |x| \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Now, given $p \in [1, \infty)$ and $f \in L_p(m_d)$, we approximate f by $f_n = f * g_n$; by 7e1, $f_n \to f$ in $L_p(m_d)$; and by 7e4, $f_n \in C^n(\mathbb{R}^d)$.

7e6 Remark. You may wonder, whether a measure μ can be thought of as the limit of functions $\mu * g_n$. Yes, in some sense it can. If μ is a finite measure on $\mathbb{R}^d$ and $p_n = \mu * g_n$, then for every bounded continuous $f : \mathbb{R}^d \to \mathbb{R}$,

$$\int f p_n \, \mathrm{d}m \to \int f \, \mathrm{d}\mu \quad \text{as } n \to \infty$$

(since functions $f * g_n$ are uniformly bounded and converge to f). This is the so-called narrow convergence[1] of measures: $p_n \cdot m \to \mu$. However, it does not mean that $(p_n \cdot m)(B) \to \mu(B)$ for Borel sets B.

We may also consider polynomials

$$P_n(x) = \mathrm{const}_{n,d} (1 - |x|^2)^{n+1} \quad \text{for } x \in \mathbb{R}^d \,,$$

the ball $B = \{x : |x| \le 1/2\} \subset \mathbb{R}^d$, and a function $f \in L_p(B)$ extended by 0 outside B. Then $f * P_n = f * g_n$ on B (since $P_n = g_n$ on $B - B$), and $f * P_n$ is a polynomial (by 7e5).

We see that polynomials are dense in L_p on a bounded measurable subset of $\mathbb{R}^d$. But we'll get more.

[1] Or "weak convergence", or "weak* convergence".

7e7 Exercise. Let a function $f : \mathbb{R}^d \to \mathbb{R}$ vanish outside $B = \{x : |x| \le 1/2\}$, and $f_n = f * P_n$.

(a) If $f \in C(\mathbb{R}^d)$, then $f * P_n \to f$ in $C(B)$ (that is, uniformly on B).

(b) If $f \in C^k(\mathbb{R}^d)$, then $f * P_n \to f$ in $C^k(B)$ (that is, uniformly on B, with all derivatives of order $\le k$).

Prove it.

7e8 Corollary (of 7e7(a) and 7c6(c))**.** Let μ be a finite compactly supported measure on $\mathbb{R}^d$.

(a) Polynomials are dense in $L_p(\mu)$ for every $p \in [1, \infty)$.

(b) There exists an orthonormal basis of $L_2(\mu)$ consisting of polynomials.

However, this claim can fail for measures without compact support (try Lebesgue measure), for σ-finite measures (recall 3f2), and for $p = \infty$ (recall 7c4(b)).

7e9 Exercise. Let $G \subset \mathbb{R}^d$ be an open set, and $f : G \to \mathbb{R}$.

(a) If G is bounded, f is bounded, and f is continuous, then there exist polynomials f_n such that $f_n \to f$ uniformly on compacta (that is, uniformly on every compact subset of G). Prove it.

(b) Does (a) hold if f is unbounded? if G is unbounded? if both are unbounded?

7e10 Exercise. Let $B = \{x \in \mathbb{R}^d : |x| \le 1\}$ and $f : B \to \mathbb{R}$.

(a) If $f \in C(B)$, then there exist polynomials f_n such that $f_n \to f$ in $C(B)$.

(b) If $f \in C^k(B)$, then there exist polynomials f_n such that $f_n \to f$ in $C^k(B)$.

Prove it.[1]

Polynomials of a given degree are a shift-invariant finite-dimensional vector space of functions on $\mathbb{R}^d$. Interestingly, each $k \in \mathbb{R}^d$ leads to a two-dimensional shift-invariant space of functions

$$\{x \mapsto a \cos\langle k, x\rangle + b \sin\langle k, x\rangle : a, b \in \mathbb{R}\} ;$$

or rather, a pair $\{-k, k\}$ for $k \ne 0$ leads to a two-dimensional space, and a single point $k = 0$ to a one-dimensional space. Even better, we may turn to complex-valued functions and get a one-dimensional space

$$\{x \mapsto c e^{\mathrm{i}\langle k,x\rangle} : c \in \mathbb{C}\} ,$$

[1] Hint: $f_\varepsilon(x) = f\big((1-\varepsilon)x\big)$.

