1 Basics of measure theory

1.1 Introduction

A central theme of measure theory is the following question. How can we assign a (nonnegative) measure to subsets of some ground set Ω? In applications, the measure can have the meaning of size, content, mass, probability etc. We are talking of probability measure if the total measure of Ω is 1.

In your probability and statistics courses you mostly studied probability measures as distributions of random variables, with the focus on two types: 'discrete' with distribution supported by some finite or countable set or 'continuous' with some density. An elementary example of different type is a random variable $\min(E, a)$, where E is exponential r.v. and $a > 0$, with distribution having atom at a and positive density on $[0, a)$.

As a function of set the measure should be *additive*. If Ω is finite or countable this is pretty straightforward, as a nonzero value $\mu(\omega)$ [(1)] can be assigned to *every* $\omega \in \Omega$ and then $\mu(A)$ defined for any $A \subset \Omega$ using the summation formula $\mu(A) = \sum_{\omega \in A} \mu(A)$.

The problem becomes more involved if Ω is uncountable like $[0, 1], \mathbb{R}^k$ or the infinite 'coin-tossing space' $\{0, 1\}^\infty = \{0, 1\} \times \{0, 1\} \times \ldots$, or the space of continuous functions $C[0, 1]$, because it is not possible to assign measure in a reasonable way to every subset. That is to say, we should be careful about the domain of definition of a measure.

A fundamental example of measure is the Lebesgue measure generalising the geometric notions of length, area and volume. As a leading example today we shall consider the Lebesgue measure in one-dimension, that is the 'length' λ defined on certain subsets in $\mathbb{R}$. The length $\lambda(I)$ of any interval $I = [a, b], (a, b], (a, b), [a, b)$ is $\lambda(I) = b - a$. For union of disjoint intervals $I_1, \ldots, I_n$ the length is

$$\lambda\left(\bigcup_{k=1}^{n} I_k\right) = \sum_{k=1}^{n} \lambda(I_k),$$

which is an instance of the property called *finite additivity*. For infinite sequence of disjoint intervals $I_1, I_2, \ldots$ the length of the union is the sum of series,

$$\lambda\left(\bigcup_{k=1}^{\infty} I_k\right) = \sum_{k=1}^{\infty} \lambda(I_k)$$

(infinite if the series diverges), which is an instance of the property called *σ-additivity*.

Using the σ-additivity we can do measure calculations for more complex sets. What is the length of the set $\mathbb{Q}$ of rational numbers? The length of a point is $\lambda(\{x\}) = 0$, and $\mathbb{Q}$ is countable, so by σ-additivity $\lambda(\mathbb{Q}) = 0$. But then the length of the set of irrational numbers $[0, 1] \setminus \mathbb{Q}$ in $[0, 1]$ should be 1 since $\lambda([0, 1] \setminus \mathbb{Q}) + \lambda(\mathbb{Q}) = \lambda([0, 1]) = 1$.

Now let us find the length of the *Cantor set* $C \subset [0, 1]$. The Cantor set can be constructed step-by-step, at each stage obtaining some union of disjoint intervals C_k. Start with removing the middle third from $[0, 1]$, thus defining $C_1 := [0, 1/3] \cup [2/3, 1]$. Then remove the middle third from $[0, 1/3]$ and do the same with $[2/3, 1]$, thus defining C_2. By induction, C_{k+1} is obtained by removing the middle third from every interval in C_k. The Cantor set is defined as the infinite intersection $C = \cap_{k=1}^{\infty} C_k$. Note that, as in the above example, for $B \subset A$ we have $\lambda(B) \leq A$, because $A = B \cup (A \setminus B)$ is a disjoint union and $\lambda(A) = \lambda(B) + \lambda(B \setminus A)$. One can calculate the length of all removed intervals

$$\lambda([0, 1] \setminus C) = \frac{1}{3} + \frac{1}{3}\left(1 - \frac{1}{3}\right) + \frac{1}{3}\left(\frac{2}{3}\right)^2 + \cdots = \frac{1}{3} \cdot \frac{1}{1 - 2/3} = 1$$

[(1)] This is a shorthand notation for $\mu(\{\omega\})$ in case of one-point sets.

to see that $\lambda(C) = 1 - 1 = 0$. Another way to derive this is to show by induction that

$$\lambda(C_{k+1}) = \frac{2}{3}\lambda(C_k), \quad \text{hence} \;\; \lambda(C_k) = \left(\frac{2}{3}\right)^k,$$

and since $C \subset C_k$, we have

$$\lambda(C) \leq \lambda(C_k) = \left(\frac{2}{3}\right)^k, \quad k = 1, 2, \dots$$

and letting $k \to \infty$ yields $\lambda(C_k) \to 0$, so $\lambda(C) = 0$. The Cantor set is uncountable (has cardinality continuum, same as the cardinality of $[0, 1]$ or $\mathbb{R}$) and, as we have shown, has length 0.

