

# 1 Measure Theory

## 1.1 Why Measure Theory?

There are two different views – not necessarily exclusive – on what “probability” means: the *subjectivist* view and the *frequentist* view. To the subjectivist, probability is a system of laws that should govern a rational person’s behavior in situations where a bet must be placed (not necessarily just in a casino, but in situations where a decision must be made about how to proceed when only imperfect information about the outcome of the decision is available, for instance, should I allow Dr. Scissorhands to replace my arthritic knee by a plastic joint?). To the frequentist, the laws of probability describe the *long-run* relative frequencies of different events in “experiments” that can be repeated under roughly identical conditions, for instance, rolling a pair of dice. For the frequentist interpretation, it is imperative that *probability spaces* be large enough to allow a description of an experiment, like dice-rolling, that is repeated infinitely many times, and that the mathematical laws should permit easy handling of *limits*, so that one can make sense of things like “the probability that the long-run fraction of dice rolls where the sum of the dice is 7”. But even for the subjectivist, the laws of probability should allow for description of situations where there might be a continuum of possible outcomes, or possible actions to be taken. Once one is reconciled to the need for such flexibility, it soon becomes apparent that measure theory is the only way to go.

## 1.2 Algebras and $\sigma$ –algebras

**Definition 1.1.** An *algebra* of subsets of  $\Omega$  is a collection of sets that contains the empty set  $\emptyset$  and is closed under *complementation* and *finite unions*. A  $\sigma$ –*algebra* is an algebra that is closed under *countable unions*.

**Exercise 1.2.** Show that

- (a) Every algebra is closed under finite intersections and set-theoretic differences.
- (b) Every  $\sigma$ –algebra  $\mathcal{F}$  is closed under countable monotone unions and intersections, i.e.,

$$\begin{aligned} A_1 \subset A_2 \subset A_3 \subset \cdots \quad \text{and} \quad A_n \in \mathcal{F} &\implies \bigcup_{n \geq 1} A_n \in \mathcal{F} \\ A_1 \supset A_2 \supset A_3 \supset \cdots \quad \text{and} \quad A_n \in \mathcal{F} &\implies \bigcap_{n \geq 1} A_n \in \mathcal{F}. \end{aligned}$$

**Observation 1.3.** If  $\{\mathcal{F}_\lambda\}_{\lambda \in \Lambda}$  is a (not necessarily countable) collection of  $\sigma$ –algebras on a set  $\Omega$ , then the intersection  $\bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda$  is also a  $\sigma$ –algebra. This is important, because it implies that for any collection  $\mathcal{S}$  of subsets of  $\Omega$  there is a *smallest*  $\sigma$ –algebra containing  $\mathcal{S}$ . This  $\sigma$ –algebra is denoted by  $\sigma(\mathcal{S})$ , and called the  $\sigma$ –*algebra generated by*  $\mathcal{S}$ .

The same argument shows that there is a smallest *algebra* containing  $\mathcal{S}$ ; this must be contained in  $\sigma(\mathcal{S})$ .

**Definition 1.4.** If  $(\mathcal{X}, d)$  is a metric space then the *Borel  $\sigma$ -algebra* on  $\mathcal{X}$ , denoted by  $\mathcal{B}_{\mathcal{X}}$ , is the  $\sigma$ -algebra generated by the collection  $\mathcal{U}$  of all open subsets of  $\mathcal{X}$ .

**Exercise 1.5.** Show that if  $\mathcal{X} = \mathbb{R}$  then the Borel  $\sigma$ -algebra  $\mathcal{B}_{\mathbb{R}}$  is the  $\sigma$ -algebra generated by the collection of open half-lines  $(a, \infty)$  where  $a \in \mathbb{Q}$ .

**Example 1.6.** Let  $\Omega = \{0, 1\}^{\mathbb{N}}$  be the space of infinite sequences of 0s and 1s. Given a finite sequence  $\varepsilon = \varepsilon_1 \varepsilon_2 \cdots \varepsilon_m$  define the *cylinder set*  $\Sigma_{\varepsilon}$  to be the set of all sequences  $x \in \Omega$  that agree with  $\varepsilon$  in the first  $m$  slots. Let  $\mathcal{B}_{\Omega}$  be the  $\sigma$ -algebra generated by the cylinder sets. The pair  $(\Omega, \mathcal{B}_{\Omega})$  will be referred to as *sequence space*.

**Exercise 1.7.** Define a metric on  $\Omega$  by setting  $d(x, y) = 2^{-n(x, y)}$  where  $n(x, y)$  is defined to be the maximum  $n$  such that  $x_i = y_i$  for all  $i \leq n$ . (a) Show that this is in fact a metric. (b) Show that the metric space  $(\Omega, d)$  is compact. (c) Show that the Borel  $\sigma$ -algebra on the metric space  $(\Omega, d)$  is the  $\sigma$ -algebra generated by the cylinder sets.