$c\mathrm{e}^{\mathrm{i}\langle k,x\rangle} = (a+b\mathrm{i})(\cos\langle k,x\rangle + \mathrm{i}\sin\langle k,x\rangle)$. Shift invariance leads to convolution invariance: for $f_k : x \mapsto \mathrm{e}^{\mathrm{i}\langle k,x\rangle}$ and a finite measure μ on $\mathbb{R}^d$ we have

$$\mu * f_k = \Big(\int \mathrm{e}^{-\mathrm{i}\langle k,\cdot\rangle} \, \mathrm{d}\mu \Big) f_k \, ,$$

since $\int f_k(x-\cdot) \, \mathrm{d}\mu = \int \mathrm{e}^{\mathrm{i}\langle k,x-\cdot\rangle} \, \mathrm{d}\mu = \mathrm{e}^{\mathrm{i}\langle k,x\rangle} \int \mathrm{e}^{-\mathrm{i}\langle k,\cdot\rangle} \, \mathrm{d}\mu$.

We restrict ourselves to $k \in \mathbb{Z}^d$ and 2π-periodic (on each coordinate) functions. Linear combinations

$$x \mapsto c_1 \mathrm{e}^{\mathrm{i}\langle k_1,x\rangle} + \dots + c_n \mathrm{e}^{\mathrm{i}\langle k_n,x\rangle}$$

are called trigonometric polynomials. They are an algebra of functions (since $\mathrm{e}^{\mathrm{i}\langle k_1,x\rangle}\mathrm{e}^{\mathrm{i}\langle k_2,x\rangle} = \mathrm{e}^{\mathrm{i}\langle k_1+k_2,x\rangle}$).

Given a function $f \in L_p\big((-\pi,\pi)^d\big)$, we may extend it to $\mathbb{R}^d$ by 2π-periodicity, and this extended function may be thought of as the convolution $\nu_{2\pi\mathbb{Z}^d} * f$ of the counting measure $\nu_{2\pi\mathbb{Z}^d}$ on the lattice $2\pi\mathbb{Z}^d$ and f (this time f is treated as 0 outside $(-\pi,\pi)^d$). By the "periodic" convolution $f *_{2\pi} g$ of two such functions f, g we mean the restriction to $(-\pi,\pi)^d$ of the 2π-periodic function[1]

$$(\nu_{2\pi\mathbb{Z}^d} * f) * g = f * (\nu_{2\pi\mathbb{Z}^d} * g) \, .$$

We introduce trigonometric polynomials

$$P_n(x_1,\dots,x_d) = \mathrm{const}_{n,d}(1+\cos x_1)^n \dots (1+\cos x_d)^n$$

choosing the constants such that $\int_{(-\pi,\pi)^d} P_n \, \mathrm{d}\mu = 1$. Similarly to the algebraic case, for every $f \in L_p\big((-\pi,\pi)^d\big)$ the trigonometric polynomials $f_n = P_n *_{2\pi} f$ satisfy $f_n \to f$ in $L_p\big((-\pi,\pi)^d\big)$. Similarly to 7e8 we conclude that trigonometric polynomials are dense in $L_p\big((-\pi,\pi)^d\big)$ for every $p \in [1,\infty)$, and there exists an orthonormal basis of $L_2\big((-\pi,\pi)^d\big)$ consisting of trigonometric polynomials. In contrast to the algebraic case, here one such basis suggests itself: just

$$\{ x \mapsto (2\pi)^{-d/2} \mathrm{e}^{\mathrm{i}\langle k,x\rangle} : k \in \mathbb{Z}^d \} \, ,$$

since these functions are evidently orthogonal! And for the real-valued functions, the basis consists of the constant function $(2\pi)^{-d/2}$ and functions $x \mapsto (2\pi)^{-d/2}\sqrt{2}\cos\langle k,x\rangle$, $x \mapsto (2\pi)^{-d/2}\sqrt{2}\sin\langle k,x\rangle$ where k runs over a half of $\mathbb{Z}^d \setminus \{0\}$ in the sense that one element should be chosen in each pair $\{-k,k\}$.