How far can we go with ascribing the length to more complex sets $A \subset \mathbb{R}$? After the founder of measure theory Henri Lebesgue, the sets for which this can be done are called *Lebesgue measurable*, and the generalised length is called *the Lebesgue measure on* $\mathbb{R}$, to be discussed in the next section. Using the Axiom of Choice from the set theory it is possible to show existence of sets that are not Lebesque-mesaurable, but it is impossible to build them up from a system of intervals in some constructive manner. In probability theory we typically consider measures on a smaller class of Borel sets, which is rich enough for all purposes.

A *probability measure* on Ω is a measure with $\mu(\Omega) = 1$. Subsets of Ω to which probability is assigned are called events, and notation $\mathbb{P}(A)$ will be used for probability of $A \subset \Omega$. For instance, the Lebesgue measure on $[0, 1]^k$ is a probability measure, used to model a point chosen uniformly at random from the cube.

In your probability courses you studied repeated Bernoulli trials (e.g. coin-tossing) with some success probability p. For fixed number n of trials the finite sample space $\{0, 1\}^n$ with 2^n elements is sufficient. However, if the number of trials is unlimited, or for infinite series of trials a suitable sample space to model possible outcomes would be

$$\Omega = \{(\omega_1, \omega_2, \dots) : \omega_i = 0 \text{ or } 1, \text{for } i = 1, 2, \dots\} = \{0, 1\}^\infty,$$

so one outcome is an infinite sequence like $(0, 1, 1, 0, \dots)$. Identifying 1 with a 'head' the event A 'first two tosses are heads' is $A = \{\omega \in \Omega : \omega_1 = \omega_2 = 1\}$ with $\mathbb{P}(A) = p^2$. More complex events are required to formulate theorems of probability theory like the Strong Law of Large Numbers, hence the same question arises: what is the reserve of events A to make sense of $\mathbb{P}(A)$?

1.2 Definition of measure

To pursue the idea that a measure is a σ-additive function of a set, the domain of definition of a measure should be a system of sets closed under the operations of taking countable union, and also intersection and complementation. In this context 'closed' means that applying operations $\cap, \cup, ^c$ to a countable selection of sets from the system will yield another set from the system. We write $A^c = \Omega \setminus A$ for the complement.

Definition 1.1. *A σ-algebra $\mathcal{F}$ on a set Ω is a family of subsets of Ω with the following properties:*

(i) $\Omega \in \mathcal{F}$,

(ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$,

(iii) $A_j \in \mathcal{F}, j \in \mathbb{N}, \Rightarrow \bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$.

A system $\mathcal{A}$ of subsets of Ω is an *algebra* if $\mathcal{A}$ satisfies (i), (ii) and (iii′): $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$. Thus algebra is closed under the operation of taking union of finitely many sets, while σ-algebra admits countable unions.

Conditions (i), (ii), (iii) is a minimal set of axioms defining σ-algebra. Using these other properties are derived. So $\varnothing \in \mathcal{F}$ by (i), (ii). Then $A_1, A_2 \in \mathcal{F} \Rightarrow A_1 \cup A_2 \in \mathcal{F}$ because we can set $A_j = \varnothing$ for $j \geq 2$ in (ii). Using complementation rules, $A_j \in \mathcal{F}, j \in \mathbb{N}, \Rightarrow \bigcap_{j=1}^{\infty} A_j \in \mathcal{F}$. And so on.

Note: operating with more than countably many sets from $\mathcal{F}$ may lead to outside of $\mathcal{F}$. Indeed, *every* subset of Ω is a union of it individual points.

The power set $\mathcal{P}(\Omega)$ (all subsets) is a σ-algebra. If Ω is countable this is often a reasonable setting for a measure.

Let $A_1, A_2, \cdots$ be a partition of Ω in disjoint subsets. Taking arbitrary unions of A_j's will define a σ-algebra.

But typically σ-algebras have too many sets to admit explicit description. However, with each collection of sets $\mathcal{S}$ we can associate a σ-algebra generated by $\mathcal{S}$, which we denote $\sigma(\mathcal{S})$. Observe that for σ-algebras $\mathcal{F}_1, \mathcal{F}_2$ also the intersection $\mathcal{F}_1 \cap \mathcal{F}_2$ is a σ-algebra. For any system of σ-algebras $(\mathcal{F}_j, j \in J)$ (possibly with uncountable index set J) also $\cap_{j \in J} \mathcal{F}_j$, is a σ-algebra. Therefore, we can specify any collection $\mathcal{S} \subset \mathcal{P}(\Omega)$ of *generators* and define $\sigma(\mathcal{S})$ to be the intersection of all σ-algebras that contain $\mathcal{S}$. Thus the $\sigma(\mathcal{S})$ generated by $\mathcal{S}$ is the smallest σ-algebra containing $\mathcal{S}$.

Examples

1. Consider $\mathcal{S} = \{\varnothing\}$. The generated σ-algebra is the smallest possible, $\{\varnothing, \Omega\}$.