### 1.3 Measures

**Definition 1.8.** A *measure* on an algebra  $\mathcal{A}$  is a set function  $\mu : \mathcal{A} \rightarrow [0, \infty]$  satisfying  $\mu(\emptyset) = 0$  that is countably additive, i.e., if  $A_n \in \mathcal{A}$  and  $\cup_{n \geq 1} A_n \in \mathcal{A}$  then

$$\mu\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mu(A_n).$$

A *probability measure* is a measure with total mass 1, that is,  $\mu(\Omega) = 1$ . Since we only rarely will deal with measures on algebras, we will adopt the convention that *unless explicitly stated*, measures are always defined on  $\sigma$ -algebras.

**Properties of Measures.** If  $\mu$  is a measure on a  $\sigma$ -algebra  $\mathcal{F}$  then

- (a)  $\mu$  is *finitely additive*.
- (b)  $\mu$  is *monotone*, i.e.,  $A \subset B$  implies  $\mu(A) \leq \mu(B)$ .
- (c)  $\mu$  is *countably subadditive*, i.e., for any sequence  $A_n \in \mathcal{F}$ ,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n).$$

- (d) If  $\mu = P$  is a *probability measure* then  $P$  is *continuous from above and below*, i.e.,

$$\begin{aligned} A_1 \subset A_2 \subset A_3 \subset \cdots &\implies P\left(\bigcup_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} P(A_n), \\ A_1 \supset A_2 \supset A_3 \supset \cdots &\implies P\left(\bigcap_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} P(A_n) \end{aligned}$$

**Example 1.9.** If  $\mathcal{X}$  is a countable set then *counting measure* on  $\mathcal{X}$  is the measure on the  $\sigma$ -algebra  $2^{\mathcal{X}}$  defined by  $\mu(A) = \text{cardinality of } A$ .

**Example 1.10.** If  $\mathcal{X}$  is an *uncountable* set then the collection  $\mathcal{G}$  consisting of all countable (including finite) and co-countable sets is a  $\sigma$ -algebra. The set function  $\mu$  on  $\mathcal{G}$  such that  $\mu(A) = 0$  or  $\infty$  depending on whether  $A$  is countable or not is a measure. This measure is not very interesting, but in many circumstances it can be used as a *counterexample*.

**Exercise 1.11.** Show by example that the second assertion of (c) need not hold for measures in general.

Continuity from above and below lead directly to one of the most useful tools in probability, the (easy half of the) *Borel-Cantelli Lemma*. It is often the case that an interesting event can be expressed in the form  $\{A_n \text{ i.o.}\}$ , where  $A_n$  is some sequence of events and the notation  $A_n \text{ i.o.}$  is shorthand for “infinitely many of the events  $A_n$  occur”, formally,

$$\{A_n \text{ i.o.}\} = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n.$$

If each of the events  $A_n \in \mathcal{F}$ , then so is the event  $\{A_n \text{ i.o.}\}$ . The Borel-Cantelli Lemma is a simple necessary condition for checking whether the event  $A_n \in \mathcal{F}$  has positive probability.

**Proposition 1.12.** (*Borel-Cantelli Lemma*) Let  $\mu$  be a measure on a  $\sigma$ -algebra  $\mathcal{F}$ . For any sequence  $A_n \in \mathcal{F}$ ,

$$\sum_{n=1}^{\infty} \mu(A_n) < \infty \implies \mu\{A_n \text{ i.o.}\} = 0.$$

*Proof.* To show that  $\mu\{A_n \text{ i.o.}\} = 0$  it is enough to show that  $\mu\{A_n \text{ i.o.}\} < \varepsilon$  for every  $\varepsilon > 0$ . Since  $\sum \mu(A_n) < \infty$ , for any  $\varepsilon > 0$  there exists  $m \geq 1$  such that  $\sum_{n=m}^{\infty} \mu(A_n) < \varepsilon$ ; then

$$\mu\{A_n \text{ i.o.}\} \leq \mu(\bigcup_{n \geq m} A_n) \leq \sum_{n=m}^{\infty} \mu(A_n) < \varepsilon.$$

□

**Example 1.13.** *Diophantine Approximation.* How closely can a real number be approximated by a rational number? – Of course, this problem as stated has a trivial answer, because the construction of the real number system guarantees that the rational numbers are *dense* in the reals. But what if we ask instead *how closely can a real number be approximated by a rational with denominator at most  $m$* , for some large  $m$ ? Fix  $\alpha > 0$ , and for each integer  $m \geq 1$  define  $G_m$  to be the set of all  $x \in [0, 1]$  such that there is an integer  $0 \leq k \leq m$  for which

$$\left| x - \frac{k}{m} \right| \leq \frac{1}{m^{2+\alpha}}.$$

Now define  $G = \{G_m i.o.\}$ ; this is the set of all  $x \in [0, 1]$  such that for infinitely many  $m \geq 1$  there is a rational approximation to  $x$  with denominator  $m$  that is within distance  $1/m^{2+\alpha}$  of  $x$ .

