

Introductory Probability Theory

Yashaswi Mohanty

Contents

Part 1. Analysis	4
Chapter 1. Measures	5
1.1. Why is measurement hard?	5
1.2. Constructing the Lebesgue measure	6
1.3. Abstract measure spaces	16
Chapter 2. Measurable functions	17
Part 2. Probability	18

Part 1

Analysis

CHAPTER 1

Measures

1.1. Why is measurement hard?

On the real line \mathbb{R} , we may want our measure to satisfy some properties that are consistent with our intuitive notion of “length”. Formally, we want a function

$$\lambda : 2^{\mathbb{R}} \longrightarrow [0, \infty]$$

that satisfies

- (1) $\lambda(\emptyset) = 0$
- (2) $\lambda([a, b]) = b - a$ for $a \leq b \in \mathbb{R}$
- (3) Countable additivity: For a countable collection of pairwise-disjoint sets $\{A_i\}_{i \in \mathbb{N}} \subseteq \mathbb{R}$

$$(1) \quad \lambda\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \lambda(A_i)$$

- (4) Translation invariance: $\lambda(A + a) = \lambda(A)$ for any $a \in \mathbb{R}$ where $A + a := \{\alpha + a \mid \alpha \in A\}$.

Quite counterintuitively, it turns out that no such function exists! To prove this assertion, we need to construct some special kinds of sets that only exist if we assume the Axiom of Choice.

EXAMPLE 1.1. Define an equivalence relation \sim on $[0, 1]$ such that

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

Note that there are uncountably many classes in such a construction as the equivalence class for any given irrational number can contain at most countably many other irrational numbers. For example,

$$\left[\frac{\pi}{4}\right] = \left\{\frac{\pi}{4} + q \pmod{1} \mid q \in \mathbb{Q}\right\}.$$

Thus, using the Axiom of Choice, we can construct a set $E \subseteq [0, 1]$ such that E consists of exactly one “representative” from each equivalence class. Next, we can define

$$E_q := \{x + q \pmod{1} \mid x \in E\}$$

so that $\{E_q\}_{q \in \mathbb{Q}}$ is a partition of $[0, 1]$. To see that the sets are disjoint, suppose for contradiction that for any distinct $q, \tilde{q} \in \mathbb{Q} \cap [0, 1]$, $E_q \cap E_{\tilde{q}} \neq \emptyset$. If $x \in E_q \cap E_{\tilde{q}}$, then $x - q \in E$ and $x - \tilde{q} \in E$. But they clearly belong to the same equivalence class and this is a contradiction given our construction of E . To see that the union of these sets is $[0, 1]$, consider an arbitrary $y \in [0, 1]$ and observe that since our equivalence relation \sim partitions $[0, 1]$, $y \in [x]$ for some $x \in E$. Then $q^* = y - x \in \mathbb{Q}$ and so $y = x + q^* \pmod{1} \in E_{q^*}$. Thus we have that

$$[0, 1] \subseteq \bigcup_{q \in \mathbb{Q}} E_q.$$

Since the reverse inclusion follows by the definition of E_q , we have that $\{E_q\}_{q \in \mathbb{Q}}$ is a partition of $[0, 1]$.

PROPOSITION 1.1. *There exists no function $\lambda : 2^{\mathbb{R}} \longrightarrow [0, \infty]$ that satisfies properties (1)-(4) described above*

PROOF. Suppose, for contradiction, that such a function λ exists. We can define the collection of sets $\{E\}_{q \in \mathbb{Q}}$ as in Example 1.1 and observe that

$$\begin{aligned} 1 = \lambda([0, 1]) &= \lambda \left[\bigcup_{q \in \mathbb{Q}} E_q \right] \\ &= \sum_{q \in \mathbb{Q}} \lambda[E_q] \\ &= \sum_{q \in \mathbb{Q}} c \end{aligned}$$

where the first equality follows from property (2), the second equality follows from the fact that $\{E\}_{q \in \mathbb{Q}}$ is a partition of $[0, 1]$, the third equality is due to property (3). The last equality follows as a consequence of translation invariance (property (4)). Since $c \in [0, 1]$

$$\sum_{q \in \mathbb{Q}} c = 0 \text{ or } \infty \neq 1$$

which is a contradiction. Thus no such function λ exists. \square

This particular example of a *non-measurable* set is called a *Vitali set*. While we used the interval $[0, 1]$ to construct such a set, it turns out that this construction can be extended to any set of positive length in the Lebesgue sense.

