6 Iterated integral, product measure

6a Iterated integral for $\mathbb{R} \times \mathbb{R}$ and $\mathbb{R}^m \times \mathbb{R}^n$

Below, "measurable" means "Lebesgue measurable", m_1 is Lebesgue measure on $\mathbb{R}$, and m_2 is Lebesgue measure on $\mathbb{R}^2$.

6a1 Theorem (Tonelli)**.** If a function $f : \mathbb{R}^2 \to [0, \infty]$ is measurable, then

(a) its section $f(x, \cdot) : y \mapsto f(x, y)$ is a measurable function $\mathbb{R} \to [0, \infty]$ for almost all $x \in \mathbb{R}$;

(b) the integral of the section

$$x \mapsto \int_{\mathbb{R}} f(x, \cdot) \, \mathrm{d}m_1$$

is a measurable function $\mathbb{R} \to [0, \infty]$ (defined almost everywhere);

(c) $$\int_{\mathbb{R}^2} f \, \mathrm{d}m_2 = \int_{\mathbb{R}} \biggl(x \mapsto \int_{\mathbb{R}} f(x, \cdot) \, \mathrm{d}m_1 \biggr) \, \mathrm{d}m_1 \in [0, \infty] \, .$$

In more traditional notation,

$$\iint f(x, y) \, \mathrm{d}x\mathrm{d}y = \int \biggl(\int f(x, y) \, \mathrm{d}y \biggr) \, \mathrm{d}x = \int \mathrm{d}x \int \mathrm{d}y \, f(x, y) \, .$$

Applying it to $(x, y) \mapsto f(y, x)$ we get

$$\int \mathrm{d}x \int \mathrm{d}y \, f(x, y) = \iint f(x, y) \, \mathrm{d}x\mathrm{d}y = \int \mathrm{d}y \int \mathrm{d}x \, f(x, y) \, .$$

6a2 Remark. Given $f : \mathbb{R} \to [0, \infty]$, we define

$$g(x, y) = \begin{cases} 1 & \text{if } 0 < y < f(x), \\ 0 & \text{otherwise,} \end{cases}$$

then $\int g(x,\cdot)\,\mathrm{d}m = f(x)$ and $\int g(\cdot,y)\,\mathrm{d}m = m\{x : f(x) > y\}$; assuming (for now) measurability of g we get

$$\text{(6a3)} \qquad \int_{\mathbb{R}} f\,\mathrm{d}m = m_2\{(x,y) : 0 < y < f(x)\} = \int_0^\infty m\{x : f(x) > y\}\,\mathrm{d}y\,,$$

that is, both "the distribution formula" 4c13 (Tao, Exer. 1.7.25) and "area interpretation of the integral" (Tao, Exer. 1.7.24) at once. Note that measurability of g implies measurability of f. The converse also holds (take simple $f_n \uparrow f$ and observe that $g_n \uparrow g$).

6a4 Remark. The "area interpretation" sheds new light on the Dominated Convergence Theorem 5d1. First, WLOG, $0 \le f_n \le g_n$ (instead of $|f_n| \le g$), since $f_n + g \to f + g$ a.e., and $0 \le f_n + g \le 2g$. Second, we introduce

$$h_n(x,y) = \begin{cases} 1 & \text{if } 0 < y < f_n(x), \\ 0 & \text{otherwise,} \end{cases} \qquad h(x,y) = \begin{cases} 1 & \text{if } 0 < y < f(x), \\ 0 & \text{otherwise} \end{cases}$$

and note that $f_n \to f$ a.e. implies $h_n \to h$ a.e. (think, why). By 5c1, $h_n \to h$ locally in measure. We treat h_n and h as indicator functions on the measure space $\{(x,y) : 0 < y < g(x)\}$ (subspace of $(\mathbb{R}^2, \mathcal{L}[\mathbb{R}^2], m_2)$). Integrability of g means finiteness of this new measure! Now, the local convergence in measure is in fact global, and implies L_1-convergence (due to boundedness).

Below, by (a), (b), (c) we mean (a), (b), (c) of Theorem 6a1. We abbreviate the phrase "f is measurable and satisfies (a), (b), (c)" to "f satisfies (a), (b), (c)". We also abbreviate the phrase "the indicator $1\!\!1_A$ of a set $A \subset \mathbb{R}^2$ satisfies (a), (b), (c)" to "set A satisfies (a), (b), (c)".

6a5 Exercise. Every box $[a,b) \times [c,d)$ satisfies (a), (b), (c).
Prove it.

6a6 Lemma. Let $f, f_1, f_2, \dots : \mathbb{R}^2 \to [0,\infty]$ and $f_k \uparrow f$. If each f_k satisfies (a), (b), (c), then f satisfies (a), (b), (c).

Proof. Measurability of f_k implies measurability of f by 3c10, and $\int f_k\,\mathrm{d}m_2 \uparrow \int f\,\mathrm{d}m_2$ by Monotone Convergence Theorem 4c11.

For every $x \in \mathbb{R}$ we have $f_k(x,\cdot) \uparrow f(x,\cdot)$, thus, (a) for f_k implies (a) for f by 3c10 (again), and $\int f_k(x,\cdot)\,\mathrm{d}m_1 \uparrow \int f(x,\cdot)\,\mathrm{d}m_1$ by 4c11 (again).

