2 Probability, random elements, random sets

After introducing probability measures we discuss random objects of two kinds.

2a Probability space, measure algebra

We know that algebras are easy but σ-algebras are not. Surprisingly, a measure makes σ-algebras easy by neglecting null sets. In particular, the Lebesgue σ-algebra is simpler than the Borel σ-algebra.

2a1 Definition. (a) A *probability measure* on a measurable space $(X, \mathcal{A})$ is a map $\mu : \mathcal{A} \to [0,1]$ such that $\mu(X) = 1$ and $\mu(A_1 \cup A_2 \cup \dots) = \mu(A_1) + \mu(A_2) + \dots$ whenever sets $A_1, A_2, \dots \in \mathcal{A}$ are pairwise disjoint.

(b) A *probability space* is a triple $(X, \mathcal{A}, \mu)$ such that $(X, \mathcal{A})$ is a measurable space and μ is a probability measure on $(X, \mathcal{A})$.

2a2 Example. One and only one probability measure μ on $([0,1], \mathcal{B}[0,1])$ satisfies $\mu([0,x]) = x$ for all $x \in [0,1]$. This μ is often called "Lebesgue measure on $[0,1]$", but I prefer to add "restricted to the Borel σ-algebra".

2a3 Definition. Let $(X, \mathcal{A}, \mu)$ be a probability space, and $A \subset X$.

(a) *Inner measure* $\mu_*(A)$ and *outer measure* $\mu^*(A)$ of A are defined by

$$\mu_*(A) = \max\{\mu(B) : B \in \mathcal{A}, B \subset A\}, \quad \mu^*(A) = \min\{\mu(B) : B \in \mathcal{A}, B \supset A\};$$

(b) A is a *null* (or *negligible*) set[1] if $\mu^*(A) = 0$;

(c) A is a *μ-measurable* set (symbolically, $A \in \mathcal{A}_\mu$) if $\mu_*(A) = \mu^*(A)$.

[1]Terence Tao calls it "sub-null set", requiring $A \in \mathcal{A}$ for null sets. For some authors, "null" means just "empty".

In contrast, sets of $\mathcal{A}$ will be called $\mathcal{A}$-measurable. Clearly, $\mathcal{A} \subset \mathcal{A}_\mu$; that is, $\mathcal{A}$-measurability implies μ-measurability. For μ of 2a2, $\mathcal{A}_\mu$ is the Lebesgue σ-algebra on $[0,1]$.

2a4 Core exercise. Prove that (a) the maximum and minimum in 2a3(a) are reached; and (b) $\mu_*(A) = 1 - \mu^*(X \setminus A)$.

2a5 Core exercise. Prove that $\mu^* : 2^X \to [0,1]$ satisfies

(a) $\mu^*(\emptyset) = 0$,

(b) $\mu^*(A) \le \mu^*(B)$ whenever $A \subset B \subset X$ ("monotonicity");

(c) $\mu^*(A_1 \cup A_2 \cup \dots) \le \mu^*(A_1) + \mu^*(A_2) + \dots$ for all $A_1, A_2, \dots \subset X$ ("countable subadditivity").

Deduce that the null sets are a *σ-ideal,* that is, every subset of a null set is a null set, and a countable union of null sets is a null set.

If A is a null set then its complement $X \setminus A$ is called a set of *full measure* (or *conegligible*), and one says that $x \notin A$ for *almost all* x, in other words, *almost everywhere* or *almost surely* ("a.s.").

2a6 Definition. Let $(X, \mathcal{A}, \mu)$ be a probability space, and $A, B \subset X$.

(a) The *distance* between A and B is[1]

$$\operatorname{dist}(A,B) = \mu^*(A \triangle B)\,;$$

here the *symmetric difference* $A \triangle B = (A \setminus B) \cup (B \setminus A) = \{x : \mathbf{1}_A(x) \ne \mathbf{1}_B(x)\}$.

(b) Sets A, B are *equivalent* or *almost equal* (symbolically, $A \overset{\mu}{\sim} B$) if $\operatorname{dist}(A,B) = 0$, that is, their symmetric difference is a null set.

Note that

$$\operatorname{dist}(A,B) = \min\{\mu^*(C) : A \cap (X \setminus C) = B \cap (X \setminus C)\}\,. \tag{2a7}$$

("By throwing C away we get $A = B$"...)

2a8 Core exercise. Prove that $\operatorname{dist} : 2^X \times 2^X \to [0,1]$ is a *pseudometric,* that is,

(a) $\operatorname{dist}(A,A) = 0$,

(b) $\operatorname{dist}(A,B) = \operatorname{dist}(B,A)$ ("symmetry"),

(c) $\operatorname{dist}(A,C) \le \operatorname{dist}(A,B) + \operatorname{dist}(B,C)$ ("triangle inequality").

[1] Not at all the distance between their closest points...

2a9 Core exercise. Let $A_n, B_n \subset X$, $\operatorname{dist}(A_n, B_n) = \varepsilon_n$, then
(a) $\operatorname{dist}(A_1 \cap A_2, B_1 \cap B_2) \le \varepsilon_1 + \varepsilon_2$,
(b) $\operatorname{dist}(A_1 \cup A_2, B_1 \cup B_2) \le \varepsilon_1 + \varepsilon_2$,
(c) $\operatorname{dist}(A_1 \setminus A_2, B_1 \setminus B_2) \le \varepsilon_1 + \varepsilon_2$,
(d) $\operatorname{dist}(A_1 \triangle A_2, B_1 \triangle B_2) \le \varepsilon_1 + \varepsilon_2$.
Prove it. Generalize (a), (b) to finite and countable operations. Deduce that $\forall n \; A_n \stackrel{\mu}{\sim} B_n$ implies $A_1 \cap A_2 \stackrel{\mu}{\sim} B_1 \cap B_2$, $A_1 \cup A_2 \stackrel{\mu}{\sim} B_1 \cup B_2$, $A_1 \setminus A_2 \stackrel{\mu}{\sim} B_1 \setminus B_2$ and $A_1 \triangle A_2 \stackrel{\mu}{\sim} B_1 \triangle B_2$. Generalize the first two relations to countable operations.

2a10 Core exercise. (a) A set is μ-measurable if and only if it is equivalent to some $\mathcal{A}$-measurable set.
(b) $\mathcal{A}_\mu$ is a σ-algebra.
(c) there exists one and only one probability measure $\bar\mu$ on $(X, \mathcal{A}_\mu)$ such that $\bar\mu|_{\mathcal{A}} = \mu$.
Prove it.

The probability space $(X, \mathcal{A}_\mu, \bar\mu)$ is called the *completion* of $(X, \mathcal{A}, \mu)$. It is *complete;* that is, the given σ-algebra contains all null sets. Sometimes I'll write $\mu(A)$ (instead of $\bar\mu(A)$) for $A \in \mathcal{A}_\mu$.

Consider the set $2^X/\stackrel{\mu}{\sim}$ of all equivalence classes. According to 2a9 we may apply finite and countable (but not uncountable) operations to equivalence classes; and the usual properties (such as $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$) hold. The relation $A \subset B$ for $A, B \in 2^X/\stackrel{\mu}{\sim}$ may be defined by $\mu^*(A \setminus B) = 0$, and is equivalent to $A = A \cap B$ (as well as $B = A \cup B$). Also the distance is well-defined on $2^X/\stackrel{\mu}{\sim}$, and is a *metric;* that is, a pseudometric satisfying

$$\operatorname{dist}(A, B) = 0 \quad \Longrightarrow \quad A = B \qquad \text{for } A, B \in 2^X/\stackrel{\mu}{\sim}.$$

Note that $\mu_*, \mu^* : 2^X/\stackrel{\mu}{\sim} \to [0, 1]$ are well-defined.