2. Consider $\mathcal{S} = \{A_1, \ldots, A_k\}$, where $A_1 \cup \cdots \cup A_k = \Omega$, A_j's are nonempty and pairwise disjoint. We speak in this situation of a partition of Ω with blocks (or atoms) A_j. Every set in $\sigma(\mathcal{S})$ is obtained by selecting some of the A_j's and taking union, e.g. $A_2 \cup A_3 \cup A_7$ (provided $k \geq 7$). There are 2^k ways to select a subset from a set with k elements, therefore $\sigma(\mathcal{S})$ has 2^k elements.

3. Taking partition $\mathcal{S} = \{A_1, A_2, \cdots\}$ into countably many (disjoint, nonempry) blocks will result in $\sigma(\mathcal{S})$ with continuum elements.

4. Consider the coin-tossing space $\Omega = \{0,1\}^\infty$. For each k and $(\epsilon_1, \ldots, \epsilon_k) \in \{0,1\}^k$ let $A(\epsilon_1, \ldots, \epsilon_k) = \{\omega \in \Omega : \omega_1 = \epsilon_1, \ldots, \omega_k = \epsilon_k\}$, a set called finite-dimensional cylinder. Let $\mathcal{F}_k$ be generated by the partition with parts $A(\epsilon_1, \ldots, \epsilon_k)$, where k is fixed; so the cardinality of $\mathcal{F}_k$ is 2^k. Observe that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ is an increasing sequence of σ-algebras, we call such sequence *filtration*. In the coin-tossing interpretation, the event $A(1,0,1,1)$ occurs when the first outcomes are $1,0,1,1$. So $\mathcal{F}_k$ incorporates the information contained in the first k coin-tosses. As more trials are observed, we get more information.

   Union of σ-algebras need not be a σ-algebra (not closed under $\cap$). So we let $\mathcal{F} = \sigma(\cup_{k=1}^{\infty} \mathcal{F}_k)$, which is the σ-algebra generated by *all* $A(\epsilon_1, \ldots, \epsilon_k)$'s, that is with k and ϵ_j's freely chosen. Think of $\mathcal{F}$ as complete information gathered after infinitely many trials.

   This $\mathcal{F}$ is rich enough to state the 'strong laws' of probability theory. For example, the event

   $$A = \{\omega \in \Omega : \lim_{k \to \infty} (\omega_1 + \cdots + \omega_k)/k = 1/2\}$$

   is in $\mathcal{F}$, but does not belong to $\mathcal{F}_k$ for any k. Indeed, we can only compute the long-run frequency of heads as infinitely many coin tosses have been observed. If $p = 1/2$ (the coin is fair), then $\mathbb{P}(A) = 1$, but $\mathbb{P}(A) = 0$ for $p \neq 1/2$. Indeed, recall the Law of Large Numbers.

σ-algebra of Borel sets Define the *Borel σ-algebra* on $\mathbb{R}$, denoted $\mathcal{B}(\mathbb{R})$, as the σ-algebra generated by the set of semi-open intervals $\{(a,b] : -\infty < a < b \leq \infty]\}$. Elements of $\mathcal{B}(\mathbb{R})$ are called *Borel-measurable* or *Borel sets*. Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is a universum of sets sufficient for all practical purposes. There are many other ways to select the set of generators for $\mathcal{B}(\mathbb{R})$: we can take for $\mathcal{S}$ all open sets, or all closed sets. A 'spare' collection of generators $\mathcal{S}$ for the Borel σ-algebra is the set of half-lines $\{(-\infty, x] : x \in \mathbb{R}\}$. This can be further reduced to the countable collection of half-lines $\{(-\infty, x] : x \in \mathbb{Q}\}$.

Sometimes it is useful to employ conditions on σ-algebras other that the defining axioms (i),(ii),(iii). Next are two commonly used characterisations.

Proposition 1.2. (The monotone class characterisation.) *If algebra $\mathcal{A}$ satisfies the conditions: for $A_n \in \mathcal{A}, n \geq 1$,*

$$A_1 \subset A_2 \subset \cdots \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A},$$
$$A_1 \supset A_2 \supset \cdots \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{A},$$

then $\mathcal{A}$ is a σ-algebra.

Proposition 1.3. (Dynkin's $\pi - \lambda$-system.) *A system $\mathcal{S}$ of subsets in Ω is a σ-algebra if it satisfies*

$$\pi\text{−system condition} \quad A_1, \ldots, A_n \in \mathcal{S} \Rightarrow \bigcap_{k=1}^{n} A_k \in \mathcal{S}$$

and

$$\lambda\text{−system conditions} \quad \begin{cases} \Omega \in \mathcal{S}, \\ A, B \in \mathcal{S}, A \subset B \Rightarrow B \setminus A \in \mathcal{S}, \\ A_n \in \mathcal{S}, n \geq 1; A_1 \subset A_2 \subset \cdots \Rightarrow \cup_{n=1}^{\infty} A_n \in \mathcal{S}. \end{cases}$$

A pair $(\Omega, \mathcal{F})$, which is set Ω endowed with a $\sigma-$algebra $\mathcal{F}$, is called *a measurable space*.

