Foundations of Probability

Axioms, σ-Algebras, Probability Spaces,
Borel Sets, Cardinality, and Non-measurable Sets

Graduate Random Processes
Department of Electrical & Computer Engineering
University of Illinois Urbana–Champaign

Lectures 1–2

Abstract

These notes accompany the first two lectures of the course. We begin with the *axiomatic method* itself—what an axiom system is and why mathematics is built on one—using Euclid's geometry and the Peano axioms as touchstones. We then develop the Kolmogorov foundations of probability: measurable spaces, σ-algebras and the σ-algebra generated by a class of events, the axioms of a probability measure, and the elementary properties that follow from them. Finally we confront the reason all of this machinery is *necessary*. We define the Borel σ-algebra, prove (via Cantor's theory of cardinality) that it is strictly smaller than the power set, and construct a Vitali set to show that one genuinely *cannot* assign a consistent "length" to every subset of the line. The moral: the domain of a probability measure must be a σ-algebra, not the power set.

Contents

Lecture 1: The Axiomatic Method and the Kolmogorov Axioms

1 Why axioms? The axiomatic method

Before we axiomatize probability, it is worth asking what an *axiom system* is and why essentially all of modern mathematics is organized around one.

An axiomatic theory begins with a small stock of *primitive* (undefined) notions and a list of *axioms*: statements about those notions that are accepted *without proof*. Every other statement of the theory—each *theorem*—must then be obtained from the axioms by the rules of logical deduction. The axioms are the load-bearing assumptions; the theorems are everything the assumptions force to be true.

Why insist on this discipline?

(i) **It makes assumptions explicit.** Everything we rely on is written down. Nothing is smuggled in from "intuition."

(ii) **It prevents circularity and infinite regress.** One cannot define every term or prove every statement; some starting points must be taken as given, or definitions and proofs would never terminate.

(iii) **It is where intuition fails.** For finite or "tame" objects, intuition is a fine guide. For infinite and uncountable objects—the continuum, in particular—intuition is unreliable and often flatly wrong. An axiom system lets us reason safely where our pictures break down. Probability on an uncountable sample space is exactly such a place.

Two classical examples

Example 1.1 (Euclid's geometry)**.** The oldest axiomatic theory. Euclid's *Elements* (c. 300 BCE) derives plane geometry from five postulates:

P1. Any two points can be joined by a straight segment.

P2. Any segment can be extended indefinitely to a straight line.

P3. Given a segment, a circle can be drawn with it as radius and one endpoint as centre.

P4. All right angles are equal.

P5. *(Parallel postulate.)* If a line crossing two lines makes the interior angles on one side sum to less than two right angles, the two lines meet on that side.

The fifth postulate is famously less "self-evident" than the others. For two millennia mathematicians tried to *deduce* it from P1–P4 and failed. In the nineteenth century Gauss, Bolyai and Lobachevsky showed why: replacing P5 with an alternative yields perfectly consistent *non-Euclidean* geometries (hyperbolic, elliptic). This is the deepest lesson of the axiomatic method—the choice of axioms is a genuine choice, and different consistent choices describe different worlds.

Example 1.2 (The Peano axioms for the natural numbers)**.** The natural numbers $\mathbb{N}$ are characterized by a primitive constant 0, a primitive *successor* operation S, and the axioms:

PA1. $0 \in \mathbb{N}$.

PA2. For every $n \in \mathbb{N}$, its successor $S(n) \in \mathbb{N}$.

PA3. There is no n with $S(n) = 0$ (so 0 starts the chain).

PA4. S is injective: if $S(m) = S(n)$ then $m = n$.

PA5. *(Induction.)* If a set A satisfies $0 \in A$ and $\big(n \in A \Rightarrow S(n) \in A\big)$, then $A = \mathbb{N}$.

Everything about the natural numbers—addition, multiplication, order, mathematical induction as a proof technique—is built on these five statements. Axiom PA5 is the engine behind *proof by induction*, which we will use repeatedly.

Remark 1.3 (Why probability needs this)*.* Naively one might hope to define, for *every* subset $A \subseteq [0,1]$, a number $\mathbb{P}(A) \in [0,1]$ measuring "how likely" a uniform random point lands in A, in a way that respects lengths and is additive. By the end of Lecture 2 we will *prove* that no such assignment on *all* subsets exists. This is not a technicality to be waved away—it is the reason probability is founded on measure theory, and the reason the events of a probability space form a σ-algebra rather than the full power set. We now build that foundation.

2 Measurable spaces and σ-algebras

Fix a nonempty set Ω, the *sample space*, whose elements $\omega \in \Omega$ are the possible outcomes of an experiment. An *event* is a subset $A \subseteq \Omega$; the event A "occurs" when the realized outcome ω lies in A. We will *not* in general take every subset of Ω to be an event. The admissible events form a distinguished collection with enough closure properties to support a consistent calculus of probabilities.