**Proposition 1.14.** *For every  $\alpha > 0$ , the Lebesgue measure of the set  $G$  is 0.*

*Proof.* Exercise. □

## 1.4 Lebesgue Measure

**Theorem 1.15.** *There is a unique measure  $\lambda$  on the Borel  $\sigma$ -algebra  $\mathcal{B}_{\mathbb{R}}$  such that for every interval  $J \subset \mathbb{R}$  the measure  $\lambda(J)$  is the length of  $J$ . The measure  $\lambda$  is translation invariant in the sense that for every  $x \in \mathbb{R}$  and every Borel set  $A \in \mathcal{B}_{\mathbb{R}}$ ,*

$$\lambda(A + x) = \lambda(A).$$

The restriction of the measure  $\lambda$  to the Borel  $\sigma$ -algebra  $\mathcal{B}_{[0,1]}$  of subsets of  $[0, 1]$  is the *uniform distribution* on  $[0, 1]$ , also known as the Lebesgue measure on  $[0, 1]$ , and also denoted (in these notes) by  $\lambda$ . The uniform distribution on  $[0, 1]$  has a related, but slightly different, translation invariance property, stemming from the fact that the unit interval  $[0, 1]$  can be identified in a natural way with the *unit circle* in the plane, which is a *group* (under rotation). To formulate this translation invariance property, we define a new addition law on  $[0, 1]$ : for any  $x, y \in [0, 1]$ , set

$$x + y \mod 1$$

to be the unique element of  $[0, 1)$  that differs from  $x + y$  (the usual sum) by an integer (which will be either 0 or 1). Then the uniform distribution on  $[0, 1)$  has the property that for every Borel subset  $A \subset [0, 1)$  and every  $x \in [0, 1)$ ,

$$\lambda(A + x \mod 1) = \lambda(A).$$

The translation (or rotation, if you prefer) invariance of the uniform distribution on the circle seems perfectly natural. But there is a somewhat unexpected development.

**Proposition 1.16.** *Assume the Axiom of Choice. Then Lebesgue measure  $\lambda$  on  $[0, 1)$  does not admit a translation invariant extension to the  $\sigma$ -algebra consisting of all subsets of  $[0, 1]$ .*

The proof has nothing to do with probability theory, so I will omit it. But the proposition should serve as a caution that the existence of measures is a more delicate mathematical issue than one might think.

Although the uniform distribution on  $[0, 1]$  seems at first sight to be a very special example, there is a sense in which *every interesting random process can be constructed on the probability space*  $([0, 1], \mathcal{B}_{[0,1]}, \lambda)$ , which is sometimes referred to as *Lebesgue space*. We will return to this when we discuss measurable transformations later. For now, observe that it is possible to construct an infinite sequence of independent, identically distributed Bernoulli-1/2 random variables on Lebesgue space. To this end, set

$$\begin{aligned} Y_1 &= 1_{(.5,1]}, \\ Y_2 &= 1_{(.25,.5]} + 1_{(.75,1]}, \\ Y_3 &= 1_{(.125,.25]} + 1_{(.375,.5]} + 1_{(.675,.75]} + 1_{(.875,1]}, \\ &\dots \end{aligned}$$

You should check that these are independent Bernoulli-1/2, and you should observe that the functions  $Y_1, Y_2, \dots$  are the terms of the binary expansion:

$$x = \sum_{n=1}^{\infty} Y_n(x)/2^n.$$

Theorem 1.15 will be deduced from a more general extension theorem of Carathéodory, discussed in section 1.6 below. Carathéodory's theorem states that any measure on an algebra  $\mathcal{A}$  extends in a unique way to a measure on the  $\sigma$ -algebra  $\sigma(\mathcal{A})$  generated by  $\mathcal{A}$ . Thus, to prove Theorem 1.15 it suffices to show that there is a measure  $\lambda$  on the algebra  $\mathcal{A}$  generated by the intervals  $J \subset [0, 1]$  such that  $\lambda(J) = \text{length of } J$  for every interval  $J$ .