Definition 1.4. *Let $(\Omega, \mathcal{F})$ be a measurable space. A measure on Ω is a nonnegative function*

$$\mu : \mathcal{F} \to [0, \infty]$$

such that $\mu(\varnothing) = 0$ and the σ-additivity property holds:

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i), \tag{1}$$

for disjoint sets $A_i \in \mathcal{F}, i \in \mathbb{N}$. The triple $(\Omega, \mathcal{F}, \mu)$ is referred to as a measure space.

By the definition $\mu(A)$ is nonnegative, and the value ∞ is allowed. If $\mu(\Omega) < \infty$ we say that μ is a finite measure. If $\mu(\Omega) = 1$ we call μ probability measure, and often use notation $\mathbb{P}$. In the probability context we call measurable sets $A \in \mathcal{F}$ events, to which probability $\mathbb{P}(A)$ is assigned.

Example For fixed $x \in \Omega$, suppose $\{x\} \in \mathcal{F}$ (the one-point set is measurable). *Dirac measure* at x is

$$\delta_x(A) = \begin{cases} 1, \text{if } x \in A, \\ 0, \text{if } x \notin A. \end{cases}$$

Example Choose $x_1, x_2, \ldots$ from Ω and let $y_1, y_2, \ldots$ be positive numbers. A *discrete* measure is defined as

$$\mu(A) = \sum_{i=1}^{\infty} y_i \delta_{x_i}(A), \quad A \in \mathcal{F}.$$

Plainly, mass y_i sits in point x_i, so to compute the measure of set A you calculate the total mass of *atoms* x_i in this set. If Ω is countable, e.g. $\Omega = \mathbb{N}$ then every measure on $(\Omega, \mathcal{P}(\Omega))$ is discrete. The set of atoms of a discrete measure on $\mathbb{R}$ need not consist of isolated points like $\mathbb{N}$ or $\mathbb{Z}$, rather may have accumulation points and even be everywhere dense. For instance, enumerate rationals $\mathbb{Q}$ and put mass 2^{-i} on the ith point; then every interval contains infinitely many atoms.

The last example points at the following simple fact: for measures $\mu_1, \mu_2, \ldots$ on $(\Omega, \mathcal{F})$ and nonnegative reals $y_1, y_2, \ldots$, the linear combination $\sum_{i=1}^{\infty} y_i \mu_i$ is also a measure on $(\Omega, \mathcal{F})$.

Next we list useful properties of measure implied by (and in fact equivalent to) the σ-additivity. Let $A_i \in \mathcal{F}, i \in \mathbb{N}$.

1. Increasing tower of sets, monotonicity:
$$A_1 \subset A_2 \subset \cdots \Rightarrow \mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mu(A_i).$$
To see this, apply the σ-additivity property to the union of disjoint sets $A_{i+1} \setminus A_i$. Note that $\mu(A_i)$ is nondecreasing in i in this case.

2. Decreasing tower of sets, monotonicity:
$$A_1 \supset A_2 \supset \cdots \Rightarrow \mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mu(A_i).$$
This is obtained from the above increasing case by passing to complements.

3. Subadditivity:
$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} \mu(A_i).$$
The measurable sets A_i here need not be disjoint. If the left-hand side finite, the measure of the union can be expressed by the inclusion-exclusion formula.

1.3 Construction of measures by extension

Having introduced the general concept of measure, we wish to return to our principal example. We have the length $\lambda(A)$ defined for intervals $[a, b] \subset \mathbb{R}$ and some other sets of relatively simple nature. Is it possible to have λ well defined for *all* Borel sets, consistently with the definition of intervals? This is the fundamental problem of *measure extension*, which we may treat in the general setting.

Recall that a system of sets $\mathcal{A} \subset \mathcal{P}(\Omega)$ is an algebra if it satisfies conditions (i),(ii) from Definition 1.1, and is closed under finite unions. A function on algebra $\mu_0 : \mathcal{A} \to [0, \infty]$ is called a *pre-measure* if it satisfies (1) whenever $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$. The difference between pre-measure and measure is that a pre-measure is defined on algebra, which need not be closed under *countable* unions.

These concepts are best seen on our main example, the set $\mathbb{R}$. Let $\mathcal{S}$ be the set of intervals $(a, b]$, this is a generator of the Borel σ-algebra. Let $\mathcal{A}$ be the collection of sets $A \subset \mathbb{R}$ representable as finite unions of disjoint intervals,
$$A = \bigcup_{i=1}^{k} (a_i, b_i],$$
one may check that $\mathcal{A}$ is an algebra. We have the length defined on $\mathcal{A}$ by the formula
$$\lambda(A) = \sum_{i=1}^{k} (b_i - a_i).$$

Note that a countable union of disjoint intervals *may* belong to $\mathcal{A}$, for example $(0, 1/2] \cup (1/2, 3/4] \cup (3/4, 7/8] \cup \cdots = (0, 1]$. The length λ (which is a *pre-measure* for a time being, until extended) is σ-additive on $\mathcal{A}$.