Definition 2.1 (σ-algebra)**.** A collection $\mathcal{F} \subseteq 2^{\Omega}$ of subsets of Ω is a *σ-algebra* (or *σ-field*) on Ω if

(F1) $\Omega \in \mathcal{F}$; *(the whole space is an event)*

(F2) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$; *(closed under complementation)*

(F3) $A_1, A_2, \dots \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$. *(closed under countable unions)*

The pair $(\Omega, \mathcal{F})$ is called a *measurable space*, and the members of $\mathcal{F}$ are the *measurable sets* or *events*.

The three axioms are deliberately spare. Everything one wants closure under follows.

Proposition 2.2 (Elementary closure properties)**.** *Let $\mathcal{F}$ be a σ-algebra on Ω. Then:*

(a) $\emptyset \in \mathcal{F}$;

(b) $\mathcal{F}$ is closed under countable intersections: $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcap_i A_i \in \mathcal{F}$;

(c) $\mathcal{F}$ is closed under finite unions and finite intersections;

(d) $\mathcal{F}$ is closed under set differences $A \setminus B$ and symmetric differences $A \triangle B$.

Proof. (a) By (F1) $\Omega \in \mathcal{F}$, so by (F2) $\emptyset = \Omega^c \in \mathcal{F}$.

(b) Let $A_1, A_2, \cdots \in \mathcal{F}$. By (F2) each $A_i^c \in \mathcal{F}$, so by (F3) $\bigcup_i A_i^c \in \mathcal{F}$, and by (F2) again its complement is in $\mathcal{F}$. By De Morgan's law,

$$\Big(\bigcup_i A_i^c\Big)^c = \bigcap_i A_i \in \mathcal{F}.$$

(c) Given finitely many $A_1, \ldots, A_n \in \mathcal{F}$, pad the sequence with $A_{n+1} = A_{n+2} = \cdots = \emptyset \in \mathcal{F}$ and apply (F3); this gives the finite union. Finite intersections follow from (b) analogously.

(d) $A \setminus B = A \cap B^c \in \mathcal{F}$ by (F2) and part (c). Then $A \triangle B = (A \setminus B) \cup (B \setminus A) \in \mathcal{F}$. □

Remark 2.3 (σ-algebra vs. algebra). If we require closure only under *finite* unions (in place of (F3)) we obtain an *algebra* (or *field*). The prefix "σ" signals the strengthening to *countable* operations. This strengthening is exactly what lets us pass to limits—to speak of $\{X_n \text{ converges}\}$, of lim sup and lim inf of events—which is indispensable in a theory of random *processes*.

Examples of σ-algebras

Example 2.4 (The two extremes)**.** For any Ω:

- the *trivial* σ-algebra $\mathcal{F} = \{\emptyset, \Omega\}$ is the smallest σ-algebra on Ω;
- the *power set* $\mathcal{F} = 2^\Omega$ (all subsets) is the largest.

Every σ-algebra on Ω lies between these two.

Example 2.5 (Generated by a single event)**.** Fix $A \subseteq \Omega$ with $A \neq \emptyset, \Omega$. The smallest σ-algebra containing A is

$$\mathcal{F} = \{\emptyset,\ A,\ A^c,\ \Omega\}.$$

One checks directly that this four-element collection satisfies (F1)–(F3), and that dropping any of its members destroys some closure property. Interpretation: this is the information "did A occur or not?"

Example 2.6 (A die)**.** Let $\Omega = \{1, 2, 3, 4, 5, 6\}$. The power set 2^Ω has $2^6 = 64$ events and is a perfectly good σ-algebra (finite sample spaces cause no trouble). A *coarser* choice records only parity:

$$\mathcal{F} = \big\{\emptyset,\ \{1, 3, 5\},\ \{2, 4, 6\},\ \Omega\big\},$$

the events observable if we are told only whether the roll is odd or even.

3 The σ-algebra generated by a class of events

Often we have in mind a family $\mathcal{C}$ of "basic" events we insist on being able to measure—say all intervals, or all cylinders of a process—and we want the *smallest* σ-algebra containing them. That such a smallest one exists is not obvious; it rests on the following stability lemma.

Proposition 3.1 (Intersections of σ-algebras are σ-algebras)**.** *Let $\{\mathcal{F}_\alpha\}_{\alpha \in I}$ be an arbitrary (possibly uncountable) nonempty family of σ-algebras on Ω. Then*

$$\mathcal{F} \;=\; \bigcap_{\alpha \in I} \mathcal{F}_\alpha \;=\; \{A \subseteq \Omega : A \in \mathcal{F}_\alpha \text{ for every } \alpha \in I\}$$

is a σ-algebra on Ω.