1.2. Constructing the Lebesgue measure

The key issue with our previous definition of a measure on \mathbb{R} is that one cannot have a set-valued function that both has our four desired properties *and* is defined on all subsets of the real line. As a convention, the canonical construction of a measure retains the desired properties in exchange for restricting the class of subsets on which the measure is defined. These subsets are called *measurable* and the standard construction of the Lebesgue measure leads to the class of measurable subsets on the real line to have a special structure of a σ -algebra. Before we define this structure it might be worthwhile looking at various types of structures a class of sets could have

1.2.1. Structures of sets. In the rest of this chapter, we assume that (\mathcal{X}, τ) is an abstract topological space.

DEFINITION 1.1. Let $\mathcal{F} \subseteq 2^{\mathcal{X}}$. We call \mathcal{F} a *ring* if

- (i) $\emptyset \in \mathcal{F}$
- (ii) $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$
- (iii) $A, B \in \mathcal{F} \Rightarrow A \setminus B \in \mathcal{F}$.

Note that the above definition implies that $A \cap B = A \setminus (A \setminus B) \in \mathcal{F}$.

DEFINITION 1.2. Let $\mathcal{F} \subseteq 2^{\mathcal{X}}$. We call \mathcal{F} an *algebra* if

- (i) \mathcal{F} is a ring
- (ii) $\mathcal{X} \in \mathcal{F}$.

For example, if we let \mathcal{X} be an arbitrary infinite set, the collection of all finite subsets of \mathcal{X} forms a ring but not an algebra.

DEFINITION 1.3. Let $\mathcal{F} \subseteq 2^{\mathcal{X}}$. We call \mathcal{F} a σ -ring if

- (i) \mathcal{F} is a ring
- (ii) \mathcal{F} is closed under countable unions.

DEFINITION 1.4. Let $\mathcal{F} \subseteq 2^{\mathcal{X}}$. We call \mathcal{F} a σ -algebra if

- (i) \mathcal{F} is an algebra.
- (ii) \mathcal{F} is closed under countable unions.

Naturally, the power set $2^{\mathcal{X}}$ is a ring, algebra, σ -ring, and σ -algebra all rolled into one.

REMARK. Algebras are sometimes referred to as *fields* in the probability literature.

As we said earlier, the notion of a σ -algebra is important because the standard Lebesgue measurable sets form a σ -algebra of subsets of \mathbb{R} . However, the other structures we have defined are also important; as we “extend” the notion of the length of an interval on the real line to more complicated sets, we shall first expand our class of measurable sets to a ring of sets.

1.2.2. Lengths of intervals. The most intuitive notion of a measure on \mathbb{R} arises from the length of an interval. Thus, in our construction of the Lebesgue measure, we start with the simplest class of sets which consists of intervals in \mathbb{R} . Define $\mathcal{L} = \{(a, b] \mid -\infty < a \leq b < \infty\}$ and let $\lambda_1 : \mathcal{L} \rightarrow [0, \infty]$ be given by $\lambda_1((a, b]) = b - a$. It turns out that our collection of half-open intervals in \mathbb{R} has the structure of a *semi-ring*.

DEFINITION 1.5. Let $\mathcal{F} \subseteq 2^{\mathcal{X}}$. We call \mathcal{F} a *semi-ring* if

- (i) $\emptyset \in \mathcal{F}$.
- (ii) $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$.
- (iii) $A, B \in \mathcal{F} \Rightarrow \exists \{A_i\}_{i=1}^n \in \mathcal{F}$ such that

$$A \setminus B = \bigcup_{i=1}^n A_i$$

PROPOSITION 1.2. \mathcal{L} is a semi-ring.

PROOF. To see (i), note that $\emptyset = (a, a] \in \mathcal{L}$. For (ii), note that for any intervals $A = (a_1, b_1]$ and $B = (a_2, b_2]$,¹

$$(a_1, b_1] \cap (a_2, b_2] = (\max(a_1, a_2), \min(b_1, b_2)] \in \mathcal{L}.$$

To see (iii), we have to consider two possible cases. First, if A, B are disjoint then $A \setminus B = A \in \mathcal{L}$. If A, B have a non-trivial intersection, then

$$\begin{aligned} A \setminus B &= A \cap B^C = (a_1, b_1] \cap \{(-\infty, a_2] \cup (b_2, \infty)\} \\ &= (a_1, b_1] \cap (-\infty, a_2] \bigcup (a_1, b_1] \cap (b_2, \infty) \\ &= (a_1, \min(b_1, a_2)] \bigcup (\max(a_1, b_2), b_1] \end{aligned}$$

where the components of the union expressed in the last equality are in \mathcal{L} . This completes the proof. \square

The fact that \mathcal{L} is a semi-ring is important because there's a relatively straightforward way to “expand” a semi-ring into a ring.

THEOREM 1.1. Let \mathcal{F} be a semi-ring and let \mathcal{B} be the set of all finite disjoint unions of sets in \mathcal{F} . Then \mathcal{B} is a ring.