Now, (b) for f_k implies (b) for f by 3c10 (once again), and

$$\int_{\mathbb{R}} \Big(x \mapsto \int_{\mathbb{R}} f_k(x,\cdot)\,\mathrm{d}m_1 \Big)\,\mathrm{d}m_1 \uparrow \int_{\mathbb{R}} \Big(x \mapsto \int_{\mathbb{R}} f(x,\cdot)\,\mathrm{d}m_1 \Big)\,\mathrm{d}m_1$$

by 4c11 (once again).

Now, (c) for f_k implies (c) for f. $\square$

6a7 Corollary. Let $A, A_1, A_2, \dots \subset \mathbb{R}^2$ and $A_k \uparrow A$. If each A_k satisfies (a), (b), (c), then A satisfies (a), (b), (c).

6a8 Exercise. Let $f, g : \mathbb{R}^2 \to [0, \infty]$ satisfy (a), (b), (c). Then

(A) $f + g$ satisfies (a), (b), (c);

(B) if, in addition, $f \le g$ and g is integrable, then $g - f$ satisfies (a), (b), (c).

Prove it.

In particular,

$$\text{(6a9)} \quad \text{if } A \subset B, m(B) < \infty, \text{ and } A, B \text{ satisfy (a), (b), (c),} \qquad \text{then } B \setminus A \text{ satisfies (a), (b), (c).}$$

6a10 Exercise. Every open set $U \subset \mathbb{R}^2$ satisfies (a), (b), (c).

Prove it.[1]

6a11 Exercise. Every compact set $K \subset \mathbb{R}^2$ satisfies (a), (b), (c).

Prove it.[2]

6a12 Lemma (sandwich)**.** Let $A \subset B \subset C \subset \mathbb{R}^2$ and $m(A) = m(C) < \infty$. If A and C satisfy (a), (b), (c), then B satisfies (a), (b), (c).

Proof. By (6a9), $C \setminus A$ satisfies (a), (b), (c). Thus, $\int \mathrm{d}x \int \mathrm{d}y \, \mathbb{1}_{C \setminus A}(x, y) = m(C \setminus A) = 0$; by 4c32, $\int \mathrm{d}y \, \mathbb{1}_{C \setminus A}(x, y) = 0$ for almost all x.

For every such x, $\mathbb{1}_B(x, \cdot)$ is measurable due to sandwich: $\mathbb{1}_A(x, \cdot) \le \mathbb{1}_B(x, \cdot) \le \mathbb{1}_C(x, \cdot)$; thus, B satisfies (a). Also, for such x, $\int \mathbb{1}_A(x, y) \, \mathrm{d}y = \int \mathbb{1}_B(x, y) \, \mathrm{d}y = \int \mathbb{1}_C(x, y) \, \mathrm{d}y$, which implies (b) and (c) for B due to sandwich. $\square$

6a13 Lemma. Every measurable set $A \subset \mathbb{R}^2$ satisfies (a), (b), (c).

Proof. WLOG, A is bounded (due to 6a7). We take compact sets K_i such that $K_i \uparrow K_\infty \subset A$ (but K_∞ need not be compact, of course), and bounded open sets $U_i \downarrow U_\infty \supset A$ (but U_∞ need not be open) such that $m(K_\infty) = m(A) = m(U_\infty)$.

The set K_∞ satisfies (a), (b), (c) by 6a11 and 6a7. The set U_∞ satisfies (a), (b), (c) by the same argument and (6a9), since compact sets $\overline{U}_1 \setminus U_i \uparrow \overline{U}_1 \setminus U_\infty$.

We have $K_\infty \subset A \subset U_\infty$ and $m(U_\infty \setminus K_\infty) = 0$. It remains to apply 6a12. $\square$

[1] Hint: recall 2b3 (and 6a7, and maybe 4c16).

[2] Hint: 6a10 and 6a9.

***Proof** of Th. 6a1.* By 6a13, (a), (b), (c) hold for indicators. By positive linearity (recall 4c15) they holds for simple functions. In general, we take simple f_k such that $f_k \uparrow f$ (recall the phrase before 4c15) and apply 6a6. $\square$

For measurable $f : \mathbb{R}^2 \to [-\infty, +\infty]$,

$$\text{(6a14)} \qquad f \text{ is integrable} \quad \Longleftrightarrow \quad \int \mathrm{d}x \int \mathrm{d}y \, |f(x,y)| < \infty \, .$$

6a15 Theorem (Fubini)**.** If a function $f : \mathbb{R}^2 \to \mathbb{R}$ is integrable, then

(a) its section $f(x, \cdot) : y \mapsto f(x,y)$ is integrable for almost all $x \in \mathbb{R}$;

(b) the integral of the section

$$x \mapsto \int_{\mathbb{R}} f(x, \cdot) \, \mathrm{d}m_1$$

is an integrable function $\mathbb{R} \to \mathbb{R}$ (defined almost everywhere);

$$\text{(c)} \qquad \int_{\mathbb{R}^2} f \, \mathrm{d}m_2 = \int_{\mathbb{R}} \biggl(x \mapsto \int_{\mathbb{R}} f(x, \cdot) \, \mathrm{d}m_1 \biggr) \, \mathrm{d}m_1 \, .$$

For integrable f, again,

$$\int \mathrm{d}x \int \mathrm{d}y \, f(x,y) = \iint f(x,y) \, \mathrm{d}x\mathrm{d}y = \int \mathrm{d}y \int \mathrm{d}x \, f(x,y) \, .$$

Below, by (a), (b), (c) we mean (a), (b), (c) of Theorem 6a15.