Proof. We verify (F1)–(F3). *(F1)* Each $\mathcal{F}_\alpha$ contains Ω, hence so does their intersection. *(F2)* If $A \in \mathcal{F}$ then $A \in \mathcal{F}_\alpha$ for every α; each $\mathcal{F}_\alpha$ is closed under complementation, so $A^c \in \mathcal{F}_\alpha$ for every α, whence $A^c \in \mathcal{F}$. *(F3)* If $A_1, A_2, \cdots \in \mathcal{F}$, then for each fixed α all the A_i lie in $\mathcal{F}_\alpha$, so $\bigcup_i A_i \in \mathcal{F}_\alpha$; as this holds for every α, $\bigcup_i A_i \in \mathcal{F}$. □

Definition 3.2 (Generated σ-algebra)**.** Let $\mathcal{C} \subseteq 2^\Omega$ be any collection of subsets. The *σ-algebra generated by $\mathcal{C}$*, written $\sigma(\mathcal{C})$, is

$$\sigma(\mathcal{C}) \;=\; \bigcap \big\{\, \mathcal{G} : \mathcal{G} \text{ is a } \sigma\text{-algebra on } \Omega \text{ and } \mathcal{C} \subseteq \mathcal{G} \,\big\}.$$

This definition is not vacuous: the family being intersected is nonempty, since the power set 2^Ω is a σ-algebra containing $\mathcal{C}$. By Proposition 3.1 the intersection is again a σ-algebra, and it plainly contains $\mathcal{C}$. The following proposition records that $\sigma(\mathcal{C})$ deserves the name "smallest."

Proposition 3.3 (Minimality)**.** *$\sigma(\mathcal{C})$ is the smallest σ-algebra containing $\mathcal{C}$; that is, $\mathcal{C} \subseteq \sigma(\mathcal{C})$, and if $\mathcal{G}$ is any σ-algebra with $\mathcal{C} \subseteq \mathcal{G}$, then $\sigma(\mathcal{C}) \subseteq \mathcal{G}$.*

Proof. That $\mathcal{C} \subseteq \sigma(\mathcal{C})$ is immediate: every σ-algebra in the defining intersection contains $\mathcal{C}$, so their intersection does too. For minimality, let $\mathcal{G} \supseteq \mathcal{C}$ be a σ-algebra. Then $\mathcal{G}$ is one of the sets in the intersection of Definition 3.2, and an intersection is contained in each of its members, so $\sigma(\mathcal{C}) \subseteq \mathcal{G}$. □

Example 3.4. $\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$, recovering the earlier example. More generally, if $\{A_1, \ldots, A_n\}$ is a finite partition of Ω, then $\sigma(\{A_1, \ldots, A_n\})$ consists of $\emptyset$ together with all $2^n - 1$ nonempty unions of the blocks A_i—a total of 2^n events.

Remark 3.5. In practice $\sigma(\mathcal{C})$ is defined "from above" (as an intersection) precisely because there is usually no way to describe its members "from below" by an explicit constructive formula. We will see this vividly with the Borel sets in Lecture 2.

4 The axioms of probability

We can now state Kolmogorov's axioms (1933), which put probability on the same rigorous footing as Euclidean geometry or arithmetic.

Definition 4.1 (Probability measure; probability space)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A *probability measure* on it is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfying:

(P1) **Non-negativity:** $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$.

(P2) **Normalization:** $\mathbb{P}(\Omega) = 1$.

(P3) **Countable (σ-)additivity:** for any sequence $A_1, A_2, \cdots \in \mathcal{F}$ of *pairwise disjoint* events (i.e. $A_i \cap A_j = \emptyset$ for $i \neq j$),

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) \;=\; \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*.

Three comments on the choices made here.

- The domain of $\mathbb{P}$ is $\mathcal{F}$, *not* 2^Ω. This is the crucial point that Lecture 2 will justify: on an uncountable Ω one cannot in general extend a reasonable $\mathbb{P}$ to all of 2^Ω.
- Axiom (P3) is stated for *countably* many disjoint sets. Merely finite additivity would be too weak to support limiting arguments; full countable additivity is what makes the theory analytically powerful (and is equivalent, given finite additivity, to a continuity property proved below).
- The series $\sum_i \mathbb{P}(A_i)$ has non-negative terms and is bounded by 1, so it converges; its value does not depend on the order of summation.

5 Properties deduced from the axioms

Everything in elementary probability follows from (P1)–(P3). We derive the standard toolkit.

Proposition 5.1 (Basic identities and inequalities)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A, B \in \mathcal{F}$. Then:*

(1) $\mathbb{P}(\emptyset) = 0$.