The next is the measure extension theorem due to Carathéodory.

Theorem 1.5. *Suppose μ_0 is a pre-measure on $(\Omega, \mathcal{A})$, where $\mathcal{A}$ is an algebra. Then there is a measure on $(\Omega, \sigma(\mathcal{A}))$ such that*

$$\mu(A) = \mu_0(A) \quad \text{for } A \in \mathcal{A}.$$

Moreover, this measure μ is unique if there exists a sequence of sets $B_1 \subset B_2 \dots$ such that $\cup_{j=1}^{\infty} B_j = \Omega$, $B_j \in \mathcal{A}$ and $\mu_0(B_j) < \infty$ for all $j \in \mathbb{N}$.

If $\cup_{j=1}^{\infty} B_j = \Omega$, for some $B_j \in \mathcal{F}, j \in \mathbb{N}$, such that $\mu(B_j) < \infty$ for all $j \in \mathbb{N}$, we call measure μ *σ-finite.* Carathéodory's Theorem entails that a σ-finite measure on $(\Omega, \mathcal{F})$ is uniquely determined by its values on some algebra of generators.

By Carathéodory's Theorem , the length λ defined initially on intervals has a unique extension to the Borel σ-algebra. The extended measure is called the *Lebesque measure on $\mathcal{B}(\mathbb{R})$*.

Example. Let us look how to define probability as a measure on $\Omega = \{0, 1\}^{\infty}$, to give a rigorous meaning to the notion of 'infinitely many independent Bernoulli trials with success probability p'.

Fix p and for each cylinder set $A(\epsilon_1, \dots, \epsilon_k)$ let

$$\mathbb{P}(A(\epsilon_1, \dots, \epsilon_k)) = p^t(1-p)^{k-t}, \quad \text{where} \quad t = \epsilon_1 + \cdots + \epsilon_k. \tag{2}$$

The union $\mathcal{A} = \cup_{k=1}^{\infty} \mathcal{F}_k$ is an algebra, and $\mathbb{P}$ is a pre-measure on $(\Omega, \mathcal{A})$. By Carathéodory's theorem there is a probability measure consistent with (2) and defined on $\mathcal{F} = \sigma(\mathcal{A})$. This probability measure is unique because $\mathbb{P}(\Omega) = 1$ is finite. In particular, it is meaningful to assign probability to the event

$$A = \{\omega \in \Omega : \lim_{k \to \infty} (\omega_1 + \cdots + \omega_k)/k = z\}$$

(which is $\mathbb{P}(A) = 1$ if $z = p$, and $\mathbb{P}(A) = 0$ if $z \neq p$).

Example Take $\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}^{\infty}$ with σ-algebra generated by finite-dimensional cylinder sets, and the probability measure making the coordinates in Ω to independent random variables $X_1, X_2, \cdots$ with uniform distribution on $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Define a random real number by the decimal expansion

$$U = \sum_{n=1}^{\infty} \frac{X_n}{10^n}.$$

The distribution of U is a probability measure on $[0, 1]$, what is this measure? Each interval of the kind $((k-1)10^{-n}, k10^{-n}]$ corresponds to a cylinder set in Ω, so has probability 10^{-n}, which suggests that the distribution of U is uniform (that is Lebesgue measure on $[0, 1]$). This can be justified by application of Carathéodory's theorem, since the intervals $((k-1)10^{-n}, k10^{-n}]$ comprise an algebra.

Construction of measures on $\mathbb{R}$ via the distribution function. Measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ which give finite mass to halflines $(-\infty, x]$ can be defined in terms of generalised distribution functions. Let $F : \mathbb{R} \to (0, \infty)$ be a nondecreasing right-continuous function with left limits and $\lim_{x \to -\infty} F(x) = 0$. We define the measure of halfline $(-\infty, x]$ to be

$$\mu(-\infty, x] = F(x). \tag{3}$$

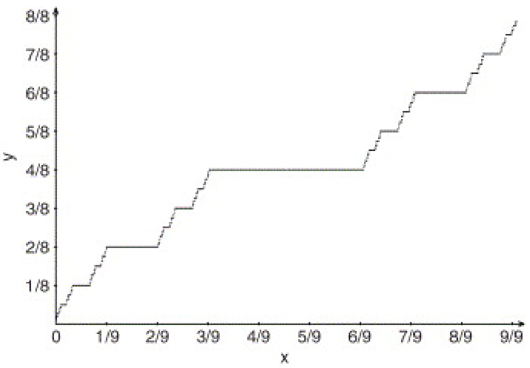
This is extended to intervals by $\mu(a, b] = F(b) - F(a)$ and is extendible to all Borel sets in a unique way by Carathéodory's theorem. If $\lim_{x \to \infty} F(x) = 1$ the measure μ is a probability measure, and F its cumulative distribution function. This method is very general, and allows one to construct both discrete distributions (e.g. supported by $\mathbb{N}$) and probability distributions with densities. The correspondence defined by (3) is invertible, in the sense that for every μ with $\mu(-\infty, x] < \infty, x \in \mathbb{R}$ the function F defined by this formula has the above properties (nondecreasing, etc).