6a16 Exercise. If f and g satisfy (a), (b), (c), then $(-f)$ and $f+g$ satisfy (a), (b), (c).

Prove it.

***Proof** of Th. 6a15.* We have $f = f^+ - f^-$, and $f^+, f^- : \mathbb{R}^2 \to [0, \infty)$ are integrable. It remains to apply 6a16. $\square$

6a17 Remark. Theorems 6a1 and 6a15 generalize readily to $\mathbb{R}^m \times \mathbb{R}^n$. The only change in the proof is 6a5: replace $[a,b) \times [c,d)$ with $B_1 \times B_2$ (the product of two boxes). Other changes are trivial replacements of $\mathbb{R}$ with $\mathbb{R}^m$, $\mathbb{R}^n$ as needed, and $\mathbb{R}^2$ with $\mathbb{R}^{m+n}$.

6b Product measure

6b1 Definition. The product $(X_1, S_1) \times (X_2, S_2)$ of two measurable spaces (X_1, S_1) and (X_2, S_2) is the measurable space (X, S) where $X = X_1 \times X_2$ and S is the σ-algebra on X generated by $\{A_1 \times A_2 : A_1 \in S_1, A_2 \in S_2\}$ (or, equivalently, by $\{A_1 \times X_2 : A_1 \in S_1\} \cup \{X_1 \times A_2 : A_2 \in S_2\}$).

6b2 Exercise. $(\mathbb{R}^m, \mathcal{B}[\mathbb{R}^m]) \times (\mathbb{R}^n, \mathcal{B}[\mathbb{R}^n]) = (\mathbb{R}^{m+n}, \mathcal{B}[\mathbb{R}^{m+n}])$.
Prove it.

6b3 Exercise. If $(X, S) = (X_1, S_1) \times (X_2, S_2)$, then for every $A \in S$ and every $x \in X_1$ the section $A_x = \{y \in X_2 : (x, y) \in A\}$ belongs to S_2.
Prove it.[1]

6b4 Exercise. $(\mathbb{R}^m, \mathcal{L}[\mathbb{R}^m]) \times (\mathbb{R}^n, \mathcal{L}[\mathbb{R}^n]) \neq (\mathbb{R}^{m+n}, \mathcal{L}[\mathbb{R}^{m+n}])$.
Prove it.[2,3]

6b5 Remark. Measurability[4] of all sections $A_x = \{y \in X_2 : (x, y) \in A\}$ and $A^y = \{x \in X_1 : (x, y) \in A\}$ is necessary but not sufficient for measurability of A. A counterexample: sets $A_f = \{\big(x, (x, f(x))\big) : x \in \mathbb{R}\} \subset \mathbb{R} \times \mathbb{R}^2$ for arbitrary functions $f : \mathbb{R} \to \mathbb{R}$ have Borel sections (of at most one point); but only a minority (a continuum) of them are Borel sets.

6b6 Theorem. Let (X_1, S_1, μ_1) and (X_2, S_2, μ_2) be two σ-finite measure spaces, and $(X, S) = (X_1, S_1) \times (X_2, S_2)$ the product of the underlying measurable spaces. Then there exists one and only one measure μ on (X, S) such that

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \quad \text{for all } A_1 \in S_1,\ A_2 \in S_2\,,$$

and this measure satisfies

$$\int_{X_1} \big(x \mapsto \mu_2(A_x)\big)\, \mathrm{d}\mu_1 = \mu(A) = \int_{X_2} \big(y \mapsto \mu_1(A^y)\big)\, \mathrm{d}\mu_2$$

for all $A \in S$ (the integrands being measurable).

How to prove existence of μ in Th. 6b6? The problem is, measurability of the integrand $x \mapsto \mu_2(A_x)$ (for arbitrary $A \in S$). We cannot express $\mu_2(A_x \cap B_x)$, $\mu_2(A_x \cup B_x)$ in terms of $\mu_2(A_x)$ and $\mu_2(B_x)$. Disjoint union is

[1] Hint: 3c2 for $y \mapsto (x, y)$.

[2] Hint: use 6b3.

[3] Moreover, $(\mathbb{R}^2, \mathcal{L}[\mathbb{R}^2])$ is not of the form $(\mathbb{R}, S_1) \times (\mathbb{R}, S_2)$; could you prove this fact?

[4] When dealing with measurable spaces (rather than measure spaces) "measurable set" means just "set that belongs to the given σ-algebra".

unproblematic (even countable), as well as $A \setminus B$ when $B \subset A$; but we need more.