PROOF. Property (i) in Definition 1.1 is trivially satisfied thus we need to prove properties (ii) and (iii). Let $A, B \in \mathcal{B}$. To prove property (iii), we first establish the weaker claim that $A \cap B \in \mathcal{B}$. Observe that

$$\begin{aligned} A &= \bigcup_{i=1}^{n_A} A_i, \quad A_i \in \mathcal{F}, A_i \cap A_j = \emptyset \text{ for } i \neq j, \\ B &= \bigcup_{i=1}^{n_B} B_i, \quad B_i \in \mathcal{F}, B_i \cap B_j = \emptyset \text{ for } i \neq j, \end{aligned}$$

¹If $\max(a_1, a_2) > \min(b_1, b_2)$, then $(\max(a_1, a_2), \min(b_1, b_2)] = \emptyset \in \mathcal{L}$

by the definition of \mathcal{B} . Then

$$\begin{aligned} A \cap B &= \left(\bigcup_{i=1}^{n_A} A_i \right) \cap \left(\bigcup_{j=1}^{n_B} B_j \right) \\ &= \bigcup_{i=1}^{n_A} \bigcup_{j=1}^{n_B} (A_i \cap B_j) \end{aligned}$$

where $\forall i, j : A_i \cap B_j \in \mathcal{F}$ as \mathcal{F} is a semi-ring. Clearly, $A_i \cap B_j$ is disjoint from $A_{i'} \cap B_{j'}$, thus proving the claim. Next, we establish property (iii) by noting that

$$\begin{aligned} A \setminus B &= \left(\bigcup_{i=1}^{n_A} A_i \right) \setminus B \\ &= \left(\bigcup_{i=1}^{n_A} A_i \right) \cap B^C \\ &= \bigcup_{i=1}^{n_A} (A_i \cap B^C) \\ &= \bigcup_{i=1}^{n_A} \left(A_i \cap \left(\bigcap_{j=1}^{n_B} B_j^C \right) \right) \\ &= \bigcup_{i=1}^{n_A} \bigcap_{j=1}^{n_B} (A_i \cap B_j^C) \\ &= \bigcup_{i=1}^{n_A} \bigcap_{j=1}^{n_B} A_i \setminus B_j \end{aligned}$$

where the $A_i \setminus B_j \in \mathcal{B}$ since $A_i, B_j \in \mathcal{F}$. By the closure under finite intersections property established earlier, $E_i = \bigcap_{j=1}^{n_B} A_i \setminus B_j \in \mathcal{B}$ for any $1 \leq i \leq n_A$. Thus we can rewrite the chain of equalities above as

$$A \setminus B = \bigcup_{i=1}^{n_A} E_i$$

where $E_i \cap E_{i'} = \emptyset$ because $A_i \cap A_{i'} = \emptyset$. Since the finite disjoint union of elements of \mathcal{B} is also a finite disjoint union of elements of \mathcal{F} , our claim follows. Finally, to establish property (ii), observe that

$$A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A)$$

which is a disjoint union of elements in \mathcal{B} and so is also in \mathcal{B} by the same argument as earlier. \square

COROLLARY 1.1. *Let \mathcal{J} be the set of all finite disjoint unions of sets in \mathcal{L} . Then \mathcal{J} is a ring.*

PROOF. By Proposition 1.2, \mathcal{L} is a semi-ring. The claim then follows by an application of Theorem 1.1. \square

Now we can extend our proto-measure λ_1 to a new proto-measure $\lambda_2 : \mathcal{J} \rightarrow [0, \infty]$ as follows:

$$\lambda_2(A) := \begin{cases} \lambda_1(A), & A \in \mathcal{L} \\ \sum_{i=1}^n \lambda_1(B_i), & A = \bigcup_{i=1}^n B_i, \{B_i\}_{i=1}^n \text{ are disjoint in } \mathcal{L} \end{cases}$$

PROPOSITION 1.3. *λ_2 is finitely additive on \mathcal{J} . That is, for any finite disjoint collection of sets $\{A_i\}_{i=1}^n \in \mathcal{J}$*

$$\lambda_2 \left(\bigcup_{i=1}^n A_i \right) = \sum_{i=1}^n \lambda_2(A_i).$$

PROOF. For clarity, we will prove finite additivity for two sets, since the general case follows by induction. Let $A, B \in \mathcal{J}$ such that $A \cap B = \emptyset$. By definition,

$$\begin{aligned} A &= \bigcup_{i=1}^{n_A} A_i, \{A_i\}_{i=1}^{n_A} \text{ are disjoint in } \mathcal{L} \\ B &= \bigcup_{i=1}^{n_B} B_i, \{B_i\}_{i=1}^{n_B} \text{ are disjoint in } \mathcal{L} \end{aligned}$$