**Lemma 1.17.** *The algebra  $\mathcal{A}$  generated by the intervals  $J \subset [0, 1]$  consists of all subsets of  $[0, 1]$  that are finite unions of intervals.*

*Proof.* Exercise. □

**Lemma 1.18.** *There is a unique measure  $\lambda$  on the algebra  $\mathcal{A}$  generated by the intervals  $J \subset [0, 1]$  that satisfies  $\lambda(J) = |J|$  for every interval  $J$ .*

*Proof.* It suffices to show that for any interval  $J = (a, b)$ , if  $J = \cup_{i=1}^{\infty} J_i$  is a countable union of pairwise disjoint intervals  $J_i$  then

$$|J| = \sum_{n=1}^{\infty} |J_n|.$$

The inequality  $\geq$  is relatively routine, because for this it suffices to show that for any finite  $m \geq 1$  we have  $|J| \geq \sum_{n=1}^m |J_n|$ , and this follows because the intervals  $J_i = (a_i, b_i)$  can be arranged so that

$$a \leq a_1 \leq b_1 \leq a_2 \leq b_2 \leq \dots \leq b_n \leq b.$$

The hard part is the reverse inequality  $\leq$ . Fix  $\varepsilon > 0$  small, and define  $J_n^+$  to be the interval obtained by fattening  $J_n$ , i.e., if  $J_n = [a_n, b_n]$  then  $J_n^+ = (a_n - \varepsilon/2^n, b_n + \varepsilon/2^n)$ . The intervals  $J_n^+$

form an open cover of  $[a + \varepsilon, b - \varepsilon]$ , so there is a finite subcover  $(J_n^+)_{1 \leq n \leq m}$ . Therefore (by an easy argument)

$$b - a - 2\varepsilon \leq \sum_{n=1}^m (|J_n| + \varepsilon/2^n).$$

Since  $\varepsilon > 0$  is arbitrary, the desired inequality follows. □

## 1.5 Monotone Class Theorem and Dynkin's $\pi - \lambda$ Lemma

One of the most common chores in measure theory is to prove that some property  $\mathcal{G}$  holds for every set in a  $\sigma$ -algebra  $\mathcal{F}$ . There are two tools designed for this: the *monotone class* theorem and the  $\pi - \lambda$  lemma. The strategy for using these tools is this: first, show that property  $\mathcal{G}$  holds for all sets  $F$  in a certain smaller collection (e.g., for intervals, or for cylinder sets); second, show that the collection of all sets  $F$  for which property  $\mathcal{G}$  holds is a monotone class, or a  $\lambda$ -system (depending on which tool you've bought); third, utter the magic incantation "by the Monotone Class Theorem" (or "by the  $\pi - \lambda$  Lemma").

**Definition 1.19.** A *monotone class* on  $\Omega$  is a collection  $\mathcal{M}$  of subsets of  $\Omega$  that is closed under monotone unions and monotone intersections, that is, for any sequence  $A_n \in \mathcal{M}$ ,

- (i) if  $A_1 \subset A_2 \subset \dots$  then  $\cup_{n \geq 1} A_n \in \mathcal{M}$ , and
- (ii) if  $A_1 \supset A_2 \supset \dots$  then  $\cap_{n \geq 1} A_n \in \mathcal{M}$ .

Clearly, the intersection of any collection of monotone classes is again a monotone class; consequently, by the same logic as in Observation 1.3, for any collection  $\mathcal{A}$  of subsets of a given space  $\Omega$  there is a *smallest* monotone class  $\mathcal{M}$  containing  $\mathcal{A}$ .

**Proposition 1.20.** (*Monotone Class Theorem*) If  $\mathcal{A}$  is an algebra, then the smallest monotone class  $\mathcal{M}$  containing  $\mathcal{A}$  is  $\sigma(\mathcal{A})$ .

*Proof.* Every  $\sigma$ -algebra is a monotone class, so the smallest monotone class containing  $\mathcal{A}$  is contained in the smallest  $\sigma$ -algebra containing  $\mathcal{A}$ . Consequently, it suffices to show that  $\mathcal{M}$  is a  $\sigma$ -algebra, that is, that  $\mathcal{M}$  is closed under complementation and countable unions.

(1) Let  $\mathcal{M}_0$  be the subset of  $\mathcal{M}$  consisting of all  $F \in \mathcal{M}$  such that  $F^c \in \mathcal{M}$ . Clearly,  $\mathcal{A} \subset \mathcal{M}_0$ , because algebras are closed under complementation. Furthermore,  $\mathcal{M}_0$  is a monotone class. (Exercise: check this.) Since  $\mathcal{M}$  is the *smallest* monotone class containing  $\mathcal{A}$ , it follows that  $\mathcal{M}_0 = \mathcal{M}$ . This proves that  $\mathcal{M}$  is closed under complementation.

(2) Let  $\mathcal{M}_1$  be the subset of  $\mathcal{M}$  consisting of all  $F \in \mathcal{M}$  such that  $F \cap A \in \mathcal{M}$  for every  $A \in \mathcal{A}$ . Once again it is obvious that  $\mathcal{A} \subset \mathcal{M}_1$ , because algebras are closed under finite

intersections. Furthermore,  $\mathcal{M}_1$  is a monotone class, by an easy argument. Hence,  $\mathcal{M}_1 = \mathcal{M}$ .