It may happen that μ_1, μ_2 are the pushforward measures $f_* m_1$, $g_* m_1$ of Lebesgue measure for some $f : [0, a) \to X_1$, $g : [0, b) \to X_2$; in this case we may take the pushforward measure $h_* m_2$ where $h(s,t) = \big(f(s), g(t)\big)$ for $(s,t) \in [0,a) \times [0,b)$. This approach works in most cases needed for probability theory. Moreover, by some tricks it is possible to reduce the general case to this special case. But we prefer a more familiar way.

We introduce the algebra of sets $\mathcal{E}$ (not σ-algebra!) on X, generated by $\{A_1 \times A_2 : A_1 \in S_1, A_2 \in S_2\}$ (or, equivalently, by $\{A_1 \times X_2 : A_1 \in S_1\} \cup \{X_1 \times A_2 : A_2 \in S_2\}$). Clearly, $\mathcal{E} \subset S$.

6b7 Lemma. In the assumptions of Th. 6b6 assume in addition that $\mu_1(X_1) < \infty$, $\mu_2(X_2) < \infty$.

(a) There exists one and only one additive set function[1] μ_0 on $\mathcal{E}$ such that $\mu_0(A \times B) = \mu_1(A)\mu_2(B)$ for all $A \in S_1$, $B \in S_2$.

(b) For every $E \in \mathcal{E}$, the function $x \mapsto \mu_2(A_x)$ is measurable, and $\mu_0(E) = \int_{X_1} \big(x \mapsto \mu_2(A_x)\big) \, \mathrm{d}\mu_1$.

Proof. Each finite partition $X_1 = A_1 \uplus \dots \uplus A_m$ of X_1 into $A_i \in S_1$ leads to a finite algebra of sets on X_1 (the 2^m unions of these parts).[2] Given also a finite partition $X_2 = B_1 \uplus \dots \uplus B_n$ of X_2 into $B_j \in S_2$, we get a product partition $X = \uplus_{i,j} A_i \times B_j$ of X into mn sets of $\mathcal{E}$, and the product algebra of sets on X (the 2^{mn} unions of these parts).

Two finite algebras on X_1 are always contained (both) in some finite algebra (since two finite partitions have a common refinement). Therefore, two finite product algebras on X are always contained (both) in some finite product algebra.

It follows that the union of all finite product algebras is an algebra; and clearly, it is $\mathcal{E}$.

On each finite product algebra, existence and uniqueness of the needed additive set function are evident. On different algebras these additive set functions conform and, taken together, give μ_0, which proves (a). Also, (b) evidently holds on each finite product algebra (since it holds for $E = A_i \times B_j$), therefore, on the whole $\mathcal{E}$. $\square$

It is possible to adapt the procedure of Sect. 2b–2d, using sets of $\mathcal{E}$ instead of Jordan sets, countable unions of sets of $\mathcal{E}$ instead of open sets, and countable intersections of sets of $\mathcal{E}$ instead of compact sets. But we do not go this way since, fortunately, a better way is well-known.

[1] It means, $\mu_0(E \uplus F) = \mu(E) + \mu(F)$ for all disjoint $E, F \in \mathcal{E}$.

[2] In fact, the general form of a finite algebra of sets.

6b8 Definition. A set M of subsets of a given set X is a *monotone class,* if for all $A_1, A_2, \dots \in M$ and $A \subset X$,

$$A_n \uparrow A \implies A \in M\,, \quad A_n \downarrow A \implies A \in M\,.$$

For example, all intervals (on $\mathbb{R}$) are a monotone class; but all closed intervals are not.

6b9 Exercise. In the assumptions of Th. 6b6 assume in addition that $\mu_1(X_1) < \infty$, $\mu_2(X_2) < \infty$. Then all sets $A \in S$ such that the function $x \mapsto \mu_2(A_x)$ is measurable are a monotone class.

Prove it.[1]

6b10 Exercise. An algebra of sets is a monotone class if and only if it is a σ-algebra.

Prove it.

6b11 Theorem (Monotone class theorem)**.** Let (X, S) be a measurable space, $\mathcal{E} \subset S$ an algebra of sets that generates S, and $M \subset S$ a monotone class containing $\mathcal{E}$. Then $M = S$.

Proof. WLOG, M is the monotone class generated by $\mathcal{E}$. It is sufficient to prove that M is an algebra of sets.

The set $\{A \in M : X \setminus A \in M\}$ is a monotone class (think, why) containing $\mathcal{E}$ (think, why), therefore it is the whole M; that is,

$$\forall A \in M \;\; X \setminus A \in M\,.$$

Given $E \in \mathcal{E}$, the set $\{A \in M : A \cup E \in M\}$ is a monotone class (since $A_n \uparrow A \implies A_n \cup E \uparrow A \cup E$ and $A_n \downarrow A \implies A_n \cup E \downarrow A \cup E$) containing $\mathcal{E}$, therefore it is the whole M; that is,

$$\forall E \in \mathcal{E} \; \forall A \in M \;\; A \cup E \in M\,.$$

Given $B \in M$, the set $\{A \in M : A \cup B \in M\}$ is a monotone class containing $\mathcal{E}$, therefore it is the whole M; that is,

$$\forall A, B \in M \;\; A \cup B \in M\,.$$

$\square$

[1] Hint: use finiteness of μ_2 when treating $A_n \downarrow A$.