and so we have that

$$\begin{aligned} \lambda_2(A \cup B) &= \lambda_2\left(\left(\bigcup_{i=1}^{n_A} A_i\right) \cup \left(\bigcup_{i=1}^{n_B} B_i\right)\right) \\ &= \sum_{i=1}^{n_A} \lambda_2(A_i) + \sum_{i=1}^{n_B} \lambda_2(B_i) \\ &= \lambda_2(A) + \lambda_2(B) \end{aligned}$$

where the second equality follows from associativity of addition along with the fact that $(\bigcup_{i=1}^{n_A} A_i) \cup (\bigcup_{i=1}^{n_B} B_i)$ is a disjoint union of sets in \mathcal{L} . \square

1.2.3. Structures generated by a class of sets. A key way to “expand” a particular class of sets into a larger structure is to look at the structure *generated* by the class of sets. This idea can be formalized in the following definition, which serves as particular example of this general concept of generation.

DEFINITION 1.6. For any $\mathcal{A} \subseteq 2^{\mathcal{X}}$, we refer to the intersection of all rings that contain \mathcal{A} as the ring *generated* by \mathcal{A} . Formally, we write

$$\text{ring}(\mathcal{A}) = \bigcap \{\mathcal{R} \subseteq 2^{\mathcal{X}} \text{ is a ring} \mid \mathcal{A} \subseteq \mathcal{R}\}.$$

PROPOSITION 1.4. For any $\mathcal{A} \subseteq 2^{\mathcal{X}}$, $\text{ring}(\mathcal{A})$ is a ring.

PROOF. First note that $\text{ring}(\mathcal{A})$ exists since $2^{\mathcal{X}}$ is a ring and so $\{\mathcal{R} \subseteq 2^{\mathcal{X}} \text{ is a ring} \mid \mathcal{A} \subseteq \mathcal{R}\}$ is non-empty. Next observe that $\emptyset \in \text{ring}(\mathcal{A})$ vacuously, so property (i) in Definition 1.1 is easily satisfied. For property (ii), let $A, B \in \text{ring}(\mathcal{A})$ and observe that $A, B \in \mathcal{R}$ for every $\mathcal{R} \in \{\mathcal{R} \subseteq 2^{\mathcal{X}} \text{ is a ring} \mid \mathcal{A} \subseteq \mathcal{R}\}$. Since \mathcal{R} is a ring, $A \cup B \in \mathcal{R}$ for every \mathcal{R} and thus $A \cup B$ is in the intersection i.e. $\text{ring}(\mathcal{A})$. A similar argument establishes property (iii) and thus we can conclude that $\text{ring}(\mathcal{A})$ is a ring (as it should, given its name). \square

In our construction of the Lebesgue measure on \mathbb{R} , we discovered that \mathcal{J} , which is the set of all disjoint unions of half-open intervals in \mathbb{R} , is a ring. It turns out that we can make a stronger statement using the language of generators developed here.

PROPOSITION 1.5. $\mathcal{J} = \text{ring}(\mathcal{L})$

PROOF. Let $A \in \mathcal{J}$ be arbitrary. Then we can write

$$A = \bigcup_{i=1}^{n_A} A_i$$

where $A_i \in \mathcal{L}$ are pairwise disjoint. Let \mathcal{R} be an arbitrary ring that contains \mathcal{L} and observe that since rings are closed under finite unions, $A \in \mathcal{R}$. Since \mathcal{R} was arbitrary, A is contained by every ring that contains \mathcal{L} and is thus contained in the intersection of all such rings i.e. $\text{ring}(\mathcal{L})$. This proves that $\mathcal{J} \subseteq \text{ring}(\mathcal{L})$.

To see reverse inclusion, recall that \mathcal{J} is a ring that contains \mathcal{L} , and so the intersection of all rings that contain \mathcal{L} is certainly contained in \mathcal{J} . This completes the proof. \square

In measure theory, the most important structure on sets is the σ -algebra, and the σ -algebra generated by a class of sets \mathcal{A} , defined analogously to Definition 1.6 about rings and notated as $\sigma(\mathcal{A})$, plays in an important role in this theory. Using a similar argument as the one shown earlier, one can conclude that $\sigma(\mathcal{A})$ is indeed a σ -algebra. In analysis and probability theory, mathematicians are interested in σ -algebras generated by a special class of sets.

DEFINITION 1.7. The σ -algebra generated by the topology τ on set \mathcal{X} is called the *Borel σ -algebra* on \mathcal{X} and is denoted $\mathcal{B}(\mathcal{X})$.