If F has a jump at x, then the corresponding μ has an atom at x of mass $\mu(\{x\}) = F(x) - \lim_{k\to\infty} F(x - 1/k)$. If F has a density, in the sense that

$$F(x) = \int_{-\infty}^{x} f(z)dz \tag{4}$$

then the measure of each point $\{x\}$ is zero, in which case we say that the measure is non-atomic (or diffuse). Conversely, if F is continuous then the associated measure is non-atomic, but this does not mean that the measure has a density!

Example Cantor distribution function (see the picture) is an example of a probability measure which is non-atomic, but has no density to represent F as integral (4). Under this measure, the Cantor set has full probability $\mu(C) = 1$ although its Lebesgue measure is $\lambda(C) = 0$; we say that the Cantor distribution is singular.



We have seen that a measure on $\mathcal{B}(\mathbb{R})$ may be discrete, may have a density or may be singular. A measure decomposition theorem says that these exhaust, in a sense, all possibilities. Specifically, if a measure is σ-finite, then the measure can be represented as sum of three component measures: discrete, absolutely continuous (having a density) and a singular measure.

Example Let $X_1, X_2, \cdots$ be i.i.d. with any distribution on $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ different from uniform but giving positive probability to each digit. Define

$$Z = \sum_{n=1}^{\infty} \frac{X_n}{10^n}.$$

The distribution of such random variable is singular. Indeed, if $p \neq 10^{-1}$ is the probability of digit j then the long-run frequency (i.e. proportion among first n as $n \to \infty$) of digit j in Z is p; which is event of zero probability under the uniform distribution.

1.4 Lebesgue measure and Lebesgue measurable sets in $\mathbb{R}^k$

The Lebesgue measure on the line has natural generalisation to Euclidean spaces $\mathbb{R}^k$. For a rectangular parallelepiped $A = [a_1, b_1] \times \cdots \times [a_k, b_k]$ its Lebesgue measure is defined as the k-dimensional volume

$$\lambda^{(k)}(A) = \prod_{i=1}^{k} (b_i - a_i).$$

The σ-algebra of Borel sets $\mathcal{B}(\mathbb{R}^k)$ in k dimensions is the σ-algebra generated by open sets in $\mathbb{R}^k$. Like in $\mathbb{R}$, there is a more spare systems of generators generalising the half-lines in one dimension

$$\mathcal{S} = \{(-\infty, x_1] \times \cdots \times (-\infty, x_k] : (x_1, \ldots, x_k) \in \mathbb{R}^k\}.$$

There is a larger than $\mathcal{B}(\mathbb{R}^k)$ σ-algebra of sets, to which the Lebesgue measure can be extended. If A is a Borel set with $\lambda^{(k)}(A) = 0$ and $B \subset A$ it is reasonable to assign to B measure 0. The σ-algebra generated by $\mathcal{B}(\mathbb{R}^k)$ and such null-subsets B is the σ-algebra of *Lebesgue-measurable* sets. This operation of adding subsets of zero-measure sets is called *completion*, that is the σ-algebra of Lebesgue-measurable sets is complete. We describe the basic steps of the completion.

Definition 1.6. A system of subset $\mathcal{S}$ in Ω is a *semiring* if

(a) $A, B \in \mathcal{S} \Rightarrow A \cap B \in \mathcal{S}$,

(b) $A, B \in \mathcal{S} \Rightarrow A \setminus B = C_1 \cup \cdots \cup C_n$ for some disjoint $C_i \in \mathcal{S}$.

Definition 1.7. Let μ be a pre-measure on semiring $\mathcal{S}$. For $A \subset \Omega$ define *exterior* measure

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

where the infinum is taken over all covers (i.e. $\cup_{n=1}^{\infty} A_n \supset A$) of A by $A_n \in \mathcal{S}$. A set A is said to be *Lebesgue-measurable* if $\mu^*(A \Delta B_n) \to 0$ for some sequence of sets $B_1, B_2, \ldots \in \mathcal{S}$. We denote $L(\mathcal{S}, \mu)$ the collection of Lebesgue-measurable sets.