***Proof** of Th. 6b6 (existence).* By 6b7(b), 6b9 and Monotone class theorem 6b11, the function $x \mapsto \mu_2(A_x)$ is measurable whenever $A \subset B \times C$, $B \in S_1$, $C \in S_2$, $\mu_1(B) < \infty$, $\mu_2(C) < \infty$ and $A \in S$. Taking $B_n \uparrow X_1$, $C_n \uparrow X_2$ we get $\mu_2\big((A \cap (B_n \times C_n))_x\big) \uparrow \mu_2(A_x)$ for all $A \in S$ (think, why); by 3c10, the latter is measurable in x. We define $\mu(A) = \int_{X_1} \big(x \mapsto \mu_2(A_x)\big)\, \mathrm{d}\mu_1$ and prove that μ is a measure. Clearly, $\mu(\emptyset) = 0$; and $\mu(A_1 \uplus A_2 \uplus \dots) = \mu(A_1) + \mu(A_2) + \dots$ by 4c16. $\square$

Now about uniqueness of the product measure, and more generally, about uniqueness of a measure with given restriction to a generating algebra of sets.

6b12 Lemma. Let (X, S, μ) be a measure space, $\mu(X) < \infty$, and $\mathcal{E} \subset S$ an algebra of sets that generates S. If ν is a measure on (X, S) such that $\nu|_{\mathcal{E}} = \mu|_{\mathcal{E}}$, then $\nu = \mu$.

Proof. By the Monotone class theorem 6b11 it is sufficient to prove that the set $M = \{A \in S : \nu(A) = \mu(A)\}$ is a monotone class. If $A_n \uparrow A$, then $\nu(A) = \lim_n \nu(A_n) = \lim_n \mu(A_n) = \mu(A)$. If $A_n \downarrow A$, then $X \setminus A_n \uparrow X \setminus A$, thus $\nu(X \setminus A) = \mu(X \setminus A)$, that is, $\nu(X) - \nu(A) = \mu(X) - \mu(A)$; and $\nu(X) = \mu(X) < \infty$, since $X \in \mathcal{E}$. $\square$

6b13 Remark. Note that $\mathcal{E}$ is an algebra, not just a set.[1] Two different measures may coincide on the generating set $\{A_1 \times X_2 : A_1 \in S_1\} \cup \{X_1 \times A_2 : A_2 \in S_2\}$ (even if they are finite). Probabilistically: the joint distribution of two random variables says much more than the pair of marginal distributions.

***Proof** of Th. 6b6 (uniqueness).* The proof of the existence part of Th. 6b6 gives two measures

$$\mu_{1,2}(A) = \int_{X_1} \big(x \mapsto \mu_2(A_x)\big)\, \mathrm{d}\mu_1\,, \quad \mu_{2,1}(A) = \int_{X_2} \big(y \mapsto \mu_1(A^y)\big)\, \mathrm{d}\mu_2$$

on (X, S), satisfying $\mu_{1,2}|_{\mathcal{E}} = \mu_0 = \mu_{2,1}|_{\mathcal{E}}$. If μ_1 and μ_2 are finite, then $\mu_{1,2} = \mu_{2,1}$ by 6b12. In general, $\mu_{1,2}(A) = \mu_{2,1}(A)$ whenever $A \subset B \times C$, $B \in S_1$, $C \in S_2$, $\mu_1(B) < \infty$, $\mu_2(C) < \infty$ and $A \in S$. We take $B_n \uparrow X_1$, $C_n \uparrow X_2$ and get $\mu_{1,2}(A) = \lim_n \mu_{1,2}\big(A \cap (B_n \times C_n)\big) = \lim_n \mu_{2,1}\big(A \cap (B_n \times C_n)\big) = \mu_{2,1}(A)$. $\square$

6b14 Example. Without σ-finiteness of μ_1 and μ_2 measures $\mu_{1,2}$ and $\mu_{2,1}$ can differ.

[1] Beware of Remark 6.9 in the textbook by Capiński and Kopp: "in order to show that two measures on a σ-field coincide it suffices to prove that they coincide on the generating sets of that σ-field".

Let (X_1, S_1, μ_1) be $\mathbb{R}$ with Lebesgue measure, and (X_2, S_2, μ_2) be $\mathbb{R}$ with the counting measure: $\forall y \in \mathbb{R}\ \ \mu_2(\{y\}) = 1$ (S_2 may be the Borel σ-algebra, or the whole $2^{\mathbb{R}}$). In the product $X = \mathbb{R} \times \mathbb{R}$ we consider the diagonal $A = \{(x, y) : x = y\}$; being a Borel set, it belongs to S. We have $A_x = \{x\}$, $A^y = \{y\}$, $\mu_2(A_x) = 1$, $\mu_1(A^y) = 0$, thus, $\mu_{1,2}(A) = \int_{\mathbb{R}} 1 \, \mathrm{d}m = \infty$, but $\mu_{2,1}(A) = \int_{\mathbb{R}} 0 \, \mathrm{d}\mu_2 = 0$.