The Borel σ -algebra is interesting because it turns that it is the σ -algebra generated by \mathcal{L} is indeed $\mathcal{B}(\mathbb{R})$, where \mathbb{R} has the usual topology. To prove this fact, we need a little lemma from an introductory course on analysis and topology.

LEMMA 1.1. *Any open set in the usual topology of \mathbb{R} can be written as a countable disjoint union of open intervals in \mathbb{R} .*

PROOF. Let O be an open set in \mathbb{R} and let $x \in O$ be arbitrary. Define $I_x \subseteq O$ to be the largest open interval that contains x (that is, I_x is the union of all open intervals in O that contain x). Note that at least one such interval exists because O is open and so there exists some $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \subseteq O$. Now for any distinct $x, y \in O$, I_x and I_y are either disjoint or equal since if they were neither, $I_x \cup I_y \subseteq O$ would be a larger interval that contains both x and y . Let \mathcal{I} denote the collection of all disjoint such intervals (that is, we get \mathcal{I} by discarding all the “redundant” intervals in $\{I_x\}_{x \in O}$). We can do this without invoking the Axiom of Choice since there are only countably many intervals in \mathcal{I} : every interval $I \in \mathcal{I}$ contains at least one rational number because the rationals are a countably dense subset of \mathbb{R} . Thus, since the intervals are disjoint, \mathcal{I} can have at most countably many intervals. Of course

$$O = \bigcup_{I \in \mathcal{I}} I$$

and so our claim follows. □

PROPOSITION 1.6. $\sigma(\mathcal{L}) = \mathcal{B}(\mathbb{R})$

PROOF. Let O be an open set in \mathbb{R} . Then, by Lemma 1.1

$$\begin{aligned} O &= \bigcup_{i=1}^{\infty} (a_i, b_i) \\ &= \bigcup_{i=1}^{\infty} \bigcup_{n=1}^{\infty} \left(a_i, b_i - \frac{1}{n} \right] \end{aligned}$$

which is in $\sigma(\mathcal{L})$ by closure under countable unions (property (ii) in Definition 1.4). Therefore the topology of \mathbb{R} is in $\sigma(\mathcal{L})$ which implies that $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{L})$. The

To see the reverse inclusion, observe that for any $(a, b] \in \mathcal{L}$, we can write

$$(a, b] = (a, b) \cup \{b\} \in \mathcal{B}(\mathbb{R})$$

since $\{b\}$ is closed in \mathbb{R} and closed sets are the complements of open sets and thus contained in $\mathcal{B}(\mathbb{R})$.² Therefore $\mathcal{L} \subseteq \mathcal{B}(\mathbb{R})$ and so $\sigma(\mathcal{L}) \subseteq \mathcal{B}(\mathbb{R})$, completing the proof. □

Now we are ready to prove that our proto-measure λ_2 is actually a countably-additive pre-measure on ring (\mathcal{L}) .

PROPOSITION 1.7. λ_2 is a countably additive pre-measure on ring (\mathcal{L}) , that is to say,

$$(i) \quad \lambda_2(\emptyset) = 0$$

² σ -algebras on \mathcal{X} are closed under complements since they are closed under set-differences and contain \mathcal{X} .

(ii) For disjoint $\{A_i\}_{i=1}^\infty \in \mathcal{J}$ such that $\bigcup_{i=1}^\infty A_i \in \mathcal{J}$

$$\lambda_2 \left(\bigcup_{i=1}^\infty A_i \right) = \sum_{i=1}^\infty \lambda_2 (A_i).$$

PROOF. Property (i) is inherited from λ_1 . To see property (ii), let $\{A_i\}_{i=1}^\infty \in \mathcal{J}$ be disjoint and write $A := \bigcup_{i=1}^\infty A_i$. First, note that if $\lambda_2(A_i) = \infty$ for any $i \in \mathbb{N}$, then $\infty = \lambda_2(A_i) \leq \lambda_2(\bigcup_{i=1}^\infty A_i) = \infty$ where the inequality is due to the monotonicity³ of λ_2 . Thus, in this case, the claim follows vacuously. So, without loss of generality, we can assume that $\lambda_2(A_i) < \infty$ for every $i \in \mathbb{N}$. First, note that for any $n \in \mathbb{N}$, $\bigcup_{i=1}^n A_i \subseteq A$ and so, by the monotonicity and finite additivity of λ_2 , we have that

$$\lambda_2(A) \geq \sum_{i=1}^n \lambda_2(A_i)$$

for every $n \in \mathbb{N}$. Taking limits, we have countable superadditivity:

$$\lambda_2(A) \geq \sum_{i=1}^\infty \lambda_2(A_i).$$

In order to deduce the reverse inequality, **TODO: Figure out proof of the reverse inequality** □