The set of all equivalence classes of measurable sets,[1]

$$\mathcal{A}/\stackrel{\mu}{\sim} = \mathcal{A}_\mu/\stackrel{\mu}{\sim},$$

is called the *measure algebra* of $(X, \mathcal{A}, \mu)$. It is closed under the finite and countable operations.[2]

Note that $\mu : \mathcal{A}/\stackrel{\mu}{\sim} \to [0, 1]$ is well-defined.

[1] "Many of the difficulties of measure theory and all the pathology of the subject arise from the existence of sets of measure zero. The algebraic treatment gets rid of this source of unpleasantness by refusing to consider sets at all; it considers sets modulo sets of measure zero instead." (P.R. Halmos, "Lectures on ergodic theory", Math. Soc. Japan 1956, p. 42.)

[2] In fact, $\mathcal{A}/\stackrel{\mu}{\sim}$ is a complete Boolean algebra, and is topologically closed in $2^X/\stackrel{\mu}{\sim}$.

Recall $\sim\mathcal{E}, \mathcal{E}_{\mathrm{d}}, \mathcal{E}_{\mathrm{s}}, \mathcal{E}_\delta, \mathcal{E}_\sigma$ introduced in Sect. 1(a,b) for arbitrary $\mathcal{E} \subset 2^X$. Now they are also well-defined for arbitrary $\mathcal{E} \subset 2^X/\overset{\mu}{\sim}$. Similarly to 1a10 and 1b8, a set $\mathcal{E} \subset 2^X/\overset{\mu}{\sim}$ is called an *algebra* if it contains $\sim\mathcal{E}$, $\mathcal{E}_{\mathrm{d}}$, $\mathcal{E}_{\mathrm{s}}$, and *σ-algebra,* if it contains $\sim\mathcal{E}$, $\mathcal{E}_\delta$, $\mathcal{E}_\sigma$.

2a11 Core exercise. Measure subalgebras, that is, sub-σ-algebras of $\mathcal{A}/\overset{\mu}{\sim}$, are in a natural bijective correspondence with sub-σ-algebras of $\mathcal{A}_\mu$ that contain all null sets.[1]

Formulate it accurately, and prove.

We know (recall 1a21 and 1c18 or the paragraph before 1b8) that $\mathcal{E}_{\mathrm{ds}} = \mathcal{E}_{\mathrm{sd}}$ but generally $\mathcal{E}_{\delta\sigma} \ne \mathcal{E}_{\sigma\delta}$ for $\mathcal{E} \subset 2^X$.

2a12 Proposition. For every algebra $\mathcal{E} \subset \mathcal{A}/\overset{\mu}{\sim}$ the set $\mathcal{E}_{\delta\sigma}$ is a σ-algebra, and $\mathcal{E}_{\delta\sigma} = \mathcal{E}_{\sigma\delta}$.

A surprise: no more $\mathcal{E}_{\sigma\delta\sigma}$ and the like!

In particular, for an arbitrary probability measure on the Cantor set we have $F_\sigma/\overset{\mu}{\sim} = G_\delta/\overset{\mu}{\sim}$ (recall 1b7).

2a13 Extra exercise. Does the equality $\mathcal{E}_{\delta\sigma} = \mathcal{E}_{\sigma\delta}$ hold for arbitrary sets (not just algebras) $\mathcal{E} \subset \mathcal{A}/\overset{\mu}{\sim}$?

The "sandwich" below implies 2a12 and gives substantially more detailed information.

2a14 Proposition. Let $\mathcal{E} \subset \mathcal{A}$ be an algebra and $B \in \sigma(\mathcal{E})$; then there exist $A \in \mathcal{E}_{\delta\sigma}$ and $C \in \mathcal{E}_{\sigma\delta}$ such that $A \subset B \subset C$ and $A \overset{\mu}{\sim} B \overset{\mu}{\sim} C$.

For an arbitrary probability measure on the Cantor set we get $A \in F_\sigma$, $C \in G_\delta$. In particular, every null set is contained in a G_δ null set. In order to get the same for a probability measure on $\mathbb{R}^d$ we generalize 2a14.

Given a set $\mathcal{F} \subset \mathcal{A}$, we introduce

$$\mu_{*\mathcal{F}}(A) = \sup\{\mu(B) : B \in \mathcal{F},\, B \subset A\} \quad \text{for } A \subset X\,.$$

2a15 Proposition. If $\mathcal{F} \subset \mathcal{A}$ satisfies $\mathcal{F}_{\mathrm{s}} \subset \mathcal{F}$ and $\mathcal{F}_\delta \subset \mathcal{F}$ then the set

$$\{\, A \in \mathcal{A}_\mu \,:\, \mu_{*\mathcal{F}}(A) = \mu(A) \,\wedge\, \mu_{*\mathcal{F}}(X \setminus A) = \mu(X \setminus A)\,\}$$

is a σ-algebra.

Let us denote this σ-algebra by Sandwich($\mathcal{F}$).

Clearly, the set $\mathcal{F}$ of all closed (or only compact) subsets of $\mathbb{R}^d$ satisfies $\mathcal{F}_{\mathrm{s}} \subset \mathcal{F}$ and $\mathcal{F}_\delta \subset \mathcal{F}$.

[1] That is, all subsets of all measure 0 sets of the whole σ-algebra $\mathcal{A}$.

2a16 Core exercise. Let $X = \mathbb{R}^d$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^d)$, $\mathcal{F}$ be the set of all closed subsets of $\mathbb{R}^d$, and μ a probability measure on $(X, \mathcal{A})$. Then

(a) $\mathcal{A} \subset \text{Sandwich}(\mathcal{F})$.

(b) $\mathcal{A}_\mu = \text{Sandwich}(\mathcal{F})$.

Deduce it from 2a15. Do the same for compact (rather than closed) subsets of $\mathbb{R}^d$.

2a17 Core exercise. If $\mathcal{F}_\delta \subset \mathcal{F}$ then the set $\mathcal{E} = \{A \in \mathcal{A}_\mu : \mu_{*\mathcal{F}}(A) = \mu(A)\}$ satisfies $\mathcal{E}_\delta = \mathcal{E}$.

Prove it.

2a18 Core exercise. If $\mathcal{F} \subset \mathcal{A}$ satisfies $\mathcal{F}_{\text{s}} \subset \mathcal{F}$ then

$$\{A \in \mathcal{A}_\mu : \mu_{*\mathcal{F}}(A) = \mu(A)\} = \{A \in \mathcal{A}_\mu : \exists B \in \mathcal{F}_\sigma \; (B \subset A \wedge B \overset{\mu}{\sim} A) \} \,.$$

Prove it.

2a19 Core exercise. For arbitrary $\mathcal{F} \subset \mathcal{A}$ the set $\mathcal{E} = \{A \in \mathcal{A}_\mu : \exists B \in \mathcal{F}_\sigma \; (B \subset A \wedge B \overset{\mu}{\sim} A) \}$ satisfies $\mathcal{E}_\sigma \subset \mathcal{E}$.

Prove it.

2a20 Core exercise. Prove Prop. 2a15.

2a21 Core exercise. Prove Prop. 2a14.

Here is another useful implication of the sandwich.

2a22 Proposition. If two probability measures on $(X, \mathcal{A})$ coincide on an algebra $\mathcal{E} \subset \mathcal{A}$ then they coincide on $\sigma(\mathcal{E})$.

Proof. Due to the relation $\mathcal{E}_{\text{d}} \subset \mathcal{E}$, for every $A \in \mathcal{E}_\delta$ there exist $A_n \in \mathcal{E}$ such that $A_n \downarrow A$ (that is, $A_1 \supset A_2 \supset \dots$ and $A_1 \cap A_2 \cap \dots = A$); thus, $\mu(A) = \lim \mu(A_n) = \lim \nu(A_n) = \nu(A)$, which shows that $\mu = \nu$ on $\mathcal{E}_\delta$. Further, due to the relation $\mathcal{E}_{\delta\text{s}} \subset \mathcal{E}_\delta$ (recall 1b1), for every $A \in \mathcal{E}_{\delta\sigma}$ there exist $A_n \in \mathcal{E}_\delta$ such that $A_n \uparrow A$, which shows that $\mu = \nu$ on $\mathcal{E}_{\delta\sigma}$. Similarly, $\mu = \nu$ on $\mathcal{E}_{\sigma\delta}$.