The incomplete product of two σ-finite measure spaces (X_1, S_1, μ_1) and (X_2, S_2, μ_2) is, by definition, the (evidently σ-finite) measure space (X, S, μ) where $(X, S) = (X_1, S_1) \times (X_2, S_2)$ and μ is as in Theorem 6b6. The completion of this measure space is the complete product. One calls this μ the *product measure* and writes $\mu = \mu_1 \times \mu_2$ and $(X, S, \mu) = (X_1, S_1, \mu_1) \times (X_2, S_2, \mu_2)$ (be it complete or incomplete).

Clearly, $\mathbb{R}^{m+n}$ with Lebesgue measure is the complete product of $\mathbb{R}^m$ and $\mathbb{R}^n$, each with Lebesgue measure.

The next two theorems (generalizations of 6a1 and 6a15) hold whenever $(X, S, \mu) = (X_1, S_1, \mu_1) \times (X_2, S_2, \mu_2)$ is the complete product of two complete σ-finite measure spaces.

6b15 Theorem (Tonelli)**.** If a function $f : X \to [0, \infty]$ is measurable, then

(a) its section $f(x, \cdot) : y \mapsto f(x, y)$ is a measurable function $X_2 \to [0, \infty]$ for almost all $x \in X_1$;

(b) the integral of the section $x \mapsto \int_{X_2} f(x, \cdot) \, \mathrm{d}\mu_2$ is a measurable function $X_1 \to [0, \infty]$ (defined almost everywhere);

(c) $$\int_X f \, \mathrm{d}\mu = \int_{X_1} \biggl(x \mapsto \int_{X_2} f(x, \cdot) \, \mathrm{d}\mu_2 \biggr) \, \mathrm{d}\mu_1 \in [0, \infty] \, .$$

6b16 Theorem (Fubini)**.** If a function $f : X \to \mathbb{R}$ is integrable, then

(a) its section $f(x, \cdot) : y \mapsto f(x, y)$ is integrable for almost all $x \in X_1$;

(b) the integral of the section $x \mapsto \int_{X_2} f(x, \cdot) \, \mathrm{d}\mu_2$ is an integrable function $X_1 \to \mathbb{R}$ (defined almost everywhere);

(c) $$\int_X f \, \mathrm{d}\mu = \int_{X_1} \biggl(x \mapsto \int_{X_2} f(x, \cdot) \, \mathrm{d}\mu_2 \biggr) \, \mathrm{d}\mu_1 \, .$$

6b17 Remark. If f is measurable on $(X_1, S_1) \times (X_2, S_2)$ (not completed), then the reservations "for almost all" and "defined almost everywhere" in these two theorems are not needed.

But in general these reservations are needed. Indeed, let $Z \subset \mathbb{R}$ be a null set and $Y \subset \mathbb{R}$ not (Lebesgue) measurable; then $f = 1\!\!1_{Z \times Y}$ is measurable, but for $x \in Z$ the section $f(x, \cdot) = 1\!\!1_Y$ is not.

Also, completeness of (X_1, S_1, μ_1) and (X_2, S_2, μ_2) is needed. Indeed, a section of a Lebesgue measurable function need not be Borel (try $f(x, y) = g(y)$); and even if it is Borel, its integral need not be Borel (try $f(x, y) = g(x)h(y)$).

Remark 6a2 generalizes readily. Also, for $f : X \to [-\infty, +\infty]$, measurability of f is equivalent to measurability of the function

$$g(x, y) = \begin{cases} -1 & \text{if } f(x) < y < 0, \\ +1 & \text{if } 0 < y < f(x), \\ 0 & \text{otherwise,} \end{cases}$$

and in this case

$$\int_X f \, \mathrm{d}\mu = (\mu \times m)\{(x, y) : 0 < y < f(x)\} - (\mu \times m)\{(x, y) : f(x) < y < 0\} = \\ = \int_0^\infty \mu\{x : f(x) > y\} \, \mathrm{d}y - \int_0^\infty \mu\{x : f(x) < -y\} \, \mathrm{d}y$$

(recall (4d3)), four cases being possible: (number) − (number), ∞ − (number), (number) − ∞, and $\infty - \infty$.

Toward proving the theorems, first we reduce the complete case to the incomplete case.

6b18 Exercise. Let (X, S, μ) be a measure space, and $(X, \overline{S}, \overline{\mu})$ its completion. Then every equivalence class of measurable functions on $(X, \overline{S}, \overline{\mu})$ contains a function measurable on (X, S).

Prove it (a) for indicator functions, (b) for simple functions, and (c) in general.[1]

In particular, every Lebesgue measurable function on $\mathbb{R}^d$ is equivalent (that is, equal almost everywhere) to some (non-unique, of course) Borel function. Moreover, the same holds for all measures on $\mathbb{R}^d$ (as before, I mean completed Borel measures, recall the end of Sect. 3e). However, different measures require different Borel functions (unless the given function is already Borel).[2]

Below, by (a), (b), (c) we mean (a), (b), (c) of Theorem 6b15.