1.2.4. Outer measures.

DEFINITION 1.8. A set valued function

$$\mu^* : 2^{\mathcal{X}} \longrightarrow [0, \infty]$$

is called an outer measure on \mathcal{X} if

- (i) $\mu^*(\emptyset) = 0$
- (ii) $A \subseteq B \in 2^{\mathcal{X}} \implies \mu^*(A) \leq \mu^*(B)$
- (iii) For $\{A_i\}_{i=1}^\infty \in 2^{\mathcal{X}}$

$$\mu^* \left(\bigcup_{i=1}^\infty A_i \right) \leq \sum_{i=1}^\infty \mu^*(A_i).$$

EXAMPLE 1.2. Given a countably additive pre-measure μ on a ring $\mathcal{A} \subseteq 2^{\mathcal{X}}$, define for any $E \subseteq \mathcal{X}$

$$\mu^*(E) := \inf \left\{ \sum_{i=1}^\infty \mu(A_i) \mid A_i \in \mathcal{A}, E \subseteq \bigcup_{i=1}^\infty A_i \right\}$$

Note that this function is defined on $2^{\mathcal{X}}$ since every bounded below subset of the (extended) real numbers has an infimum.

In order to prove that the function in our example is indeed an outer measure, the following lemma is useful.

LEMMA 1.2 (Tonelli for series). *Let $\{x_{ij}\}_{i,j \in \mathbb{N} \times \mathbb{N}}$ be a sequence of non-negative (extended) real numbers. Then*

$$\sum_{i,j \in \mathbb{N}^2} x_{ij} = \sum_{i=1}^\infty \sum_{j=1}^\infty x_{ij} = \sum_{j=1}^\infty \sum_{i=1}^\infty x_{ij}.$$

PROOF. We will prove the first equality since the second then follows by symmetry. Let $F \subset \mathbb{N}^2$ be arbitrary and finite. Then, there exists some $N \in \mathbb{N}$ such that $F \subseteq \{1, 2, \dots, N\}^2$ and so, by the non-negativity of x_{ij}

$$\sum_{i,j \in F} x_{ij} \leq \sum_{i,j \in \{1, 2, \dots, N\}^2} x_{ij} = \sum_{i=1}^N \sum_{j=1}^N x_{ij} \leq \sum_{i=1}^\infty \sum_{j=1}^\infty x_{ij}.$$

³For any $A, B \in \text{ring}(\mathcal{L})$ such that $A \subseteq B$, $\lambda_2(B) = \lambda_2(A) + \lambda_2(B \setminus A) \geq \lambda_2(A)$

This inequality holds for any finite $F \subset \mathbb{N}^2$ and so it holds for the supremum of all such finite sums. That is to say,

$$\sup_{F \subset \mathbb{N}^2 | F \text{ is finite}} \sum_{i,j \in F} x_{ij} \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_{ij}.$$

But recall that for any $\{a_i\}_{i \in \mathcal{I}} \in [0, \infty]$ where \mathcal{I} is any index set

$$\sum_{i \in \mathcal{I}} a_i := \sup_{I \subset \mathcal{I} | I \text{ is finite}} \sum_{i \in I} a_i,$$

and so we have that

$$\sum_{i,j \in \mathbb{N}^2} x_{ij} \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_{ij}.$$

To derive the other inequality, observe that it is sufficient to prove that

$$\sum_{i,j \in \mathbb{N}^2} x_{ij} \geq \sum_{i=1}^I \sum_{j=1}^{\infty} x_{ij}$$

for every $I \in \mathbb{N}$. Fix $I = I_0$ and note that

$$\sum_{i=1}^{I_0} \sum_{j=1}^{\infty} x_{ij} = \sum_{i=1}^{I_0} \lim_{J \rightarrow \infty} \sum_{j=1}^J x_{ij} = \lim_{J \rightarrow \infty} \sum_{i=1}^{I_0} \sum_{j=1}^J x_{ij}.$$

Thus to prove $\sum_{i,j \in \mathbb{N}^2} x_{ij} \geq \sum_{i=1}^{I_0} \sum_{j=1}^{\infty} x_{ij}$ we need to prove that

$$\sum_{i,j \in \mathbb{N}^2} x_{ij} \geq \sum_{i=1}^{I_0} \sum_{j=1}^J x_{ij}$$

for every $J \in \mathbb{N}$. Fix $J = J_0$ and then observe that

$$\sum_{i=1}^{I_0} \sum_{j=1}^{J_0} x_{ij} = \sum_{i,j \in \{1,2,\dots,I_0\} \times \{1,2,\dots,J_0\}} x_{ij} \leq \sum_{i,j \in \mathbb{N}^2} x_{ij}$$

where the inequality follows due to non-negativity of x_{ij} . This concludes the proof. \square

REMARK. This lemma is a special case of Tonelli's theorem, a fundamental theorem that allows us to construct measures on Cartesian products of measure spaces from the measures on those spaces themselves. This theorem will be motivated and proved in Chapter 5.