Given $B \in \sigma(\mathcal{E})$, 2a14 gives us $A \in \mathcal{E}_{\delta\sigma}$ and $C \in \mathcal{E}_{\sigma\delta}$ such that $A \subset B \subset C$ and $\mu(A) = \mu(B) = \mu(C)$. Thus, $\mu(B) = \mu(A) = \nu(A) \le \nu(B) \le \nu(C) = \mu(C) = \mu(B)$. □

Uniqueness in 2a2 follows easily.

2a23 Warning. Let $(X, \mathcal{A}, \mu)$ be a probability space and $\mathcal{B} \subset \mathcal{A}$ a sub-σ-algebra, then $(X, \mathcal{B}, \mu|_{\mathcal{B}})$ is another probability space. Be careful with the completed σ-algebra $\mathcal{B}_{\mu|_{\mathcal{B}}}$; it need not contain all null sets of $(X, \mathcal{A}, \mu)$. That is, $\mu^*(A) = 0$ does not imply $(\mu|_{\mathcal{B}})^*(A) = 0$. Here is a counterexample. Let $X = [0,1] \times [0,1]$ with the two-dimensional Lebesgue measure μ on the Borel σ-algebra $\mathcal{A}$. The first coordinate $[0,1] \times [0,1] \ni (x,y) \mapsto x \in [0,1]$ generates a sub-σ-algebra $\mathcal{B} \subset \mathcal{A}$. The diagonal $A = \{(x,y) : x = y\}$ satisfies $\mu^*(A) = 0$ but $(\mu|_{\mathcal{B}})^*(A) = 1$ (think, why).

We have three generally different σ-algebras: $\mathcal{B}$, $\mathcal{B}_1 = \mathcal{B}_{\mu|_{\mathcal{B}}}$, and $\mathcal{B}_2$ consisting of all sets equivalent to $\mathcal{B}$-measurable sets; clearly, $\mathcal{B}/\overset{\mu}{\sim} = \mathcal{B}_1/\overset{\mu}{\sim} = \mathcal{B}_2/\overset{\mu}{\sim}$. On the other hand, if $\mathcal{B} = \mathcal{A}$ then $\mathcal{B}_1 = \mathcal{B}_2$ by 2a10(a).

2b Standard models

The Cantor set provides standard models again.

Similarly to 1b17, every subset $\mathcal{E} \subset \mathcal{A}/\overset{\mu}{\sim}$ of the measure algebra generates a measure subalgebra $\sigma(\mathcal{E})/\overset{\mu}{\sim} \subset \mathcal{A}/\overset{\mu}{\sim}$, defined as the intersection of all measure subalgebras that contain $\mathcal{E}$. The corresponding (in the sense of 2a11) σ-algebra $\sigma(\mathcal{E}) \subset \mathcal{A}_\mu$ (containing all null sets) is the σ-algebra *generated* by $\mathcal{E}$. Similarly to 1d28, $\mathcal{A}/\overset{\mu}{\sim}$ is called countably generated, if $\mathcal{A}/\overset{\mu}{\sim} = \sigma(A_1, A_2, \dots)/\overset{\mu}{\sim}$ for some $A_1, A_2, \dots \in \mathcal{A}/\overset{\mu}{\sim}$.

2b1 Core exercise. If $\mathcal{A}$ is countably generated then $\mathcal{A}/\overset{\mu}{\sim}$ is countably generated (for every μ on $\mathcal{A}$).

Prove it.

By 1d29 and 2b1, every probability measure on $\mathbb{R}^d$ leads to a countably generated measure algebra.

The converse to 2b1 is generally wrong. (The Lebesgue σ-algebra on $[0,1]$ is countably separated but, in fact, not countably generated.) Rather, $\mathcal{A}/\overset{\mu}{\sim}$ is countably generated if and only if $\mathcal{A}/\overset{\mu}{\sim} = \mathcal{B}/\overset{\mu}{\sim}$ for some countably generated σ-algebra $\mathcal{B} \subset \mathcal{A}$.

Using again the idea of (1a18) and (1d36),

$$\varphi(x) = \big(\mathbf{1}_{A_1}(x), \mathbf{1}_{A_2}(x), \dots\big) \tag{2b2}$$

we see, similarly to 1d38, that $\mathcal{A}/\overset{\mu}{\sim}$ is countably generated if and only if $\mathcal{A}/\overset{\mu}{\sim}$ is generated by some measurable $\varphi : X \to \{0,1\}^\infty$. It means that φ is measurable from $(X, \mathcal{A})$ to $(Y, \mathcal{B})$, where $Y = \{0,1\}^\infty$, $\mathcal{B}$ is the usual σ-algebra on $\{0,1\}^\infty$ (as in 1d16), and $\Phi(\mathcal{B})/\overset{\mu}{\sim} = \mathcal{A}/\overset{\mu}{\sim}$ (recall 1d13(a)). In this case we introduce the image measure $\nu = \mu \circ \Phi$ on $\{0,1\}^\infty$,

$$\nu(B) = \mu\big(\varphi^{-1}(B)\big) \quad \text{for } B \in \mathcal{B}\,,$$

and observe that Φ establishes an isomorphism of measure algebras $\mathcal{A}/\overset{\mu}{\sim}$ and $\mathcal{B}/\overset{\nu}{\sim}$.

However, it does not mean that φ is an isomorphism of probability spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$, for two reasons.

First, in general φ need not be one-to-one. Though, if $(X, \mathcal{A})$ is a Borel space (recall 1d33) then φ is one-to-one (recall 1d39).

Second, in general φ need not map X onto Y, and worse, the image $\varphi(X) \subset Y$ need not be of full measure; rather, $\nu^*(\varphi(X)) = 1$ (think, why), but $\nu_*(\varphi(X))$ need not be 1, which is sometimes called "the image measure catastrophe".

We may replace $(X, \mathcal{A}, \mu)$ with $(Y, \mathcal{B}, \nu)$ when dealing with events (treated up to null sets). What about random variables?

Depending on context, by a (real-valued) random variable one means either a measurable function $X \to \mathbb{R}$ or an equivalence class of such functions.

2b3 Definition. Let $(X, \mathcal{A}, \mu)$ be a probability space.

(a) Two μ-measurable (that is, $\mathcal{A}_\mu$-measurable) functions $f, g : X \to \mathbb{R}$ are *equivalent* if $f(x) = g(x)$ for almost all x;

(b) $L_0(X, \mathcal{A}, \mu)$ is the set of all equivalence classes of μ-measurable functions $X \to \mathbb{R}$.

Functions defined almost everywhere may be used equally well. The well-known spaces L_p are subsets of L_0 (endowed with topologies).

Clearly, we have an embedding

$$L_0(Y, \mathcal{B}, \nu) \ni g \mapsto f \in L_0(X, \mathcal{A}, \mu) , \quad f(\cdot) = g(\varphi(\cdot)) ,$$

and these f, g are *identically distributed,* that is,

$$\mu\{x : f(x) \in B\} = \nu\{y : g(y) \in B\} \quad \text{for every } B \in \mathcal{B}(\mathbb{R}) \tag{2b4}$$

(think, why). It is less clear whether $L_0(Y, \mathcal{B}, \nu)$ is thus mapped onto $L_0(X, \mathcal{A}, \mu)$. Given f, we cannot construct g by just $g(\cdot) = f(\varphi^{-1}(\cdot))$, since φ need not be invertible. And nevertheless...

2b5 Proposition. [1] For every $f \in L_0(X, \mathcal{A}, \mu)$ there exists $g \in L_0(Y, \mathcal{B}, \nu)$ such that $f(\cdot) = g(\varphi(\cdot))$ a.s.[2]

2b6 Lemma. Prop. 2b5 holds for f, g with values in the Cantor set.