(2) (Finite additivity.) *If $A_1, \dots, A_n \in \mathcal{F}$ are pairwise disjoint, then $\mathbb{P}\big(\bigcup_{i=1}^n A_i\big) = \sum_{i=1}^n \mathbb{P}(A_i)$.*

(3) (Complement.) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(4) (Monotonicity.) *If $A \subseteq B$ then $\mathbb{P}(A) \le \mathbb{P}(B)$, and moreover $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.*

(5) (Range.) $0 \le \mathbb{P}(A) \le 1$.

(6) (Inclusion–exclusion.) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

(7) (Union bound / Boole's inequality.) *For any $A_1, A_2, \dots \in \mathcal{F}$,*

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) \le \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Proof. (1) Apply (P3) to the disjoint sequence $A_1 = A_2 = \cdots = \emptyset$. It gives $\mathbb{P}(\emptyset) = \sum_{i=1}^\infty \mathbb{P}(\emptyset)$. Writing $c := \mathbb{P}(\emptyset) \in [0,1]$, this reads $c = \sum_{i=1}^\infty c$. If $c > 0$ the right side is $+\infty$, impossible since $c \le 1$; hence $c = 0$.

(2) Given disjoint $A_1, \dots, A_n$, extend by $A_{n+1} = A_{n+2} = \cdots = \emptyset$. The extended sequence is pairwise disjoint, so (P3) and part (1) give $\mathbb{P}\big(\bigcup_{i=1}^n A_i\big) = \sum_{i=1}^\infty \mathbb{P}(A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$.

(3) A and A^c are disjoint with union Ω. By part (2) and (P2), $\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(\Omega) = 1$.

(4) If $A \subseteq B$, then $B = A \dot\cup (B \setminus A)$ is a disjoint union of members of $\mathcal{F}$. By part (2), $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$. Since $\mathbb{P}(B \setminus A) \ge 0$ by (P1), we get $\mathbb{P}(B) \ge \mathbb{P}(A)$; rearranging gives the stated difference formula.

(5) $\mathbb{P}(A) \ge 0$ is (P1). Since $A \subseteq \Omega$, monotonicity (4) and (P2) give $\mathbb{P}(A) \le \mathbb{P}(\Omega) = 1$.

(6) Decompose $A \cup B$ and B into disjoint pieces:

$$A \cup B = A \dot\cup (B \cap A^c), \qquad B = (B \cap A) \dot\cup (B \cap A^c).$$

By finite additivity, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$ and $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \cap A^c)$. Solving the second for $\mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$ and substituting yields the claim.

(7) Disjointification. Set $B_1 = A_1$ and, for $n \geq 2$,

$$B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i \in \mathcal{F}.$$

The B_n are pairwise disjoint, $B_n \subseteq A_n$, and $\bigcup_n B_n = \bigcup_n A_n$. Hence by (P3) and monotonicity,

$$\mathbb{P}\Big(\bigcup_n A_n\Big) = \mathbb{P}\Big(\bigcup_n B_n\Big) = \sum_n \mathbb{P}(B_n) \leq \sum_n \mathbb{P}(A_n). \qquad \square$$

The next result—continuity of probability along monotone sequences of events—is the payoff of insisting on *countable* additivity, and is the workhorse behind limit theorems.

Proposition 5.2 (Continuity from below and above)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.*

(a) (From below.) *If $A_1 \subseteq A_2 \subseteq \cdots$ and $A = \bigcup_n A_n$, then $\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n)$.*

(b) (From above.) *If $A_1 \supseteq A_2 \supseteq \cdots$ and $A = \bigcap_n A_n$, then $\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n)$.*

Proof. (a) Put $B_1 = A_1$ and $B_n = A_n \setminus A_{n-1}$ for $n \geq 2$. Because the A_n increase, the B_n are pairwise disjoint, $\bigcup_{k=1}^n B_k = A_n$, and $\bigcup_k B_k = A$. By (P3),

$$\mathbb{P}(A) = \sum_{k=1}^{\infty} \mathbb{P}(B_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(B_k) = \lim_{n\to\infty} \mathbb{P}(A_n),$$

the middle equality being the definition of an infinite sum as the limit of its partial sums, and the last using finite additivity.