PROPOSITION 1.8. *The function $\mu^* : 2^{\mathcal{X}} \rightarrow [0, \infty]$ defined in Example 1.2 is an outer measure*

PROOF. For (i), observe that $\emptyset \in \mathcal{A}$ and so $\mu^*(\emptyset) = \mu(\emptyset) = 0$. Next, let $A \subseteq B \subseteq \mathcal{X}$ and observe that

$$\left\{ \sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, B \subseteq \bigcup_{i=1}^{\infty} A_i \right\} \subseteq \left\{ \sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, A \subseteq \bigcup_{i=1}^{\infty} A_i \right\}$$

and so

$$\mu^*(B) = \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, B \subseteq \bigcup_{i=1}^{\infty} A_i \right\} \geq \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, A \subseteq \bigcup_{i=1}^{\infty} A_i \right\} = \mu^*(A)$$

which gives us (ii). For (iii), let $\{E_i\}_{i=1}^{\infty} \in 2^{\mathcal{X}}$ be disjoint and assume that $\sum_{i=1}^{\infty} \mu^*(E_i) < \infty$ since otherwise the claim is trivial. Fix $\epsilon > 0$ and choose $A_{ij} \in \mathcal{A}$ such that $E_i \subseteq \bigcup_{j=1}^{\infty} A_{ij}$ and

$$\mu^*(E_i) \leq \sum_{j=1}^{\infty} \mu(A_{ij}) < \mu^*(E_i) + \frac{\epsilon}{2^i}$$

for every $i \in \mathbb{N}$ ⁴. Observe that

$$E := \bigcup_{i=1}^{\infty} E_i \subseteq \bigcup_{i=1}^{\infty} \bigcup_{j=1}^{\infty} A_{ij}$$

and so

$$\mu^*(E) \leq \sum_{i,j \in \mathbb{N}^2} \mu(A_{ij}) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \mu(A_{ij}) \leq \sum_{i=1}^{\infty} \mu^*(E_i) + \epsilon$$

where the equality follows by Lemma 1.2 and the second inequality is due to properties of the geometric series. Since ϵ was arbitrary, the claim follows. \square

PROPOSITION 1.9. *Let \mathcal{A} , μ , and μ^* be defined as in Example 1.2. Then, for any $A \in \mathcal{A}$*

$$\mu^*(A) = \mu(A).$$

PROOF. First, observe that A is a cover for itself and that $\emptyset \in \mathcal{A}$ and so

$$\mu^*(A) = \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, A \subseteq \bigcup_{i=1}^{\infty} A_i \right\} \leq \sum_{i=1}^{\infty} \mu(A_i)$$

where $A_1 = A$ and $A_i = \emptyset$ for $i \neq 1$. Therefore,

$$\mu^*(A) \leq \mu(A).$$

To see the reverse inequality, let $\{A_i\}_{i \in \mathbb{N}} \in \mathcal{A}$ be an arbitrary cover of A . Define,

$$B_i := A \cap \left(A_i \setminus \bigcup_{j=1}^{i-1} A_j \right)$$

and notice that the $\{B_i\}$ is a pairwise disjoint collections whose union is A such that $B_i \subseteq A_i$ for every $i \in \mathbb{N}$. By countable additivity and monotonicity,

$$\mu(A) = \sum_{i=1}^{\infty} \mu(B_i) \leq \sum_{i=1}^{\infty} \mu(A_i).$$

Since $\{A_i\} \subseteq \mathcal{A}$ is an arbitrary cover of A , we have that

$$\mu(A) \leq \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, A \subseteq \bigcup_{i=1}^{\infty} A_i \right\} = \mu^*(A)$$

which completes the proof. \square

Now we are (finally!!) ready to extend our pre-measure to a bona-fide measure on a σ -algebra, using the following theorem.

THEOREM 1.2 (Caratheodory's Extension Theorem). *Given a countably-additive pre-measure μ on ring $\mathcal{A} \subseteq 2^{\mathcal{X}}$ with outer measure μ^* as in Example 1.2, define the set*

$$\mathcal{C} = \{A \subseteq \mathcal{X} \text{ such that } \mu^*(E) = \mu^*(A \cap E) + \mu^*(A^C \cap E) \forall E \in 2^{\mathcal{X}}\}.$$

Then

- (i) $\mathcal{A} \subseteq \mathcal{C}$.
- (ii) \mathcal{C} is a σ -algebra.
- (iii) $\mu^*|_{\mathcal{C}}$ is a countably additive measure on \mathcal{C} .