[1]This is basically the Doob-Dynkin lemma.

[2]This "a.s." is rather inappropriate, since f and g are equivalence classes... but this is a usual and convenient abuse of language. From now on it will occur without notice.

Proof. Treating the Cantor set as $\{0,1\}^\infty$ we have

$$f(x) = \big(\mathbf{1}_{A_1}(x), \mathbf{1}_{A_2}(x), \dots\big) \quad \text{a.s.}$$

for some $A_1, A_2, \dots \in \mathcal{A}_\mu$. We take $B_1, B_2, \dots \in \mathcal{B}_\nu$ such that $A_n \overset{\mu}{\sim} \varphi^{-1}(B_n)$, that is, $\mathbf{1}_{A_n}(\cdot) = \mathbf{1}_{B_n}(\varphi(\cdot))$ a.s., define g by $g(y) = \big(\mathbf{1}_{B_1}(y), \mathbf{1}_{B_2}(y), \dots\big)$ and get $g(\varphi(x)) = \big(\mathbf{1}_{B_1}(\varphi(x)), \mathbf{1}_{B_2}(\varphi(x)), \dots\big) = \big(\mathbf{1}_{A_1}(x), \mathbf{1}_{A_2}(x), \dots\big) = f(x)$ for almost all x. □

2b7 Lemma. Let E be a Borel subset of the Cantor set; then Prop. 2b5 holds for f, g with values in E.

Proof. First, 2b6 gives $g : Y \to C$ (the Cantor set) such that $f(\cdot) = g(\varphi(\cdot))$ a.s. By (2b4), $\nu\{y : g(y) \in E\} = \mu\{x : f(x) \in E\} = \mu(X) = 1$; that is, $g(\cdot) \in E$ a.s. □

2b8 Lemma. The measurable space $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ is isomorphic to a Borel subset of the Cantor set.

Proof. We replace $\mathbb{R}$ with the interval $(0,1)$, since they are homeomorphic. We use binary digits:

$$x = \sum_{n=1}^\infty 2^{-n} \beta_n(x) \quad \text{for } x \in (0,1)\,,$$

$\beta_n : (0,1) \to \{0,1\}$, $\liminf_n \beta_n(x) = 0$. Each β_n, being a step function, is Borel measurable. Therefore the map

$$(0,1) \ni x \mapsto \big(\beta_1(x), \beta_2(x), \dots\big) \in \{0,1\}^\infty$$

is Borel measurable (recall 1d18 and the phrase after it). The inverse map is Borel measurable by 1d8 applied to binary intervals $\big[k \cdot 2^{-n}, (k+1) \cdot 2^{-n}\big)$. □

Proof of Proposition 2b5. Follows from 2b7 and 2b8. □

2b9 Definition. A measurable space $(X, \mathcal{A})$ is *standard,* if it is isomorphic to some Borel subset of the Cantor set.

Clearly, such space is a Borel space (recall 1d33); thus, it is usually called "standard Borel space".

Recall 1c2, 1c3.

2b10 Definition. A *subspace* of a measurable space $(X, \mathcal{A})$ is a measurable space of the form $(A, \mathcal{A}|_A)$ where $A \in \mathcal{A}$, and $\mathcal{A}|_A = \{A \cap B : B \in \mathcal{A}\} = \{B \in \mathcal{A} : B \subset A\}$.

2b11 Core exercise. (a) A subspace of a standard Borel space is standard;
(b) the product of two standard Borel spaces is standard;
(c) the product of countably many standard Borel spaces is standard.
Prove it.

By 2b8, $\mathbb{R}$ is standard. By 2b11(b), $\mathbb{R}^d$ is standard. By 2b11(a), every Borel subset of $\mathbb{R}^d$ is a standard Borel space.

Proposition 2b5 (with its proof) still holds for f, g with values in a given standard Borel space.

2c Random elements

Random functions etc.

A remark on terminology. Some authors call a measurable map φ from a probability space to a measurable space $(Y, \mathcal{B})$ "random element of Y" with special cases like "random vector", "random sequence", "random function" etc. That is nice, but two objections arise. First, no one says "random real number" instead of "random variable"! Second, some people insist that, say, a "random vector" must be a single vector chosen at random, or somehow typical, but surely not a map!

Other authors call such φ "Y-valued random variable" which avoids the two objections mentioned above but leads to cumbersome phrases like "$\mathbb{R}^T$-valued random variable" instead of "random function". Also, one complains that a so-called random variable is neither random nor variable!

Another choice is, between functions and equivalence classes. I hesitate but I must choose…

2c1 Definition. Let $(X, \mathcal{A}, \mu)$ be a probability space and $(Y, \mathcal{B})$ a measurable space.

(a) A *random element* of Y is an equivalence class (w.r.t. the equivalence relation "equal almost everywhere") of measurable functions $X \to Y$.

(b) The *distribution* of a random element φ is the measure ν_φ on $(Y, \mathcal{B})$ defined by

$$\nu_\varphi(B) = \mu\big(\Phi(B)\big) \,,$$

where $\Phi(B)$ is the equivalence class of $\{x : \varphi(x) \in B\}$.

(c) The measure subalgebra $\sigma(\varphi)/\overset{\mu}{\sim} \subset \mathcal{A}/\overset{\mu}{\sim}$ *generated* by φ is defined by

$$\sigma(\varphi)/\overset{\mu}{\sim} = \Phi(\mathcal{B}) = \{\Phi(B) : B \in \mathcal{B}\}$$

(where Φ is as in (b)); the corresponding (in the sense of 2a11) σ-algebra $\sigma(\varphi) \subset \mathcal{A}_\mu$ (containing all null sets) is the σ-algebra *generated* by φ.

2c2 Example. Let $(X, \mathcal{A}, \mu)$ be the interval $(0, 1)$ with Lebesgue measure, $(Y, \mathcal{B})$ the Cantor set $\{0, 1\}^\infty$, and $\varphi : X \to Y$ the "binary digits" map used in the proof of 2b8. Its distribution $\nu = \nu_\varphi$ satisfies

$$\nu\{x : x(1) = a_1, \dots, x(n) = a_n\} = 2^{-n} \tag{2c3}$$

for all $n = 1, 2, \dots$ and $a_1, \dots, a_n \in \{0, 1\}$. Infinite tossing of a fair coin! (Known also as Bernoulli process.) It follows from 2a22 that (2c3) is satisfied by only one measure ν on $\{0, 1\}^\infty$ (sometimes called Lebesgue measure on the Cantor set); thus, $\big(\{0, 1\}^\infty, \nu\big)$ is an example of the *infinite product* of probability spaces.[1]

The Borel space $\big(\{0, 1\}^\infty\big)^\infty = \{0, 1\}^{\infty\times\infty}$ is still $\{0, 1\}^{(\text{a countable set})}$ (recall the paragraph before (1c21)). Thus, the product ν^∞ of countably many copies of ν is also well-defined;

$$\nu^\infty\{(x_1, x_2, \dots) : x_1 \in B_1, \dots, x_n \in B_n\} = \nu(B_1) \dots \nu(B_n)$$

for all n and Borel sets $B_1, \dots, B_n \subset \{0, 1\}^\infty$.

The "almost inverse" map

$$\{0, 1\}^\infty \ni x \mapsto \sum_n 2^{-n} x(n) \in [0, 1]$$

sends ν to Lebesgue measure μ. Applying it to each x_n we get the product map $\big(\{0, 1\}^\infty\big)^\infty \to [0, 1]^\infty$ that sends ν^∞ to μ^∞, the infinite product of copies of Lebesgue measure on $[0, 1]$ (or $(0, 1)$, it is the same), sometimes called Lebesgue measure on $(0, 1)^\infty$; existence of μ^∞ is thus proved (while its uniqueness follows from 2a22).