Theorem 1.8. (Lebesgue's theorem.) *The family of sets $L(\mathcal{S}, \mu)$ is a σ-algebra, and μ^* is a measure on $L(\mathcal{S}, \mu)$ extending μ on $\mathcal{S}$.*

In our main example $\Omega = \mathbb{R}$, where the finite unions $A = \cup_{j=1}^{n} (a_j, b_j]$ of disjoint intervals comprise a semiring $\mathcal{S}$. The Lebesgue measure is thus extendible to the family of Lebesgue-measurable sets $L(\mathcal{S}, \lambda)$. If $A \subset \mathcal{B}(\mathbb{R})$ is a null-set, with $\lambda(A) = 0$, then every $B \subset A$ belongs to $L(\mathcal{S}, \lambda)$ and has $\lambda(B) = 0$. Let $\mathcal{N}$ denote the family of such sets B that appear as subsets of Borel null-sets. Then $L(\mathcal{S}, \lambda) = \sigma(\mathcal{B}(\mathbb{R}), \mathcal{N})$, that is Lebesgue-measurable sets comprise the σ-algebra generated by Borel sets and their null subsets.

It is important to note that $\mathcal{B}(\mathbb{R})$ is defined regardless of any measure, while $L(\mathcal{S}, \mu)$ depends on how the measure μ is chosen.

Using transfinite induction, it can be shown that the cardinality of the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is continuum. On the other hand, by Cantor's theorem from Set Theory, for Cantor set C the cardinality of the power-set $\mathcal{P}(C)$ is bigger than continuum, and each $A \subset C$ is Lebesgue-measurable. It follows that there are more Lebesgue-measurable sets than Borel sets. Hence many Lebesgue-measurable non-Borel sets exist, although they do not admit a constructive description.

1.5 Measurable spaces $(\mathcal{R}^\infty, \mathcal{B}(\mathcal{R}^\infty))$ and $(\mathcal{R}^T, \mathcal{B}(\mathcal{R}^\infty))$

The product space $\mathbb{R}^\infty$ is the space of sequences $(x_1, x_2, \ldots), x_k \in \mathbb{R}$. Let for $B \in \mathcal{B}(\mathbb{R}^n)$

$$C_n(B) = \{(x_1, x_2, \ldots) : (x_1, \ldots, x_n) \in B\},$$

which is a finite-dimensional cylinder set. Disjoint unions of such cylinder sets (of same or different dimensions) comprise an algebra, as is easy to check. The σ-algebra generated by the cylinder sets is the Borel σ-algebra denoted $\mathcal{B}(\mathbb{R}^\infty)$. A smaller set of generators is the set of parallelepipeds $B = (a_1, b_1] \times \cdots \times (a_n, b_n]$.

In applications, the measurable space $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ endowed with some probability measure $\mathbb{P}$ models a sequence of (in general, dependent) outcomes $X_1, X_2, \ldots$ of a series of random experiments. In practice, however, we are given some way to describe the joint distribution P_n of $(X_1, \ldots, X_n)$ for each n. This begs the question if the finite-dimensional distributions $P_n, n \geq 1$, indeed determine a probability measure on the infinite-dimensional space $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$. The key concept here is consistency. A cylinder $C_n(B)$ can be assigned probability $P_n(B)$ in terms of P_n, but also in terms of P_{n+1} as $P_{n+1}(B \times \mathbb{R})$.

Definition 1.9. Let P_n be probability measures on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), n \geq 1$. The measures are said to be *consistent* if for all $n \geq 1, B \in \mathcal{B}(\mathbb{R}^n)$

$$P_{n+1}(B \times \mathbb{R}) = P_n(B).$$

Theorem 1.10. (Kolmogorov's measure extension theorem.) *Let P_n be consistent probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ for $n = 1, 2, \ldots$. There exists a unique probability measure $\mathbb{P}$ on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ such that for every n*

$$\mathbb{P}(C_n(B)) = P_n(B), \quad B \in \mathcal{B}(\mathbb{R}^n).$$

Thus, for discrete-time random process $X_1, X_2, \ldots$ the probability law of the whole process is uniquely determined by consistent finite-dimensional distributions of $(X_1, \ldots, X_n), n \geq 1$.

The space $\mathbb{R}^T$ is the space of functions (x_t) from the index set T to $\mathbb{R}$. The Borel σ-algebra $\mathcal{B}(\mathbb{R}^T)$ is generated by cylinder sets of the form

$$C_{t_1,\ldots,t_n}(B) = \{(x_t) : (x_{t_1}, \ldots, x_{t_n}) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}^n),\ n \geq 1,$$

where $\{t_1, \ldots, t_n\} \subset T$ is any collection of distinct t_i's. In fact, every $A \in \mathcal{B}(\mathbb{R}^T)$ can be represented as 'infinite-dimensional cylinder' of the form

$$A = \{(x_t) : (x_{t_1}, x_{t_2}, \ldots) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}^\infty)$$

for some t_i's and B.