6b19 Exercise. Let $(X, \overline{S}, \overline{\mu}) = (X_1, S_1, \mu_1) \times (X_2, S_2, \mu_2)$ be the complete product of two complete σ-finite measure spaces, the completion of their incomplete product (X, S, μ).

[1] Hint: (c) $f_k \uparrow f$.

[2] Try atomic measures (and a universally measurable function).

(A) A set $A \in S$ is a null set if and only if almost all sections A_x are null sets.

(B) If (a), (b), (c) hold for a function f, then they hold for every function equivalent to f.
Prove it.[1]

6b20 Exercise. Generalize 6a6.

***Proof** of Th. 6b15.* By 6b18 and 6b19(B), WLOG, f is measurable on $(X_1, S_1) \times (X_2, S_2)$ (not completed). (In this case we'll prove a bit more, as promised in 6b17.) By 6b6, (a), (b), (c) hold for indicators. The rest is similar to the proof of Th. 6a1. □

6b21 Exercise. Generalize 6a16.

***Proof** of Th. 6b16.* Similar to the proof of Th. 6a15. □

By default, by product space we mean the complete product space.

6b22 Exercise. Let $(X, S, \mu) = (X_1, S_1, \mu_1) \times (X_2, S_2, \mu_2)$.

(a) If $f \in L_1(\mu_1)$, $g \in L_1(\mu_2)$ and $h(x, y) = f(x)g(y)$ for $x \in X_1$, $y \in X_2$, then $h \in L_1(\mu)$, $\|h\| = \|f\|\|g\|$ and $\int h \, d\mu = \big(\int f \, d\mu_1\big)\big(\int g \, d\mu_2\big)$.

(b) If $f_1 : X_1 \to [0, \infty)$ and $f_2 : X_2 \to [0, \infty)$ are measurable, then $(f_1 \cdot \mu_1) \times (f_2 \cdot \mu_2) = g \cdot \mu$ where $g(x, y) = f_1(x) f_2(y)$.

Prove it.

For instance, the product of two copies of the standard Gaussian measure $\gamma_1 = \varphi_1 \cdot m_1$ on $\mathbb{R}$ (recall 4c29), where $\varphi_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for $x \in \mathbb{R}$, is the standard Gaussian measure $\gamma_2 = \varphi_2 \cdot m_2$ on $\mathbb{R}^2$, where $\varphi_2(x) = \frac{1}{2\pi} e^{-|x|^2/2}$ for $x \in \mathbb{R}^2$. Interestingly, γ_2 is invariant under rotations.

6c Introduction to independence [2]

6c1 Definition. Two random variables $X, Y : \Omega \to \mathbb{R}$ on a probability space $(\Omega, \mathcal{F}, P)$ are *independent,* if

$$P_{X,Y} = P_X \times P_Y \,.$$

[1] Hint: (A) use 6b6; (B) use (A).

[2] "At a purely formal level, one could call probability theory the study of measure spaces with total measure one, but that would be like calling number theory the study of strings of digits which terminate. At a practical level, the opposite is true: just as number theorists study concepts (e.g. primality) that have the same meaning in every numeral system that models the natural numbers, we shall see that probability theorists study concepts (e.g. independence) that have the same meaning in every measure space that models a family of events or random variables." Terence Tao, A review of probability theory; also Terence Tao Quotes.

Here $P_{X,Y}$ is the joint distribution (recall Sect. 3d).

For instance, if the joint distribution $P_{X,Y}$ is the standard Gaussian measure γ_2, then X, Y are independent. Moreover, in this case[1] $X+Y$ and $X-Y$ are also independent.

Independence means that $P_{X,Y}(A \times B) = P_X(A)P_Y(B)$, that is, $\mathbb{P}\big(X \in A, Y \in B\big) = \mathbb{P}\big(X \in A\big)\mathbb{P}\big(Y \in B\big)$ for all Borel sets $A, B \in \mathbb{R}$. The Borel σ-algebra $\mathcal{B}[\mathbb{R}]$ is generated by the algebra of sets $\mathcal{E}_1$ generated by intervals; and $\mathcal{B}[\mathbb{R}^2]$ is generated by the algebra of sets $\mathcal{E}_2$ generated by boxes (rectangles). By 6b12, it is sufficient to check the equality $P_{X,Y} = P_X \times P_Y$ on $\mathcal{E}_2$. That is,

$$\begin{gathered} X, Y \text{ are independent} \quad \text{if and only if} \\ \text{for all intervals } I, J \subset \mathbb{R} \quad \mathbb{P}\big(X \in I, Y \in J\big) = \mathbb{P}\big(X \in I\big)\mathbb{P}\big(Y \in J\big) \,. \end{gathered}$$

Moreover, intervals $(-\infty, a]$ (and their products) are enough (think, why); thus,

$$X, Y \text{ are independent} \quad \text{if and only if} \quad \forall x, y \in \mathbb{R} \;\; F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

where $F_X(x) = \mathbb{P}\big(X \le x\big)$, $F_Y(y) = \mathbb{P}\big(Y \le y\big)$, and $F_{X,Y}(x,y) = \mathbb{P}\big(X \le x, Y \le y\big)$ are the so-called cumulative distribution functions.