Taking into account that every probability measure on $\mathbb{R}$ is the image of Lebesgue measure on $(0, 1)$ under an increasing function $\varphi : \mathbb{R} \to \mathbb{R}$ well-known as the quantile function (or inverse cumulative distribution function) of the measure, we get existence (and uniqueness, as before) of $\mu_1 \times \mu_2 \times \dots$ for arbitrary probability measures $\mu_1, \mu_2, \dots$ on $\mathbb{R}$.

Every probability measure ν on $(Y, \mathcal{B})$ is the distribution of some random element, for a trivial reason: just take $(X, \mathcal{A}, \mu) = (Y, \mathcal{B}, \nu)$ and $\varphi(x) = x$. It is less clear that X can be the Cantor set whenever $(Y, \mathcal{B})$ is standard. For the proof, embed Y into the Cantor set: $\psi : Y \to \psi(Y) \subset C$, consider the image measure $\mu(\cdot) = \nu(\psi^{-1}(\cdot))$ on C, and take measurable $\varphi : C \to Y$ such that $\varphi(c) = y$ whenever $\psi(y) = c$.

[1] Only finite product is defined for measure spaces in general; but for an infinite sequence of probability spaces the product is well-defined (though I neither prove nor use it).

Moreover, (C, μ) can be replaced with $(0,1)$ and Lebesgue measure. Indeed, every probability measure μ on $\mathbb{R}$ (in particular, on $C \subset \mathbb{R}$) is the image of Lebesgue measure on $(0,1)$. Thus:

2c4 Proposition. Every probability measure on a standard Borel space is the distribution of some (at least one) random element defined on $(0,1)$ with Lebesgue measure.

2c5 Corollary. The infinite product of probability measures on standard Borel spaces is well-defined.[1]

Here is another useful feature of standard Borel spaces, — a generalization of the fact that each Lebesgue measurable function on $[0,1]$ is equivalent to some Borel measurable function.

2c6 Core exercise. Let $(X, \mathcal{A}, \mu)$ be a probability space, $(Y, \mathcal{B})$ a standard Borel space, φ a random element of Y, and $\mathcal{A}_1 \subset \mathcal{A}$ a sub-σ-algebra such that $\sigma(\varphi)/\overset{\mu}{\sim} \subset \mathcal{A}_1/\overset{\mu}{\sim}$ (that is, each set of $\sigma(\varphi)$ is equivalent to some set of $\mathcal{A}_1$; note that $\mathcal{A}_1$ need not contain all null sets in its completion, recall 2a23). Then the equivalence class φ contains some $\mathcal{A}_1$-measurable map.

Prove it.

2c7 Extra exercise. Does 2c6 hold for arbitrary (not just standard) measurable space $(Y, \mathcal{B})$?

By 2b8 and 2b11, in order to prove that $(Y, \mathcal{B})$ is standard it is sufficient to find appropriate "coordinates" on Y, that is, measurable functions $f_1, f_2, \dots : Y \to \mathbb{R}$ such that the map $f : Y \to \mathbb{R}^\infty$, $f(y) = \big(f_1(y), f_2(y), \dots\big)$, is an isomorphism between Y and its image $f(Y) \subset \mathbb{R}^\infty$, and the image is a Borel set.

The Hilbert space l_2 (over $\mathbb{R}$) consists of all sequences $x = (x_1, x_2, \dots) \in \mathbb{R}^\infty$ such that $\|x\|^2 = x_1^2 + x_2^2 + \dots < \infty$. Its Borel σ-algebra $\mathcal{B}(l_2)$ may be defined as generated by linear functionals $x \mapsto \langle x, y \rangle = x_1 y_1 + x_2 y_2 + \dots$ for all $y \in l_2$.

2c8 Core exercise. (a) l_2 is a measurable subspace of $\mathbb{R}^\infty$; that is, the σ-algebra on l_2 induced from $\mathbb{R}^\infty$ is equal to $\mathcal{B}(l_2)$.

(b) The unit ball $\{x \in l_2 : \|x\| \le 1\}$ is a Borel subset of $\mathbb{R}^\infty$.

(c) l_2 is a Borel subset of $\mathbb{R}^\infty$, and therefore a standard Borel space.

Prove it.

[1] In fact, it is well-defined for arbitrary (not just standard) measurable spaces (as was noted before).

Every separable Hilbert space is isomorphic to l_2 and therefore is also a standard Borel space.

The set Q of all rational numbers of $[0,1]$ being countable, $\mathbb{R}^Q$ is isomorphic to $\mathbb{R}^\infty$, therefore standard.

2c9 Core exercise. Each of the following subsets of $\mathbb{R}^Q$ is measurable, and is therefore a standard Borel space:

(a) all increasing functions $Q \to \mathbb{R}$;

(b) all increasing right-continuous functions $Q \to \mathbb{R}$ (that is, $f(q) = \lim_{Q \ni r \to q+} f(r)$ for all $q \in Q \setminus \{1\}$);

(c) all uniformly continuous functions $Q \to \mathbb{R}$.

Prove it.

The measurable space $\mathbb{R}^{[0,1]}$ (with the product σ-algebra) is of little use (recall 1d26) but contains useful measurable subspaces.

2c10 Core exercise. Each of the following measurable subspaces of $\mathbb{R}^{[0,1]}$ is a standard Borel space:

(a) all increasing right-continuous functions $[0,1] \to \mathbb{R}$;

(b) all increasing functions $f : [0,1] \to \mathbb{R}$ such that $f(x) = \big(f(x-) + f(x+)\big)/2$ for all $x \in (0,1)$;

(c) all continuous functions $[0,1] \to \mathbb{R}$.

Prove it.

2c11 Core exercise. The following measurable subspace of $\mathbb{R}^{[0,1]}$ is separated but not countably separated, therefore not a Borel space:

(a) all increasing functions $[0,1] \to \mathbb{R}$.

Prove it.

A wonder: the notion "a random increasing function" is problematic, but the notion "a random increasing right-continuous function" is not.

2c12 Extra exercise. (a) On the set of all increasing functions $[0,1] \to \mathbb{R}$ invent a better σ-algebra that turns it into a Borel space;

(b) is this Borel space standard?

A function $f : \mathbb{R} \to \mathbb{R}$ is called a *corlol*[1] function (or RCLL[2] or càdlàg[3] function) if it has two limits $f(x-)$, $f(x+)$ at every $x \in \mathbb{R}$, and $f(x) = f(x+)$ for every $x \in \mathbb{R}$.[4] For a function $f : [0,1] \to \mathbb{R}$ the definition is similar but,

[1] "Continuous on (the) right, limit on (the) left".

[2] "Right continuous with left limits".

[3] "Continue à droite, limite à gauche" (French).

[4] Sample functions of many stochastic processes in continuous time (martingales, Markov processes etc.) are corlol functions.

of course, without $f(0-)$ and $f(1+)$. Waiving the right continuity (but retaining $f(x+)$) we get a "function without discontinuities of the second kind".

2c13 Extra exercise. Each of the following measurable subspaces of $\mathbb{R}^{[0,1]}$ is a standard Borel space:

(a) all corlol functions $[0,1] \to \mathbb{R}$;

(b) all functions $f : [0,1] \to \mathbb{R}$ without discontinuities of the second kind, satisfying $f(x) = \big(f(x-) + f(x+)\big)/2$ for all $x \in (0,1)$.

Prove it.

2d Random sets

Strangely enough, a random Borel set is not a random element of the Borel σ-algebra but a measurable subset of a product space.

2d1 Core exercise. Let $(X, \mathcal{A}, \mu)$ be a probability space, and $M \subset L_0(X, \mathcal{A}, \mu)$ be such that

(a) $c \cdot \mathbf{1}_A \in M$ for all $c \in [0,1]$ and $A \in \mathcal{A}$;

(b) if $f \in M$ then $\mathbf{1} - f \in M$;

(c) if $f_1, f_2, \dots \in M$ and $f = \sup_n f_n$ (a.s.) then $f \in M$.

Then M contains all $f \in L_0(X, \mathcal{A}, \mu)$ such that $0 \le f \le 1$ a.s.

Prove it.