(b) The complements increase: $A_1^c \subseteq A_2^c \subseteq \cdots$ with union $\bigcup_n A_n^c = \big(\bigcap_n A_n\big)^c = A^c$. Applying part (a) and the complement rule (3),

$$1 - \mathbb{P}(A) = \mathbb{P}(A^c) = \lim_n \mathbb{P}(A_n^c) = \lim_n \big(1 - \mathbb{P}(A_n)\big) = 1 - \lim_n \mathbb{P}(A_n),$$

so $\mathbb{P}(A) = \lim_n \mathbb{P}(A_n)$. (Finiteness of $\mathbb{P}$ is what makes the step to complements legitimate here.) $\square$

Example 5.3 (A concrete probability space)**.** Let $\Omega = [0,1]$, let $\mathcal{F} = \mathcal{B}([0,1])$ be the Borel σ-algebra (defined next lecture), and let $\mathbb{P}$ be *Lebesgue measure*, so that $\mathbb{P}\big((a,b]\big) = b - a$ for $0 \leq a \leq b \leq 1$. This is the uniform distribution on $[0,1]$. As a sanity check on continuity from above, the nested intervals $A_n = (0, 1/n]$ satisfy $\bigcap_n A_n = \emptyset$ and $\mathbb{P}(A_n) = 1/n \to 0 = \mathbb{P}(\emptyset)$, consistent with Proposition 5.2(b).

Lecture 2: Borel Sets, Cardinality, and Non-measurable Sets

We have taken the domain of a probability measure to be a σ-algebra $\mathcal{F}$, not the power set 2^Ω. On a countable Ω this restriction is harmless: one may always take $\mathcal{F} = 2^\Omega$. The point of this lecture is that on $\Omega = \mathbb{R}$ (or $[0,1]$) the restriction is *forced*. We introduce the canonical σ-algebra on the line, use Cantor's theory of cardinality to show it is strictly smaller than the power set, and finally exhibit a set that *cannot* be assigned a length at all.

6 The Borel σ-algebra

Definition 6.1 (Borel σ-algebra)**.** The *Borel σ-algebra* on $\mathbb{R}$ is the σ-algebra generated by the open sets:

$$\mathcal{B}(\mathbb{R}) \;=\; \sigma\big(\{\, U \subseteq \mathbb{R} : U \text{ open} \,\}\big).$$

Its members are called *Borel sets*. More generally, on any topological space one defines $\mathcal{B}(\cdot) = \sigma(\text{open sets})$.

By closure, $\mathcal{B}(\mathbb{R})$ contains far more than the open sets: every closed set (a complement of an open set), every countable intersection of open sets (G_δ), every countable union of closed sets (F_σ), and so on up a transfinite hierarchy. In particular every interval, every singleton $\{x\} = \bigcap_n (x - \frac{1}{n}, x + \frac{1}{n})$, and hence (by countable unions) every countable set such as $\mathbb{Q}$, is Borel. For all practical purposes—every set an engineer writes down—is Borel; producing a non-Borel set explicitly requires real effort.

It is convenient that $\mathcal{B}(\mathbb{R})$ has many equivalent descriptions.

Proposition 6.2 (Equivalent generators)**.** *Each of the following collections generates the same σ-algebra $\mathcal{B}(\mathbb{R})$:*

$$\mathcal{C}_1 = \{(a,b) : a < b\}, \quad \mathcal{C}_2 = \{(-\infty, a] : a \in \mathbb{R}\}, \quad \mathcal{C}_3 = \{(-\infty, a) : a \in \mathbb{R}\}.$$

Proof sketch. Every open subset of $\mathbb{R}$ is a countable union of open intervals with rational endpoints (each point of an open set sits in such a rational interval inside the set, and there are only countably many rational intervals). Hence the open sets lie in $\sigma(\mathcal{C}_1)$, giving $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{C}_1)$; the reverse inclusion is clear since open intervals are open. Thus $\sigma(\mathcal{C}_1) = \mathcal{B}(\mathbb{R})$.

For $\mathcal{C}_2$: half-lines $(-\infty, a]$ are closed, hence Borel, so $\sigma(\mathcal{C}_2) \subseteq \mathcal{B}(\mathbb{R})$. Conversely

$$(a,b] = (-\infty, b] \setminus (-\infty, a] \in \sigma(\mathcal{C}_2), \qquad (a,b) = \bigcup_{n=1}^{\infty} \Big(a,\; b - \tfrac{1}{n}\Big] \in \sigma(\mathcal{C}_2),$$

so all open intervals lie in $\sigma(\mathcal{C}_2)$, whence $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C}_1) \subseteq \sigma(\mathcal{C}_2)$. The argument for $\mathcal{C}_3$ is analogous, using $(-\infty, a] = \bigcap_n (-\infty, a + \frac{1}{n})$ and $(-\infty, a) = \bigcup_n (-\infty, a - \frac{1}{n}]$. □

The description via $\mathcal{C}_2$ is the one that connects to distribution functions: a real random variable X is exactly a function for which every event $\{X \le a\} = X^{-1}\big((-\infty, a]\big)$ is measurable.

The central question of this lecture: *is $\mathcal{B}(\mathbb{R})$ all of $2^{\mathbb{R}}$?* The answer is *no*, and to prove it cleanly we need Cantor's calculus of the infinite.