(3) Let  $\mathcal{M}_2$  be the subset of  $\mathcal{M}$  consisting of all  $F \in \mathcal{M}$  such that  $F \cap M \in \mathcal{M}$  for every  $M \in \mathcal{M}$ . By (2), the collection  $\mathcal{M}_2$  contains  $\mathcal{A}$ . Furthermore,  $\mathcal{M}_2$  is a monotone class (exercise). Consequently,  $\mathcal{M}_2 = \mathcal{M}$ . This proves that  $\mathcal{M}$  is closed under finite intersections.

(4) By (1) and (3), the collection  $\mathcal{M}$  is closed under complements and finite intersections, and hence also under finite unions. It is now easy to see that  $\mathcal{M}$  is closed under countable unions: if  $M_1, M_2, \dots \in \mathcal{M}$ , then for each  $n \geq 1$ ,

$$\bigcup_{i=1}^n M_i := N_n \in \mathcal{M},$$

because  $\mathcal{M}$  is closed under finite unions; now

$$N_n \uparrow \bigcup_{i=1}^{\infty} M_i,$$

so it follows that  $\bigcup_{i \geq 1} M_i \in \mathcal{M}$ . □

Following is a typical application of the Monotone Class Theorem.

**Proposition 1.21.** *If  $\mu$  and  $\nu$  are probability measures on  $\sigma(\mathcal{A})$  that agree on  $\mathcal{A}$ , where  $\mathcal{A}$  is an algebra of subsets of  $\Omega$ , then  $\mu = \nu$  on  $\sigma(\mathcal{A})$ .*

*Proof.* Let  $\mathcal{H}$  be the collection of all  $H \in \sigma(\mathcal{A})$  such that  $\mu(H) = \nu(H)$ . We must show that  $\mathcal{H} = \sigma(\mathcal{A})$ . Since  $\mathcal{H}$  contains the algebra  $\mathcal{A}$ , it will suffice, by the Monotone Class Theorem, to show that  $\mathcal{H}$  is a monotone class.

Suppose that  $H_1 \subset H_2 \subset \dots$  is a non-decreasing sequence of sets in  $\mathcal{H}$ , that is, such that  $\mu(H_i) = \nu(H_i)$  for every  $i$ . Since probability measures are continuous from below,

$$\begin{aligned} \mu\left(\bigcup_{i=1}^{\infty} H_i\right) &= \lim_{i \rightarrow \infty} \mu(H_i) \quad \text{and} \\ \nu\left(\bigcup_{i=1}^{\infty} H_i\right) &= \lim_{i \rightarrow \infty} \nu(H_i); \end{aligned}$$

consequently,

$$\mu\left(\bigcup_{i=1}^{\infty} H_i\right) = \nu\left(\bigcup_{i=1}^{\infty} H_i\right).$$

This proves that  $\mathcal{H}$  is closed under increasing unions. A similar argument, using the continuity of probability measures from above, shows that  $\mathcal{H}$  is closed under decreasing unions. □

This proposition implies that Lebesgue measure on  $[0, 1]$  is unique, because any measure  $\mu$  such that  $\mu(J) = |J|$  for all intervals  $J$  is completely determined on the algebra generated by intervals, by additivity, and this algebra generates the Borel  $\sigma$ -algebra.

**Example 1.22.** This example shows that uniqueness need not hold for *infinite* measures. Let  $\Omega = [0, 1]$  and  $\mathcal{B}$  the collection of all Borel subsets of  $[0, 1]$ . The algebra  $\mathcal{A}$  consisting of all finite unions of *dyadic* intervals (that is, intervals of positive length with endpoints  $k/2^n$ ) generates  $\mathcal{B}$ . Define measures  $\mu, \nu$  on  $\mathcal{B}$  as follows:

$$\begin{aligned}\mu(B) &= \#B (= \text{cardinality of } B) \\ \nu(B) &= \infty \quad \forall B \neq \emptyset, \text{ and } \nu(\emptyset) = 0.\end{aligned}$$

The measures  $\mu$  and  $\nu$  agree on  $\mathcal{A}$ , because every non-empty set  $A \in \mathcal{A}$  has infinite cardinality. On the other hand, they disagree on *finite* sets  $B$ , so they are not the same set function on  $\mathcal{B}$ .

To use the Monotone Class Theorem to show that a property  $\mathcal{G}$  holds for all sets in a  $\sigma$ -algebra  $\mathcal{F}$ , one must start by showing that  $\mathcal{G}$  holds for all sets in an algebra  $\mathcal{A}$  that generates  $\mathcal{F}$ . This can sometimes be a bit of a slog; it is often easier to show that  $\mathcal{G}$  holds for all sets in a  $\pi$ -system.