We may replace "$M \subset L_0(X, \mathcal{A}, \mu)$" with "$M \subset 2^X/\overset{\mu}{\sim}$" and consider the least M satisfying (a), (b), (c); this M is exactly the set of all $f \in L_0(X, \mathcal{A}, \mu)$ such that $0 \le f \le 1$ a.s.

In the same spirit we may define a random Borel set. Similarly to 2c1(a), it is an equivalence class of maps $X \to \mathcal{B}(\mathbb{R})$, the equivalence being equality almost everywhere; measurability is hidden, similarly to 2d1.

2d2 Definition. Let $(X, \mathcal{A}, \mu)$ be a probability space.

(a) The set $\mathrm{RBS}(X, \mathcal{A}, \mu)$ is the least set RBS of equivalence classes of maps $X \to \mathcal{B}(\mathbb{R})$ such that

(a1) for arbitrary $A \in \mathcal{A}$ and $B \in \mathcal{B}(\mathbb{R})$ the map

$$\omega \mapsto \begin{cases} B & \text{for } \omega \in A, \\ \emptyset & \text{for } \omega \in X \setminus A \end{cases}$$

belongs to RBS;

(a2) if $S \in \mathrm{RBS}$ then the map

$$\omega \mapsto \mathbb{R} \setminus S(\omega)$$

belongs to RBS;

(a3) if $S_1, S_2, \dots \in \text{RBS}$ and $S(\cdot) = \cup_n S_n(\cdot)$ (a.s.) then $S \in \text{RBS}$.

(b) Elements of $\text{RBS}(X, \mathcal{A}, \mu)$ will be called *random Borel sets* (on $(X, \mathcal{A}, \mu)$).

Clearly, the intersection of all sets satisfying (a1), (a2), (a3) is a set satisfying (a1), (a2), (a3).

The "random Borel set" may seem to be a new notion, but it is rather the old notion "set of the product σ-algebra" in disguise! In order to see this fact we note a natural bijective correspondence between maps $X \to 2^Y$ (for now, X and Y are just two sets) and subsets of $X \times Y$. Every set $S \subset X \times Y$ leads to the map $x \mapsto \{y : (x, y) \in S\}$. And every map $f : X \to 2^Y$ leads to the set $\{(x, y) : x \in X \wedge y \in f(x)\}$.

Reformulating 2d2(a) in terms of subsets of $X \times \mathbb{R}$ we get

$$\begin{gathered} A \in \mathcal{A},\, B \in \mathcal{B}(\mathbb{R}) \quad \Longrightarrow \quad A \times B \in \text{RBS}\,, \\ S \in \text{RBS} \quad \Longrightarrow \quad (X \times \mathbb{R}) \setminus S \in \text{RBS}\,, \\ S_1, S_2, \dots \in \text{RBS} \quad \Longrightarrow \quad S_1 \cup S_2 \cup \dots \in \text{RBS}\,. \end{gathered}$$

Comparing this with 1d15 (and the phrase after it) we see that $\text{RBS} = \mathcal{A} \times \mathcal{B}(\mathbb{R})$, up to equivalence. This time, a subset of $X \times \mathbb{R}$ is negligible if it is contained in $A \times \mathbb{R}$ for some null set $A \subset X$; and "equivalent" means "differ on a negligible set only". Thus we have an equivalent definition.

2d3 Definition. Let $(X, \mathcal{A}, \mu)$ be a probability space. A *random Borel set* (on $(X, \mathcal{A}, \mu)$) is an equivalence class of maps $X \to \mathcal{B}(\mathbb{R})$ that contains the map

$$x \mapsto \{y : (x, y) \in S\}$$

for some $(\mathcal{A} \times \mathcal{B}(\mathbb{R}))$-measurable $S \subset X \times \mathbb{R}$.

It is convenient to denote by S both the set and the map (and the equivalence class); just denote $S(x) = \{y : (x, y) \in S\}$.

2d4 Core exercise. (a) Prove that Definitions 2d2 and 2d3 are equivalent.

(b) Prove that "$B \in \mathcal{B}(\mathbb{R})$" may be replaced with "$B \subset \mathbb{R}$ is an interval" in 2d2(a1); that is, the modified definition is equivalent to the original one.

(c) Prove that "$\mathcal{A} \times \mathcal{B}(\mathbb{R})$" may be replaced with "$\mathcal{A}_\mu \times \mathcal{B}(\mathbb{R})$" in 2d3.

2d5 Core exercise. (a) *"random interval"* Let $f, g \in L_0(X, \mathcal{A}, \mu)$ and

$$S(x) = \{y \in \mathbb{R} : f(x) \le y \le g(x)\} \quad \text{for } x \in X\,;$$

prove that S is a random Borel set.

(b) *"random countable set"* Let $f_1, f_2, \dots \in L_0(X, \mathcal{A}, \mu)$ and

$$S(x) = \{f_1(x), f_2(x), \dots\} \quad \text{for } x \in X\,;$$

prove that S is a random Borel set.

2d6 Extra exercise. Let $f \in L_0(X, \mathcal{A}, \mu)$, $B \in \mathcal{B}(\mathbb{R})$, and

$$S(x) = f(x) + B = \{f(x) + y : y \in B\} \quad \text{for } x \in X\,;$$

prove that S is a random Borel set.

You may wonder, why not treat a random Borel set as just a random element of $\mathcal{B}(\mathbb{R})$. You may suggest some appropriate σ-algebras on $\mathcal{B}(\mathbb{R})$. But, strangely enough, you cannot succeed! That is, you never get a definition equivalent to 2d2, 2d3, no matter which σ-algebra on $\mathcal{B}(\mathbb{R})$ is used. The problem is that in general a random Borel set fails to generate a σ-algebra, in contrast to a random element.

2d7 Core exercise. Let $(X, \mathcal{A}, \mu)$, $(Y, \mathcal{B})$, a random element $\varphi : X \to Y$ and its generated σ-algebra $\sigma(\varphi) \subset \mathcal{A}_\mu$ be as in 2c1. Prove that $\sigma(\varphi)$ is the least among all σ-algebras $\mathcal{A}_1 \subset \mathcal{A}_\mu$ such that $\mathcal{A}_1$ contains all null sets and φ is also a random element on $(X, \mathcal{A}_1, \mu|_{\mathcal{A}_1})$.

2d8 Extra exercise. The reservation "$\mathcal{A}_1$ contains all null sets" in 2d7 cannot be dropped. Moreover: let $(X, \mathcal{A}, \mu)$, $(Y, \mathcal{B})$ and $\varphi : X \to Y$ be as in 2d7. Consider all σ-algebras $\mathcal{A}_1 \subset \mathcal{A}_\mu$ such that φ is also a random element on $(X, \mathcal{A}_1, \mu|_{\mathcal{A}_1})$ (but this time $\mathcal{A}_1$ is not required to contain all null sets). What about the least among these $\mathcal{A}_1$? Prove by a counterexample that it need not exist.

Given a random Borel set S on $(X, \mathcal{A}, \mu)$, we may consider all σ-algebras $\mathcal{A}_1 \subset \mathcal{A}_\mu$ such that $\mathcal{A}_1$ contains all null sets and S is also a random Borel set on $(X, \mathcal{A}_1, \mu|_{\mathcal{A}_1})$. In this case we say that S is $\mathcal{A}_1$-measurable. If there exists the least among these $\mathcal{A}_1$, we call it the σ-algebra *generated* by S, and denote it $\sigma(S)$. But in general it need not exist! Therefore (by 2d7) the notion "random Borel set" is not equivalent to "random element of $(\mathcal{B}(\mathbb{R}), \mathcal{C})$", no matter which σ-algebra $\mathcal{C}$ on $\mathcal{B}(\mathbb{R})$ is used.

2d9 Core exercise. If $\sigma(S)$ exists then $\{x : S(x) \ni y\} \in \sigma(S)$ for all $y \in \mathbb{R}$. Prove it.