Note also that (by 6b12 again)

(6c2) $$P_X \text{ is uniquely determined by } F_X \,.$$

For every Borel set $B \in \mathbb{R}$, the probability $\mathbb{P}\big(X \in B\big)$ can be calculated out of the probabilities $\mathbb{P}\big(X \le x\big)$; but this calculation may involve a lot of "zigzags" (recall Sect. 3b).

Similarly (irrespective of independence),

(6c3) $$P_{X,Y} \text{ is uniquely determined by } F_{X,Y} \,.$$

The change of variable (4c23) gives

$$\mathbb{E}\, f(X) = \int_{\mathbb{R}} f \, \mathrm{d}P_X \,, \quad \mathbb{E}\, g(Y) = \int_{\mathbb{R}} g \, \mathrm{d}P_Y \,, \quad \mathbb{E}\, h(X,Y) = \int_{\mathbb{R}^2} h \, \mathrm{d}P_{X,Y}$$

for Borel $f, g : \mathbb{R} \to \mathbb{R}$, $h : \mathbb{R}^2 \to \mathbb{R}$. (The four cases are possible...) For independent X, Y, taking $h(x,y) = f(x)g(y)$, we get

(6c4) $$\mathbb{E}\,\big(f(X)g(Y)\big) = \big(\mathbb{E}\, f(X)\big)\big(\mathbb{E}\, g(Y)\big)$$

by 6b22(a), assuming integrability of $f(X)$ and $g(Y)$. In particular,

(6c5) $$\mathbb{E}\,(XY) = (\mathbb{E}\, X)(\mathbb{E}\, Y) \quad \text{for independent integrable } X, Y \,.$$

[1] And only in this case!

6d Introduction to conditioning

Consider the product $(\Omega, \mathcal{F}, P) = (\Omega_1, \mathcal{F}_1, P_1) \times (\Omega_2, \mathcal{F}_2, P_2)$ of two probability spaces. For every event $A \in \mathcal{F}$, 6b15 gives

$$P(A) = \mathbb{E}_1 X \quad \text{where} \quad X(\omega_1) = P_2(A_{\omega_1}) \text{ for } \omega_1 \in \Omega_1$$

(the expectation being taken on $(\Omega_1, \mathcal{F}_1, P_1)$). In particular, for (finite or) countable Ω_1 (assuming all points of Ω_1 are not null) we have

$$P_2(A_{\omega_1}) = \frac{P(A \cap B_{\omega_1})}{P(B_{\omega_1})}$$

where sets $B_{\omega_1} = \{\omega_1\} \times \Omega_2$ are a (finite or) countable partition of Ω. In the discrete probability framework the ratio above is called the conditional probability of the event A given the event B_{ω_1};

$$X(\omega_1) = P_2(A_{\omega_1}) = \mathbb{P}\big(A \big| B_{\omega_1}\big) \, ; \quad X(\cdot) = \mathbb{P}\big(A \big| \cdot\big) \, ;$$

thus,

$$\mathbb{P}\big(A\big) = \mathbb{E}_1 \mathbb{P}\big(A \big| \cdot\big) = \sum_{\omega_1 \in \Omega_1} \mathbb{P}\big(A \big| B_{\omega_1}\big) \mathbb{P}\big(B_{\omega_1}\big) \, ,$$

the so-called total probability formula.

In general, P_2 need not be atomic; the denominator may vanish; and still,

$$\mathbb{P}\big(A\big) = \mathbb{E}_1 \mathbb{P}\big(A \big| \cdot\big)$$

if we define

$$\mathbb{P}\big(A \big| B_{\omega_1}\big) = P_2(A_{\omega_1}) = P_2\{\omega_2 : (\omega_1, \omega_2) \in A\} \, .$$

More generally, given an integrable random variable $X : \Omega \to \mathbb{R}$, we have the total expectation formula

$$\mathbb{E}\, X = \mathbb{E}_1 \mathbb{E}\big(X \big| \cdot\big)$$

where the conditional expectation $\mathbb{E}\big(X \big| \omega_1\big)$ is defined as $\int_{\Omega_2} X(\omega_1, \cdot) \, \mathrm{d}P_2$. In particular, for $X = 1\!\!1_A$ we return to the conditional probability and the total probability.

Also,

$$\mathbb{E}\,(X 1\!\!1_{A_1 \times \Omega_2}) = \mathbb{E}_1\big(\mathbb{E}\big(X \big| \cdot\big) 1\!\!1_{A_1}\big) \quad \text{for all } A_1 \in \mathcal{F}_1 \, ;$$

assuming $X \ge 0$ we may rewrite it as

$$(X \cdot P)(A_1 \times \Omega_2) = \big(\mathbb{E}\big(X \big| \cdot\big) \cdot P_1\big)(A_1) \, ,$$

that is,

$$\varphi_*(X \cdot P) = \mathbb{E}\big(X \big| \cdot\big) \cdot P_1 \tag{6d1}$$

where $\varphi : \Omega \to \Omega_1$ is the projection, $\varphi(\omega_1, \omega_2) = \omega_1$.

Index