**Definition 1.23.** A  $\pi$ -system is a collection of subsets of a set  $\Omega$  that is closed under finite intersections. A  $\lambda$ -system is a collection  $\Lambda$  of subsets of  $\Omega$  such that

- (i)  $\Omega \in \Lambda$ ;
- (ii)  $A \in \Lambda$  implies  $A^c \in \Lambda$ ; and
- (iii)  $A_n \in \Lambda$  and  $A_n \cap A_m = \emptyset$  for  $n \neq m$  implies  $\cup_{n=1}^{\infty} A_n \in \Lambda$ .

**Proposition 1.24.** (Dynkin's  $\pi$  -  $\lambda$  Lemma) Any  $\lambda$ -system that contains a  $\pi$ -system  $\Pi$  contains  $\sigma(\Pi)$ . In particular, the smallest  $\lambda$ -system that contains a  $\pi$ -system  $\Pi$  is  $\sigma(\Pi)$ .

The proof is another somewhat tedious exercise in set theory. See Billingsley sec. 1.3 for the gruesome details.

The  $\pi$  -  $\lambda$  lemma is often easier to work with than the monotone class theorem, because  $\pi$ -systems are often small and easily manageable: for instance, the set of all intervals  $(a, b)$  is a  $\pi$ -system; so is the set of all rectangles  $(a, b) \times (c, d)$  in  $\mathbb{R}^2$ . Moreover, it is often the case that if one can show that a property  $\mathcal{G}$  holds for a set  $A$  then it holds for  $A^c$ , and if it holds for pairwise disjoint sets  $A_1, A_2, \dots$  then it holds for  $\cup_{n=1}^{\infty} A_n$ ; thus the collection of sets for which  $\mathcal{G}$  holds is a  $\lambda$ -system, provided it holds for  $\Omega$ .

## 1.6 Carathéodory's Extension Theorem

There are two important existence theorems for measures, the *Carathéodory Extension Theorem* and the *Riesz Representation Theorem*. Neither is easy to prove. For the purposes of probability theory, the more useful of the two is Carathéodory's Theorem, and so for now I will only state this one.



**Theorem 1.25.** (*Carathéodory*) Any finite measure  $\mu$  on an algebra  $\mathcal{A}$  extends in a unique way to a measure on the  $\sigma$ -algebra  $\sigma(\mathcal{A})$  generated by  $\mathcal{A}$ .

See Billingsley sec. 1.3 for the proof. Observe that the existence of Lebesgue measure is an immediate consequence, by Lemma 1.17. A similar argument can be used to establish the existence of Lebesgue measure in  $d \geq 2$  dimensions; the key step is the following exercise.

**Exercise 1.26.** Define a  $d$ -dimensional *rectangle* to be a Cartesian product  $A_1 \times A_2 \times \cdots \times A_d$ , where each  $A_i$  is an *interval* of the real line. Let  $\mathcal{A}_d$  be the algebra generated by all rectangles. Prove that there is a unique measure  $\lambda = \lambda_d$  on  $\mathcal{A}_d$  such that for any rectangle,

$$\lambda(A_1 \times A_2 \times \cdots \times A_d) = \prod_{i=1}^d |A_i|,$$

where  $|A_i|$  is the length of the interval  $A_i$ .

## 1.7 Measurable Transformations

**Definition 1.27.** If  $\mathcal{F}$  is a  $\sigma$ -algebra on a set  $\mathcal{X}$  then the pair  $(\mathcal{X}, \mathcal{F})$  is called a *measurable space*. If  $(\mathcal{X}, \mathcal{F})$  and  $(\mathcal{Y}, \mathcal{G})$  are two measurable spaces and  $T : \mathcal{X} \rightarrow \mathcal{Y}$  is a function with domain  $\mathcal{X}$  and range  $\mathcal{Y}$  then  $T$  is said to be a *measurable transformation* if for every  $G \in \mathcal{G}$  the inverse image  $T^{-1}(G) \in \mathcal{F}$ . In the special case where  $\mathcal{Y} = \mathbb{R}$  and  $\mathcal{G} = \mathcal{B}_{\mathbb{R}}$ , the term *real random variable*, or simply *random variable*, is often used instead of *measurable transformation*. Similarly, when the range is  $\mathcal{Y} = \mathbb{R}^d$  and  $\mathcal{G} = \mathcal{B}_{\mathbb{R}^d}$ , the term *random vector* is often used.

Definition 1.27 should be somewhat reminiscent of the definition of a continuous function (inverse images of open sets are open). Thus, measurable transformations are in some sense the measure-theoretic analogue of continuous functions in topology. Like continuous functions, measurable transformations respect composition: if  $T : \mathcal{X} \rightarrow \mathcal{Y}$  and  $S : \mathcal{Y} \rightarrow \mathcal{Z}$  are measurable transformations (with respect to  $\sigma$ -algebras  $\mathcal{F}$ ,  $\mathcal{G}$ , and  $\mathcal{H}$ ) then the composition  $S \circ T : \mathcal{X} \rightarrow \mathcal{Z}$  is a measurable transformation.