7 Interlude: Cantor's aleph numbers

Cantor discovered that "infinite" is not a single size but the bottom of an endless ladder of sizes. The organizing notion is a bijection.

Definition 7.1 (Equinumerosity and cardinality)**.** Two sets A, B have the *same cardinality*, written $|A| = |B|$, if there is a bijection $A \to B$. We write $|A| \le |B|$ if there is an injection $A \to B$. A set is *countable* if it is finite or has the same cardinality as $\mathbb{N}$; the cardinality of $\mathbb{N}$ is denoted $\aleph_0$ ("aleph-nought"), the smallest infinite cardinal.

The integers $\mathbb{Z}$ and the rationals $\mathbb{Q}$ are countable ($|\mathbb{Z}| = |\mathbb{Q}| = \aleph_0$), which already strains intuition—$\mathbb{Q}$ is dense in $\mathbb{R}$, yet no larger than $\mathbb{N}$. The real numbers are a different, strictly greater size.

Theorem 7.2 (Cantor's diagonal argument)**.** *The interval* $(0,1)$ *is uncountable. Consequently* $\mathbb{R}$ *is uncountable.*

Proof. Suppose, for contradiction, that $(0,1)$ were countable, and enumerate its elements as $x_1, x_2, x_3, \ldots$. Write each in a decimal expansion, choosing the non-terminating form when there is a choice (so, e.g., $0.5 = 0.4999\ldots$), to make expansions unique:

$$x_n = 0.\, d_{n1} d_{n2} d_{n3} \cdots, \qquad d_{nk} \in \{0, 1, \ldots, 9\}.$$

Define a new number $y = 0.\, e_1 e_2 e_3 \cdots$ by choosing each digit to differ from the diagonal one and to avoid 0 and 9:

$$e_k = \begin{cases} 5, & d_{kk} \neq 5, \\ 6, & d_{kk} = 5. \end{cases}$$

Then $y \in (0,1)$ (its digits lie in $\{5,6\}$, so no terminating-expansion ambiguity arises), yet for every n we have $e_n \neq d_{nn}$, so y differs from x_n in the n-th decimal place and thus $y \neq x_n$. Hence y is a member of $(0,1)$ absent from the enumeration—contradicting that the list was exhaustive. Therefore $(0,1)$ is uncountable. □

The cardinality of $\mathbb{R}$ is called the *cardinality of the continuum*, denoted $\mathfrak{c}$. A short argument via binary expansions shows $\mathfrak{c} = 2^{\aleph_0}$, the cardinality of the power set of $\mathbb{N}$. This is a special case of the next, completely general, theorem.

Theorem 7.3 (Cantor's theorem)**.** *For every set* A, $|A| < |2^A|$*; that is, there is an injection* $A \to 2^A$ *but no surjection* $A \to 2^A$*.*

Proof. The map $a \mapsto \{a\}$ is an injection $A \to 2^A$, so $|A| \leq |2^A|$. It remains to rule out any surjection. Let $f : A \to 2^A$ be arbitrary and consider the "diagonal" set

$$D = \{\, a \in A : a \notin f(a) \,\} \in 2^A.$$

Suppose $D = f(a_0)$ for some $a_0 \in A$. Ask whether $a_0 \in D$:

$$a_0 \in D \iff a_0 \notin f(a_0) \iff a_0 \notin D,$$

a contradiction. Hence D is not in the range of f, so f is not surjective. As f was arbitrary, no surjection $A \to 2^A$ exists, and therefore $|A| < |2^A|$. □

Iterating Cantor's theorem produces an unbounded tower of infinities

$$\aleph_0 = |\mathbb{N}| \;<\; 2^{\aleph_0} = |2^{\mathbb{N}}| = \mathfrak{c} \;<\; 2^{\mathfrak{c}} = |2^{\mathbb{R}}| \;<\; 2^{2^{\mathfrak{c}}} \;<\; \cdots.$$

The *aleph numbers* $\aleph_0 < \aleph_1 < \aleph_2 < \cdots$ enumerate the infinite cardinals in increasing order ($\aleph_1$ being the smallest uncountable cardinal, and so on). Where does $\mathfrak{c}$ sit in this list? Cantor conjectured $\mathfrak{c} = \aleph_1$—the *Continuum Hypothesis*. Remarkably, Gödel (1940) and Cohen (1963) showed this question is *independent* of the standard axioms (ZFC): it can neither be proved nor disproved from them. This is itself a striking vindication of the axiomatic method—whether $\mathfrak{c}$ equals $\aleph_1$ depends on which additional axioms one is willing to adopt.

We need only the coarse fact that $2^{\mathbb{R}}$ is strictly larger than $\mathfrak{c}$.