To be definite, we may focus on $T = [0, \infty)$ thought of as time span of some random process. Suppose we are given a family $P_{t_1,\ldots,t_n}$ of probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ for each choice of times $\{t_1, \ldots, t_n\} \subset T$. The family is called *consistent* if for $\{s_1, \ldots, s_k\} \subset \{t_1, \ldots, t_n\}$ and $B \in (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ it holds that

$$P_{t_1,\ldots,t_n}(\{(x_{t_1}, \ldots, x_{t_n}) : (x_{s_1}, \ldots, x_{s_k}) \in B\}) = P_{s_1,\ldots,s_k}(B).$$

Another Kolmogorov's theorem, generalising Theorem 1.10, states that for consistent family of probability measures $P_{t_1,\ldots,t_n}$ there exists a unique probability measure $\mathbb{P}$ on $(\mathcal{R}^T, \mathcal{B}(\mathcal{R}^\infty))$ such that

$$\mathbb{P}(C_{t_1,\ldots,t_n}(B)) = P_{t_1,\ldots,t_n}(B), \quad B \in (\mathcal{R}^n, \mathcal{B}(\mathcal{R}^n)).$$

The latter is sometimes called 'theorem about existence of the process': consistent finite-dimensional distributions uniquely determine the probability law of the process as a whole.

Exercises

1. For $A \subset \Omega$ proper subset, describe $\sigma(\{A\})$.
2. Let $\Omega = [0, 1]$. Describe the σ-algebra generated by $\{[0, 1/4), (3/4, 1]\}$ by listing all its elements.
3. Show that the increasing monotonicity property is equivalent to σ-additivity.
4. Let $\mathcal{A}$ be the family of sets $A \in \mathcal{B}(\mathbb{R})$ with the property that there exists a limit
$$\mu(A) = \lim_{n\to\infty} n^{-1}\lambda(A \cap [0, n]).$$
Show that $\mathcal{A}$ is an algebra. Is μ σ-additive on $\mathcal{A}$?
5. Consider the space of functions $x : T \to \mathbb{R}$ on some set T. Show that the sets of the form
$$A = \{x : (x(t_1), \dots, x(t_k)) \in D\}$$
for some k, $t_1 < \dots < t_k$ and $D \in \mathcal{B}(\mathbb{R}^k)$ comprise an algebra.
6. Let $\Omega = [0, 1]$, and for any rational interval $I = (a, b) \cap \mathbb{Q}, (a, b] \cap \mathbb{Q}, [a, b) \cap \mathbb{Q}, [a, b] \cap \mathbb{Q}$ with $a, b \in \mathbb{Q}$ let $\mu(I) = b - a$. Consider algebra $\mathcal{A}$ consisting of finite disjoint unions of such intervals. Show that μ as a function on $\mathcal{A}$ is finitely additive, but not σ-additive.
7. For $A \subset \mathbb{R}$ define $x + A := \{x + a, a \in A\}$. Prove translation invariance of the Lebesgue measure: $\lambda(x + A) = \lambda(A), A \in \mathcal{B}(\mathbb{R})$. Extend the property to Lebesgue-measurable sets A.
8. Explain why the distribution function of a random variable is right-continuous with left limits.
9. Show that every probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ admits a representation $\mu + \nu$, where μ is a discrete measure and ν is a diffuse measure (with $\nu(\{x\}) = 0$ for $x \in \mathbb{R}$).
10. Let $\mu = \sum_{j=1}^{\infty} 2^{-j}\delta_j$. Is it a probability measure? Sketch the graph of its cumulative distribution function.
11. Let $\Omega = \{0, 1\}^{\infty}$. Using set-teoretic operations $\cup, \cap, ^c$ express the event
$$A = \{\omega \in \Omega : \lim_{k\to\infty} (\omega_1 + \dots + \omega_k)/k = z\}$$
in terms of events $A(\epsilon_1, \dots, \epsilon_k)$.
12. (First half of the Borel-Cantelli lemma) Let $A_j, j \in \mathbb{N}$, be events in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\sum_{j=1}^{\infty} \mathbb{P}(A_j) < \infty$. Prove that $\mathbb{P}(\cap_{n=1}^{\infty} \cup_{j=n}^{\infty} A_j) = 0$.
13. Consider $\mathcal{S} := \{\{x\} : x \in \mathbb{R}\}$. Show that for $A \in \sigma(\mathcal{S})$, either A is countable (i.e. either finite or countably infinite) or A^c is countable. Now let $\mu(x) = 1$ for every $x \in \mathbb{R}$. What are possible values of $\mu(A)$? When $\mu(A) = \infty$?
14. For Borel sets $A, B \in \mathcal{B}(\mathbb{R})$ let $d(A, B) = \lambda(A\Delta B)$. Show that $d(A, B)$ is a metric on $\mathcal{B}(\mathbb{R})$ (in particular, satisfies the triangle inequality).

Literature

1. S. Resnick, A probability path, Springer 2003.

2. R. Schilling, Measures, integrals and martingales, CUP 2005.