2d10 Extra exercise. Let f, g and S be as in 2d5, and $f(\cdot) < g(\cdot)$ a.s. Prove that $\sigma(S)$ exists and is equal to $\sigma(f, g)$.

Here is a counterpart of 2c6 for random Borel sets.

2d11 Core exercise. If S is $\mathcal{A}_1$-measurable and $\mathcal{A}_2 \subset \mathcal{A}$ is a sub-σ-algebra such that $\mathcal{A}_1/\overset{\mu}{\sim} \subset \mathcal{A}_2/\overset{\mu}{\sim}$ (that is, each set of $\mathcal{A}_1$ is equivalent to some set of $\mathcal{A}_2$; note that $\mathcal{A}_2$ need not contain all null sets, in contrast to $\mathcal{A}_1$) then there exists $S_2 \in \mathcal{A}_2 \times \mathcal{B}(\mathbb{R})$ such that $S(x) = S_2(x)$ for almost all x.

Prove it.

2d12 Example. *"Unordered infinite sample"* Let $(X, \mathcal{A}, \mu) = (0,1)^\infty$ be the product of countably many copies of $(0,1)$ with Lebesgue measure, and

$$S(x_1, x_2, \dots) = \{x_1, x_2, \dots\} \quad \text{for } x_1, x_2, \dots \in (0,1)\,;$$

this is a special case of 2d5(b). We'll see that $\sigma(S)$ does not exist. To this end we introduce transformations $T_n : X \to X$ that swap $x_1, \dots, x_n$ and $x_{n+1}, \dots, x_{2n}$:

$$T_n(x_1, x_2, \dots) = (x_{n+1}, \dots, x_{2n},\, x_1, \dots, x_n,\, x_{2n+1}, x_{2n+2}, \dots)\,;$$

these transformations are measure preserving, that is, $\mu(T_n^{-1}(A)) = \mu(A)$ for $A \in \mathcal{A}$. We consider the σ-algebra $\mathcal{B}_n \subset \mathcal{A}_\mu$ of all T_n-invariant measurable sets:

$$\mathcal{B}_n = \{A \in \mathcal{A}_\mu : T_n(A) \overset{\mu}{\sim} A\}\,;$$

clearly, $\mathcal{B}_n$ contains all null sets.

The function $x \mapsto \max(x_1, \dots, x_{2n})$ is $\mathcal{B}_n$-measurable (since it is $\mathcal{A}$-measurable and T_n-invariant). The same holds for so-called order statistics $x_{(1)}^{(2n)}, \dots, x_{(2n)}^{(2n)}$ defined by

$$x_{(1)}^{(2n)} \le \dots \le x_{(2n)}^{(2n)} \quad \text{and} \quad \{x_{(1)}^{(2n)}, \dots, x_{(2n)}^{(2n)}\} = \{x_1, \dots, x_{2n}\}\,.$$

Thus, the random Borel set $x \mapsto \{x_1, \dots, x_{2n}\}$ is $\mathcal{B}_n$-measurable. Also $\{x_{2n+1}, x_{2n+2}, \dots\}$ is $\mathcal{B}_n$-measurable (since functions $x \mapsto x_{2n+k}$ are). We see that S is $\mathcal{B}_n$-measurable for all n. Therefore $\sigma(S) \subset \cap_n \mathcal{B}_n$ if $\sigma(S)$ exists. It remains to check that $\cap_n \mathcal{B}_n$ is the trivial σ-algebra[1] (only null sets and full measure sets), see below, and apply 2d11 to $\mathcal{A}_2 = \{\emptyset, X\}$.

We need some useful general facts.

Let $(X, \mathcal{A}, \mu)$ be a probability space and $A, A_1, A_2, \dots \in \mathcal{A}/\overset{\mu}{\sim}$. What about convergence, $A_n \to A$ as $n \to \infty$? It appears, we have two different, but closely related modes of convergence. One is convergence almost sure:

$$\mathbf{1}_{A_n}(\cdot) \to \mathbf{1}_A(\cdot) \text{ a.s.}$$

[1]Basically, the Hewitt-Savage zero-one law.

The other is topological convergence:

$$\operatorname{dist}(A_n, A) \to 0$$

where $\operatorname{dist}(A_n, A) = \mu(A_n \triangle A)$ (recall 2a6).

For monotone sequences these modes coincide (think, why).

2d13 Lemma. Almost sure convergence implies topological convergence.

Proof.

$$\operatorname{dist}(A_n, A) = \mu(A_n \triangle A) = \int \mathbf{1}_{A_n \triangle A} \, \mathrm{d}\mu \to 0$$

by the bounded convergence theorem, since $\mathbf{1}_{A_n \triangle A}(x) = 0$ if and only if $\mathbf{1}_{A_n}(x) = \mathbf{1}_A(x)$. $\square$

2d14 Lemma. If $\sum_n \operatorname{dist}(A_n, A) < \infty$ then $\mathbf{1}_{A_n} \to \mathbf{1}_A$ a.s.[1]

Proof. $\infty > \sum_n \operatorname{dist}(A_n, A) = \sum_n \int \mathbf{1}_{A_n \triangle A} \, \mathrm{d}\mu = \int \sum_n \mathbf{1}_{A_n \triangle A} \, \mathrm{d}\mu$ by the monotone convergence theorem; therefore $\sum_n \mathbf{1}_{A_n \triangle A} < \infty$ a.s. $\square$

But if $\sum_n \varepsilon_n = \infty$ then there exist intervals $A_n \subset [0, 1]$ of length ε_n such that $\limsup_n \mathbf{1}_{A_n} = 1$ everywhere (think, why).

For arbitrary $\mathcal{E} \subset \mathcal{A}/\overset{\mu}{\sim}$ its *topological closure* is

$$\mathcal{E}_{\mathrm{t}} = \{ A \in \mathcal{A}/\overset{\mu}{\sim} : \inf_{E \in \mathcal{E}} \operatorname{dist}(A, E) = 0 \} .$$

2d15 Lemma. $\mathcal{E}_{\mathrm{t}} \subset \mathcal{E}_{\delta\sigma} \cap \mathcal{E}_{\sigma\delta}$ for all $\mathcal{E} \subset \mathcal{A}/\overset{\mu}{\sim}$.

Proof. Given $A \in \mathcal{E}_{\mathrm{t}}$, we choose $E_n \in \mathcal{E}$ such that $\sum_n \operatorname{dist}(E_n, A) < \infty$. By 2d14, $\mathbf{1}_{E_n}(\cdot) \to \mathbf{1}_A(\cdot)$ a.s., therefore $\mathbf{1}_A = \limsup_n \mathbf{1}_{E_n} = \inf_n \sup_k \mathbf{1}_{E_{n+k}}$ a.s.; thus, $A = \cap_n \cup_k E_{n+k} \in \mathcal{E}_{\sigma\delta}$. Similarly (using lim inf), $A \in \mathcal{E}_{\delta\sigma}$. $\square$

2d16 Core exercise. Let $\mathcal{E} \subset \mathcal{A}/\overset{\mu}{\sim}$ be an algebra. Prove that

(a) $\mathcal{E}_{\mathrm{t}}$ is an algebra;

(b) moreover, $\mathcal{E}_{\mathrm{t}}$ is a σ-algebra;

(c) moreover, $\mathcal{E}_{\mathrm{t}} = \sigma(\mathcal{E})$.

We return to Example 2d12.

Treating the coordinates $x_1, x_2, \dots$ of $x \in (0, 1)^\infty$ as random variables we have $\sigma(x_1, x_2, \dots) = \mathcal{A}_\mu$. Applying 2d16 to the algebra $\mathcal{E} = \cup_n \sigma(x_1, \dots, x_n)$ we get

$$\inf_{E \in \mathcal{E}} \operatorname{dist}(A, E) = 0$$

[1] This is basically the first Borel-Cantelli lemma.

for all $A \in \mathcal{A}_\mu$. This is instructive: functions of finitely many variables can approximate every function of infinitely many variables. Though, we need only indicator functions.