8 The Borel σ-algebra is *not* the power set

Theorem 8.1 (Cardinality of the Borel sets)**.** $|\mathcal{B}(\mathbb{R})| = \mathfrak{c} = 2^{\aleph_0}$.

Proof sketch. Lower bound. The map $a \mapsto (-\infty, a]$ injects $\mathbb{R}$ into $\mathcal{B}(\mathbb{R})$, so $|\mathcal{B}(\mathbb{R})| \geq \mathfrak{c}$.

Upper bound. One builds the Borel sets by transfinite recursion along the countable ordinals. Let Σ_1^0 be the open sets; there are exactly $\mathfrak{c}$ of them (each is a countable union of rational-endpoint intervals, and $\mathfrak{c}^{\aleph_0} = \mathfrak{c}$). Having formed the sets available before stage ξ, let stage ξ adjoin all *countable unions of complements* of earlier sets. Each stage therefore multiplies the count by at most $\mathfrak{c}^{\aleph_0} = \mathfrak{c}$, so if every earlier stage had at most $\mathfrak{c}$ sets, so does stage ξ. Every Borel set appears by some countable ordinal stage, and there are $\aleph_1 \leq \mathfrak{c}$ such stages. Hence

$$|\mathcal{B}(\mathbb{R})| \leq \aleph_1 \cdot \mathfrak{c} \leq \mathfrak{c} \cdot \mathfrak{c} = \mathfrak{c}.$$

Combining the two bounds gives $|\mathcal{B}(\mathbb{R})| = \mathfrak{c}$. □

Corollary 8.2 (Most subsets of $\mathbb{R}$ are not Borel)**.** $\mathcal{B}(\mathbb{R}) \subsetneq 2^{\mathbb{R}}$. *In fact the collection of non-Borel subsets of* $\mathbb{R}$ *has cardinality* $2^{\mathfrak{c}}$*, strictly greater than the cardinality* $\mathfrak{c}$ *of* $\mathcal{B}(\mathbb{R})$*.*

Proof. By Cantor's Theorem 7.3, $|2^{\mathbb{R}}| = 2^{\mathfrak{c}} > \mathfrak{c} = |\mathcal{B}(\mathbb{R})|$ by Theorem 8.1. A proper subset relation follows immediately: a set of size $\mathfrak{c}$ cannot equal one of size $2^{\mathfrak{c}}$. Since removing a set of size $\mathfrak{c}$ from a set of size $2^{\mathfrak{c}}$ leaves $2^{\mathfrak{c}}$ elements, the non-Borel sets number $2^{\mathfrak{c}}$. □

This is a purely cardinality-theoretic proof that non-Borel sets exist—indeed that they vastly outnumber the Borel sets—without exhibiting a single one. It explains why we could not simply take $\mathcal{F} = 2^{\Omega}$ and be done: the power set is enormously larger than any σ-algebra we can describe by generation from intervals. But cardinality alone does not yet show that taking 2^{Ω} would be *inconsistent* with wanting a length-like probability. For that we turn to an explicit construction.

9 Vitali sets: a set with no consistent length

We now prove the theorem promised in Lecture 1: there is *no* way to assign a length to every subset of the line while keeping the two properties any reasonable notion of "length" must have. The construction is due to Giuseppe Vitali (1905). It relies essentially on the *Axiom of Choice*.

Work on the interval $[0, 1)$ with addition *modulo* 1: for $x, y \in [0, 1)$ write $x \oplus y = (x + y) \bmod 1 \in [0, 1)$, and for a set $V \subseteq [0, 1)$ and $q \in [0, 1)$ write $V \oplus q = \{v \oplus q : v \in V\}$. Lebesgue measure on $[0, 1)$ is invariant under $\oplus q$ (rotating the circle $[0, 1)$ by q does not change lengths). We isolate the two properties we require of any candidate "length" $\mu : 2^{[0,1)} \to [0, \infty]$:

(M1) **Countable additivity** and $\mu([0, 1)) = 1$;

(M2) **Translation invariance:** $\mu(V \oplus q) = \mu(V)$ for all $V \subseteq [0, 1)$ and $q \in [0, 1)$.

Theorem 9.1 (Vitali, 1905)**.** *There is no function* $\mu : 2^{[0,1)} \to [0, \infty]$ *defined on* all *subsets of* $[0, 1)$ *satisfying both (M1) and (M2).*

Proof. Define a relation on $[0, 1)$ by

$$x \sim y \iff x - y \in \mathbb{Q}.$$

This is an equivalence relation (reflexive, symmetric, and transitive because $\mathbb{Q}$ is closed under negation and addition). It partitions $[0, 1)$ into equivalence classes; two points share a class exactly when they differ by a rational.