[1] This is the convolution on the torus $\mathbb{R}^d/2\pi\mathbb{Z}^d$, in disguise.

7f Outside a small set

A null set can be ignored in Lebesgue's theory; but the terrible function of Sect. 7a is equally terrible outside any null set. It is natural to ask, what can be achieved by excluding a set of small measure.

7f1 Theorem (Lusin[1]). Let $A \subset \mathbb{R}^d$ be a measurable set, $m(A) < \infty$, and $f : A \to \mathbb{R}$ a measurable function. Then for every $\varepsilon > 0$ there exists a compact set $K \subset A$ such that $m(A \setminus K) < \varepsilon$ and the restriction $f|_K$ is continuous (on K).[2]

7f2 Theorem (Egorov[3]). Let (X, S, μ) be a measure space, $\mu(X) < \infty$, and $f_1, f_2, \dots : X \to \mathbb{R}$ measurable functions such that $f_n \to 0$ a.e. Then for every $\varepsilon > 0$ there exists a set $E \in S$ such that $\mu(X \setminus E) < \varepsilon$ and $f_n \to 0$ uniformly on E.

Every convergent sequence of functions is nearly uniformly convergent. (Littlewood)

***Proof** of Th. 7f2.* Similar to the proof of 5c1.

First, we consider monotone convergence: $f_n \downarrow 0$ a.e. For every $\delta > 0$ we have $\mu\{x : f_n(x) > \delta\} \downarrow 0$. We choose $\varepsilon_k > 0$ such that $\sum_k \varepsilon_k \le \varepsilon$, and $\delta_k \to 0$. For each k we take n_k such that

$$\mu \underbrace{\{x : f_{n_k}(x) > \delta_k\}}_{E_k} \le \varepsilon_k \, .$$

The set $E = \cup_k E_k$ satisfies $\mu(E) \le \varepsilon$, and $\sup_{x \in X \setminus E} f_n(x) \le \delta_k$ whenever $n \ge n_k$, which shows that $f_n \to 0$ uniformly on $X \setminus E$.

Second, the general case ($f_n \to 0$ a.e.) reduces to the monotone case by taking $h_n = \sup(|f_n|, |f_{n+1}|, \dots)$ and noting that $|f_n| \le h_n \downarrow 0$ a.e. $\square$

***Proof** of Th. 7f1.* WLOG, f is bounded (otherwise, replace A with $\{x \in A : |f(x)| \le n\}$ where n satisfies $m\{x \in A : |f(x)| > n\} < \varepsilon/2$). By 7c6(a) there exist continuous $f_n : \mathbb{R}^d \to \mathbb{R}$ such that $f_n \to f$ in $L_1(A)$. WLOG, $f_n \to f$ a.e. on A (choose a subsequence, as in 5c5). Th. 7f2 gives a measurable $E \subset A$ such that $m(A \setminus E) < \varepsilon/2$ and $f_n \to f$ uniformly on E. Regularity 2c3 gives a compact $K \subset E$ such that $m(E \setminus K) < \varepsilon/2$. Finally, the limit $f|_K$ of uniformly convergent sequence of continuous functions $f_n|_K$ is continuous. $\square$

---

[1] Or "Luzin".

[2] It does not mean that "the whole f" is continuous at every point of K.

[3] Or "Severini-Egorov", or "Egoroff".

7f3 Exercise. Let (X, S, μ) be a σ-finite measure space, and $f_1, f_2, \dots : X \to \mathbb{R}$ measurable functions such that $f_n \to 0$ a.e. Then there exist sets $E_1, E_2, \dots \in S$ such that $\mu(X \setminus \cup_i E_i) = 0$ and $f_n \to 0$ uniformly on each E_i.

Prove it.

7f4 Exercise. Let μ be a locally finite measure on $\mathbb{R}^d$, and $f : \mathbb{R}^d \to \mathbb{R}$ a μ-measurable function. Then there exist compact sets $K_1, K_2, \dots \subset \mathbb{R}^d$ such that $\mu(\mathbb{R}^d \setminus \cup_i K_i) = 0$ and the restriction $f|_{K_i}$ is continuous (on K_i) for each i.

Prove it.

Index