To check that a given transformation  $T$  is measurable using Definition 1.27 directly is, in many situations, impractical. The next (very easy) lemma gives a way of proving measurability on the cheap.

**Lemma 1.28.** A transformation  $T : \mathcal{X} \rightarrow \mathcal{Y}$  is measurable if there is a collection  $\mathcal{K} \subset \mathcal{G}$  such that  $\mathcal{G} = \sigma(\mathcal{K})$  such that for every  $K \in \mathcal{K}$ ,

$$T^{-1}(K) \in \mathcal{F}.$$

**Example 1.29.** Let  $\mathcal{X} = [0, 1]$  and  $\mathcal{F} = \mathcal{B}_{[0,1]}$ , and take  $\mathcal{Y} = \{0, 1\}^{\mathbb{N}}$  with  $\mathcal{G}$  = the  $\sigma$ -algebra generated by the cylinder sets. Define  $T : [0, 1] \rightarrow \{0, 1\}^{\mathbb{N}}$  be the binary expansion map. (For those  $x \in [0, 1]$  that have two binary expansions, let  $T(x)$  be the one that ends in all 1s.) Then  $T$  is a measurable transformation, because for any cylinder set the inverse image is an interval.

**Example 1.30.** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be two metric spaces with Borel  $\sigma$ -algebras  $\mathcal{B}_{\mathcal{X}}$  and  $\mathcal{B}_{\mathcal{Y}}$ , respectively. If  $T : \mathcal{X} \rightarrow \mathcal{Y}$  is continuous, then it is measurable.

**Corollary 1.31.** Let  $X_1, X_2, \dots, X_m$  be random variables all defined on a measurable space  $(\Omega, \mathcal{F})$ . Then every linear combination  $\sum_{i=1}^m a_i X_i$  is also a random variable.

*Proof.* If  $X$  is a random variable and  $a \in \mathbb{R}$  is a scalar, then  $aX$  is a random variable, because it is the composition of  $X$  with the mapping  $M_a : \mathbb{R} \rightarrow \mathbb{R}$  defined by  $M_a(y) = ay$ . (This mapping is continuous, hence measurable.) Thus, to prove the corollary, it suffices to show that sums of random variables are random variables.

Suppose that  $X_1, X_2, \dots, X_m$  are random variables defined on  $(\Omega, \mathcal{F})$ , that is, that each  $X_i : \Omega \rightarrow \mathbb{R}$  is a measurable transformation. I claim that the mapping  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^m$  defined by  $\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \dots, X_m(\omega))$  is measurable. To see this, observe that by Lemma 1.28 it suffices to show that for any  $m$ -dimensional rectangle  $A = A_1 \times A_2 \times \dots \times A_m$ ,

$$\mathbf{X}^{-1}(A) \in \mathcal{F}.$$

But  $\mathbf{X}^{-1}(A) = \bigcap_{i=1}^m X_i^{-1}(A_i)$ ; since each  $X_i$  is measurable, each set  $X_i^{-1}(A_i)$  is in  $\mathcal{F}$ , and so the intersection is also in  $\mathcal{F}$ , as  $\sigma$ -algebras are closed under finite intersection. Thus,  $\mathbf{X}$  is measurable.

Now consider the mapping  $S : \mathbb{R}^m \rightarrow \mathbb{R}$  defined by addition of coordinates, i.e.,

$$S(x_1, x_2, \dots, x_m) = \sum_{i=1}^m x_i.$$

This mapping is continuous, and so it is measurable. Consequently,  $S(\mathbf{X}) : \Omega \rightarrow \mathbb{R}$  is measurable.  $\square$

The next proposition is nearly trivial, but it is of central importance, as it allows us to define the *distribution* of a random variable or random vector. It also shows that the existence of all sorts of interesting probability measures, on all sorts of different probability spaces, follows from the existence of Lebesgue measure. Thus, in most circumstances, the use of Caratheodory's theorem can be avoided.

**Proposition 1.32.** Let  $(\mathcal{X}, \mathcal{F})$  and  $(\mathcal{Y}, \mathcal{G})$  be two measurable spaces and let  $T : \mathcal{X} \rightarrow \mathcal{Y}$  be a measurable transformation. If  $P$  is a probability measure on  $\mathcal{F}$  then  $Q := P \circ T^{-1}$  is a probability measure on  $\mathcal{G}$ ; it is called the induced measure, or in some contexts the distribution of  $T$ .