Choice of representatives. By the Axiom of Choice, form a set $V \subseteq [0,1)$ containing exactly one element from each equivalence class. This V is a *Vitali set*.

Enumerate the rationals in $[0,1)$ as $q_0, q_1, q_2, \ldots$ (there are countably many), and consider the translates

$$V_k := V \oplus q_k, \qquad k = 0, 1, 2, \ldots$$

We claim these are *pairwise disjoint* and *cover* $[0,1)$.

Disjointness. Suppose $z \in V_j \cap V_k$ for some $j \neq k$. Then $z = v \oplus q_j = v' \oplus q_k$ for some $v, v' \in V$. Hence $v - v' \equiv q_k - q_j \pmod 1$, so $v - v' \in \mathbb{Q}$, i.e. $v \sim v'$. But V contains only one representative per class, forcing $v = v'$ and therefore $q_j \equiv q_k \pmod 1$; as $q_j, q_k \in [0,1)$ this means $q_j = q_k$, contradicting $j \neq k$.

Covering. Let $z \in [0,1)$ be arbitrary. Its equivalence class has a representative $v \in V$, so $z - v \in \mathbb{Q}$; reducing modulo 1, there is a rational $q_k \in [0,1)$ with $z = v \oplus q_k$, i.e. $z \in V_k$. Thus $\bigcup_k V_k = [0,1)$.

Now suppose a μ as in the statement existed. Applying countable additivity (M1) to the disjoint cover, then translation invariance (M2),

$$1 = \mu\big([0,1)\big) = \mu\Big(\bigsqcup_{k=0}^{\infty} V_k\Big) = \sum_{k=0}^{\infty} \mu(V_k) = \sum_{k=0}^{\infty} \mu(V).$$

The final series has every term equal to the single number $\mu(V)$. There are only two cases, and both are impossible:

$$\mu(V) = 0 \;\Rightarrow\; \sum_k \mu(V) = 0 \neq 1, \qquad \mu(V) > 0 \;\Rightarrow\; \sum_k \mu(V) = +\infty \neq 1.$$

Either way we contradict $\mu([0,1)) = 1$. Hence no such μ exists. □

Remark 9.2 (Reading the theorem correctly). Theorem 9.1 does *not* say "$\mu(V)$ is some particular forbidden value." It says the very *existence* of a translation-invariant, countably additive length on *all* of $2^{[0,1)}$ is contradictory. The escape is not to weaken additivity or invariance—both are non-negotiable for a notion worthy of the name "probability" or "length"—but to *shrink the domain*. Lebesgue measure is defined on the Lebesgue σ-algebra (which contains $\mathcal{B}(\mathbb{R})$); the Vitali set V is simply not a member of it. This is the precise, unavoidable reason the events of a probability space form a σ-algebra rather than the power set.

Remark 9.3 (On the Axiom of Choice). The construction used the Axiom of Choice to pick one representative from each of uncountably many classes simultaneously. This is essential: Solovay (1970) showed that it is consistent with the other axioms of set theory (ZF, without Choice, together with a suitable large-cardinal assumption) that *every* subset of $\mathbb{R}$ is Lebesgue measurable. In other words, non-measurable sets cannot be constructed without some form of Choice. The Vitali set is thus a place where a foundational *axiom* has direct, tangible consequences for what probability can and cannot do—bringing us back to the theme with which Lecture 1 began.

Summary

- An *axiom system* fixes primitive notions and unproved starting assumptions; theorems are their logical consequences. Euclid and Peano are the classical templates. Probability needs this rigor because intuition fails on the uncountable.

- A *σ-algebra* $\mathcal{F}$ on Ω is closed under complements and countable unions (Def. 2.1); $(\Omega, \mathcal{F})$ is a *measurable space*. The σ-algebra $\sigma(\mathcal{C})$ *generated* by a class $\mathcal{C}$ is the smallest one containing it, obtained as an intersection (Def. 3.2).
- A *probability measure* satisfies non-negativity, normalization, and countable additivity (Def. 4.1). From these follow $\mathbb{P}(\emptyset) = 0$, the complement rule, monotonicity, inclusion–exclusion, the union bound, and continuity along monotone sequences (§5).
- The *Borel* σ-algebra $\mathcal{B}(\mathbb{R}) = \sigma(\text{open sets})$ contains every set one meets in practice, yet by a cardinality argument $|\mathcal{B}(\mathbb{R})| = \mathfrak{c} < 2^{\mathfrak{c}} = |2^{\mathbb{R}}|$, so $\mathcal{B}(\mathbb{R}) \subsetneq 2^{\mathbb{R}}$ (§8).
- A *Vitali set* shows there is no translation-invariant, countably additive length on all of $2^{[0,1)}$ (Thm. 9.1). Hence the domain of a probability measure *must* be a σ-algebra, not the power set.