Given $B \in \cap_n \mathcal{B}_n$, we take $A_n \in \sigma(x_1, \dots, x_n)$ such that $\operatorname{dist}(A_n, B) \to 0$. We note that $\operatorname{dist}(T_n(A_n), B) = \operatorname{dist}(T_n(A_n), T_n(B)) = \operatorname{dist}(A_n, B)$ (since $T_n(B) \overset{\mu}{\sim} B$). By 2a9(a), $\operatorname{dist}(A_n \cap T_n(A_n), B) = \operatorname{dist}(A_n \cap T_n(A_n), B \cap B) \le \operatorname{dist}(A_n, B) + \operatorname{dist}(T_n(A_n), B) \to 0$. Therefore $\mu(A_n \cap T_n(A_n)) \to \mu(B)$. On the other hand, $\mu(A_n \cap T_n(A_n)) = \mu(A_n)\mu(T_n(A_n)) = (\mu(A_n))^2 \to (\mu(B))^2$ (since $T_n(A_n) \in \sigma(x_{n+1}, \dots, x_{2n})$). We get $\mu(B) = (\mu(B))^2$, which means that $\mu(B)$ is either zero or one!

We see that the random Borel set of Example 2d12 fails to generate a σ-algebra. And therefore random Borel sets in general cannot be treated as random elements of the Borel σ-algebra.

In particular, we cannot reduce an arbitrary probability space $(X, \mathcal{A}, \mu)$ to the Cantor set via 2b5. But we can do it in another way.

Let S be a random Borel set on $(X, \mathcal{A}, \mu)$. According to 2d3 we may treat S as a $(\mathcal{A} \times \mathcal{B}(\mathbb{R}))$-measurable subset of $X \times \mathbb{R}$. By 1d41, $S \in \mathcal{A}_1 \times \mathcal{B}(\mathbb{R})$ for some countably generated $\mathcal{A}_1 \subset \mathcal{A}$. By 1d38, $\mathcal{A}_1 = \sigma(\varphi)$ for some $\varphi : X \to \{0,1\}^\infty$. The map

$$X \times \mathbb{R} \ni (x, y) \mapsto \big(\varphi(x), y\big) \in \{0,1\}^\infty \times \mathbb{R}$$

generates a σ-algebra that contains both $\{A \times \mathbb{R} : A \in \mathcal{A}_1\}$ and $\{X \times B : B \in \mathcal{B}(\mathbb{R})\}$, therefore it contains $\mathcal{A}_1 \times \mathcal{B}(\mathbb{R})$.[1] We take a Borel set $S_1 \subset \{0,1\}^\infty \times \mathbb{R}$ such that $S = \{(x, y) : \big(\varphi(x), y\big) \in S_1\}$. Similarly to Sect. 2b we introduce the image measure ν on $\{0,1\}^\infty$ (the distribution of φ), treat $\big(\{0,1\}^\infty, \mathcal{B}(\{0,1\}^\infty), \nu\big)$ as another probability space $(Y, \mathcal{B}, \nu)$, and S_1 as a random Borel set on $(Y, \mathcal{B}, \nu)$. We get

$$S(x) = \{y : (x, y) \in S\} = \{y : \big(\varphi(x), y\big) \in S_1\} = S_1(\varphi(x)) \,,$$

which is the counterpart of 2b5 for random Borel sets, formulated below.

2d17 Proposition. For every random Borel set S on $(X, \mathcal{A}, \mu)$ there exists a random Borel set S_1 on $(Y, \mathcal{B}, \nu)$ such that

$$S(\cdot) = S_1(\varphi(\cdot)) \quad \text{a.s.}$$

2d18 Remark. Why only random Borel subsets of $\mathbb{R}$? For an arbitrary measurable space $(Y, \mathcal{B})$, a random measurable subset of Y is defined similarly to 2d2, 2d3. All general statements (and their proofs) of Sect. 2d (including

[1] Moreover, it is equal to $\mathcal{A}_1 \times \mathcal{B}(\mathbb{R})$, but we do not need it now.

2d11 and 2d17), stated for the special case $(Y, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, generalize immediately to the general case. However, some items (namely, 2d4(b), 2d5, 2d6, 2d10, 2d12) use the structure of $\mathbb{R}$ (at least in proofs).

2d19 Core exercise. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be probability spaces, $\varphi : X \to Y$ a map measurable from $(X, \mathcal{A})$ to $(Y, \mathcal{B})$ and measure preserving (that is, $\mu(\varphi^{-1}(B)) = \nu(B)$ for $B \in \mathcal{B}$). Then φ is measurable from $(X, \mathcal{A}_\mu)$ to $(Y, \mathcal{B}_\nu)$.

Prove it.

2d20 Core exercise. The following two conditions on a measurable space $(Y, \mathcal{B})$ are equivalent:

(a) for every probability space $(X, \mathcal{A}, \mu)$ and every random measurable subset S of Y, defined on $(X, \mathcal{A}, \mu)$, the set $\{x \in X : S(x) \neq \emptyset\}$ is μ-measurable;

(b) the same, but only for $X = \{0, 1\}^\infty$ and $\mathcal{A} = \mathcal{B}(\{0, 1\}^\infty)$ (and arbitrary μ).

Prove it.

Later we'll see that this condition is satisfied by all standard Borel spaces.

2d21 Extra exercise. Let $(X, \mathcal{A}, \mu)$ be a probability space such that $(X, \mathcal{A})$ is countably separated. Then for every (not just μ-measurable) set $Z \subset X$ there exists a measurable space $(Y, \mathcal{B})$ and a random measurable subset S of Y, defined on $(X, \mathcal{A}, \mu)$, such that $\{x \in X : S(x) \neq \emptyset\} = Z$.

Prove it. (In fact, you do not need more than one point in each $S(x)$.)

Hints to exercises

2a4: (a) $B = \cup_n B_n$.

2a5: (c) $\mu(B_n) \le \mu^*(A_n) + 2^{-n}\varepsilon$.

2a8: (c) use (2a7).

2a9: use (2a7).

2a10: (b) use (2a9).

2a11: consider the union of these equivalence classes.

2a16: (a) $\mathcal{F} \subset \text{Sandwich}(\mathcal{F})$ since $\sim\mathcal{F} \subset \mathcal{F}_\sigma$; (b) using (a) prove that null sets belong to Sandwich($\mathcal{F}$).

2a17: $\mu(B_n) \ge \mu(A_n) - 2^{-n}\varepsilon$.

2a18: $\mu(B_1 \cup \dots \cup B_n) \uparrow \mu(B_1 \cup B_2 \cup \dots)$.

2a20: use 2a17 and 2a19.

2a21: use 2a15.

2b11: (b) $\{0,1\}^\infty \times \{0,1\}^\infty$ is isomorphic to $\{0,1\}^\infty$.

2c6: Similarly to the proof of 2b5, first prove it for the special case of the Cantor set Y.

2c8: (a) $\langle x, y\rangle = \lim_n (x_1 y_1 + \dots + x_n y_n)$;
(b) $\|x\| = \sup\{\langle x, y\rangle : \|y\| \le 1\}$; take a dense countable set of y;
(c) use (b).

2c9: $f \mapsto f(q_2) - f(q_1)$ is measurable.

2c10: use 2c9.

2c11: increasing functions need not be continuous.

2d1: approximate f by $f_n : X \to \{0, \frac{1}{n}, \dots, \frac{n-1}{n}\}$; condition (b) is not needed.

2d4: take RBS according to one definition and check that it satisfies the other.

2d5: approximate the given functions similarly to the hint to 2d1.

2d7: use 2a11.

2d9: $x \mapsto (x, y)$ is measurable from $(X, \mathcal{A})$ to $(X, \mathcal{A}) \times (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

2d11: this is simpler than 2c6; you do not need standardness of $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Just consider all S that have the needed property and check that they are a σ-algebra.

2d16: (a) use 2a9; (b) use monotone convergence and (a); (c) use 2d15 and (b).

2d20: use 2d17 and 2d19.

Index