*Proof.* Exercise. □

**Example 1.33.** Let  $T : (0, 1) \rightarrow (0, \infty)$  be defined by  $T(x) = -\log x$ , and let  $\lambda$  be the uniform distribution (Lebesgue measure) on (the Borel subsets of)  $(0, 1)$ . Then the induced measure  $Q = \lambda \circ T^{-1}$  is the *unit exponential distribution*:

$$Q(y, \infty) = e^{-y}.$$

**Example 1.34.** Let  $T : [0, 1] \rightarrow \{0, 1\}^{\mathbb{N}}$  be the measurable transformation of Example 1.29 and let  $P = \lambda$  be the uniform distribution on  $[0, 1]$ . Then the induced measure  $P \circ T^{-1}$  on sequence space is the *product Bernoulli- $\frac{1}{2}$*  measure, sometimes denoted  $P_{\frac{1}{2}}$ .

**Example 1.35.** The preceding construction can be generalized to *product Bernoulli- $p$*  measure, denoted  $P_p$ . Fix  $0 < p < 1$  and set  $q = 1 - p$ . Define random variables  $X_n : (0, 1) \rightarrow \{0, 1\}$  as follows:

$$\begin{aligned} X_1 &= \mathbf{1}_{[q, 1]}; \\ X_2 &= \mathbf{1}_{[q^2, q]} + \mathbf{1}_{[q+pq, 1]}; \\ &\text{etc.,} \end{aligned}$$

so that for any sequence  $e_1, e_2, \dots, e_n$  of 0s and 1s

$$\lambda\{X_i = e_i \text{ for each } 1 \leq i \leq n\} = p^{\sum_{i=1}^n e_i} q^{n - \sum_{i=1}^n e_i}. \quad (1.1)$$

Then define a mapping from the unit interval to sequence space  $T : (0, 1) \rightarrow \{0, 1\}^{\mathbb{N}}$  by  $T = (X_1, X_2, \dots)$ . The induced probability measure  $P_p = \lambda \circ T^{-1}$  is called *product Bernoulli- $p$*  measure, because equation (1.1) implies that the random variables  $X_1, X_2, \dots$  are independent, each with the Bernoulli- $p$  distribution.

## 1.8 Lebesgue-Stieltjes Measure

**Definition 1.36.** A *cumulative distribution function*, or *distribution function* for short, is a right-continuous, nondecreasing function  $F : \mathbb{R} \rightarrow \mathbb{R}$  such that

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

If  $(\Omega, \mathcal{F}, P)$  is a probability space and  $X$  is a real random variable defined on  $\Omega$  then the cumulative distribution function (c.d.f.) of  $X$  is the distribution function

$$F_X(x) = P\{X \leq x\}. \quad (1.2)$$

That  $F_X(x)$  is in fact a distribution function follows from the upper and lower continuity of probability measures (Property (d), sec. 1.3).

**Proposition 1.37.** *For any cumulative distribution function  $F$  on  $\mathbb{R}$  there is a random variable  $X$  defined on Lebesgue space (the unit interval with Lebesgue measure) with distribution function  $F_X = F$ . Consequently, there is a unique probability measure  $P_F$  on  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  (called the Lebesgue-Stieltjes measure with c.d.f.  $F$ ) such that for every interval  $(a, b]$ ,*

$$P_F((a, b]) = F(b) - F(a). \quad (1.3)$$

*Proof of Existence.* The construction is the natural generalization of that in Example 1.33 above. It uses a trick known as the *quantile transform*: the idea, as in Example 1.33, is to construct a monotone transformation  $T : [0, 1] \rightarrow \mathbb{R}$  in such a way that  $T$  moves each quantile of the uniform distribution to the corresponding quantile of the distribution function  $F$ . This is accomplished as follows:

$$T(\alpha) = \inf\{x : F(x) \geq \alpha\}.$$

It is easily checked (exercise!) that  $T$  is a non-decreasing function of  $\alpha$ , and therefore is measurable (why?). Hence, by Proposition 1.32, the distribution of  $T$  under  $\lambda$  is a probability measure on  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ , which we now denote by  $P_F := \lambda \circ T^{-1}$ . Finally, to prove (1.3), observe that for any  $b \in \mathbb{R}$ ,

$$P_F((-\infty, b]) = \lambda(T^{-1}((-\infty, b])) = F(b).$$

□

*Proof of Uniqueness.* The uniqueness of the measure  $P_F$  in Proposition 1.37 is a consequence of the uniqueness assertion in the Caratheodory Extension Theorem (see Proposition 1.21). In particular, the equation (1.3) implies that there is only one extension to the *algebra* of sets generated by the half-open intervals  $(a, b]$  (since this algebra consists of all finite unions of half-open intervals), and so Proposition 1.21 implies that there is only one extension to the  $\sigma$ -algebra generated by these sets. □