⁴This is possible due to the assumption that $\mu^*(E_i) < \infty$, which implies that the set $\left\{ \sum_{j=1}^{\infty} \mu(A_{ij}) \mid A_{ij} \in \mathcal{A}, E_i \subseteq \bigcup_{j=1}^{\infty} A_{ij} \right\}$ is non-empty. The definition of an infimum then implies that such a cover $\{A_{ij}\}$ exists.

PROOF. First we will show (i). Let $A \in \mathcal{A}$ be arbitrary. By the countable subadditivity of μ^* , we know that

$$\mu^*(E) = \mu^*\left((A \cap E) \cup (A^C \cap E)\right) \leq \mu^*(A \cap E) + \mu^*(A^C \cap E)$$

for every $E \subseteq \mathcal{X}$. To deduce the reverse inequality, fix E such that $\mu^*(E) < \infty$ because otherwise the claim follows trivially. Pick an $\epsilon > 0$ and find a cover $\{A_i\}_{i=1}^\infty \in \mathcal{A}$ of E such that

$$\mu^*(E) \leq \sum_{i=1}^\infty \mu(A_i) < \mu^*(E) + \epsilon$$

As in the proof of Proposition 1.8, this is possible because $\mu^*(E) < \infty$ and the definition of an infimum. Next, observe that

$$\begin{aligned} E \cap A &\subseteq \bigcup_{i=1}^\infty (A_i \cap A), \\ E \cap A^C &\subseteq \bigcup_{i=1}^\infty (A_i \cap A^C) \end{aligned}$$

and so

$$\begin{aligned} \mu^*(E \cap A) &\leq \mu^*\left(\bigcup_{i=1}^\infty (A_i \cap A)\right) \leq \sum_{i=1}^\infty \mu^*(A_i \cap A) \\ \mu^*(E \cap A^C) &\leq \mu^*\left(\bigcup_{i=1}^\infty (A_i \cap A^C)\right) \leq \sum_{i=1}^\infty \mu^*(A_i \cap A^C) \end{aligned}$$

where the first inequality follows due to monotonicity and the second due to subadditivity. Together, these inequalities imply that

$$\begin{aligned} \mu^*(A \cap E) + \mu^*(A^C \cap E) &\leq \sum_{i=1}^\infty \mu^*(A_i \cap A) + \mu^*(A_i \cap A^C) \\ &= \sum_{i=1}^\infty \mu(A_i \cap A) + \mu(A_i \cap A^C) \\ &= \sum_{i=1}^\infty \mu(A_i) \\ &< \mu^*(E) + \epsilon \end{aligned}$$

where the first equality is due to the fact that rings are closed under intersections and set-differences along with Proposition 1.9 and the second equality is due to the countable additivity of μ . Since ϵ and E are arbitrary, we have that

$$\mu^*(A \cap E) + \mu^*(A^C \cap E) \geq \mu^*(E)$$

for every $E \subseteq \mathcal{X}$, establishing that $\mathcal{A} \subseteq \mathcal{C}$.

Next we show (ii); that is, we prove \mathcal{C} is a σ -algebra. Recall Definition 1.4 and notice that it is sufficient to prove that (1) $\emptyset, \mathcal{X} \in \mathcal{C}$; (2) if $A \in \mathcal{C}$ then $A^C \in \mathcal{C}$; (3) if $\{A_i\}_{i=1}^\infty \in \mathcal{C}$ then $\bigcup_{i=1}^\infty A_i \in \mathcal{C}$. Note that $\emptyset, \mathcal{X} \in \mathcal{C}$ because, trivially,

$$\mu^*(E \cap \mathcal{X}) + \mu^*(E \cap \emptyset) = \mu^*(E).$$

Symmetry between A and A^C in the definition of \mathcal{C} establishes (2). For (3), we first establish closure under finite unions and bootstrap this weaker result to yield the stronger claim. Let $A, B \in \mathcal{C}$ and let $E \subseteq \mathcal{X}$ be arbitrary. Then

$$\begin{aligned}
\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^C) \\
&= \mu^*((E \cap A) \cap B) + \mu^*((E \cap A) \cap B^C) + \mu^*(E \cap A^C) \\
&= \mu^*(E \cap A \cap B) + \mu^*(E \cap (A \cap B)^C \cap A) + \mu^*(E \cap (A \cap B)^C \cap A^C) \\
&= \mu^*(E \cap A \cap B) + \mu^*(E \cap (A \cap B)^C)
\end{aligned}$$

where the first equality is due to the definition of \mathcal{C} and the fact that $B \in \mathcal{C}$, the second equality is due to the identities

$$\begin{aligned}
(A \cap B)^C \cap A &= (A^C \cup B^C) \cap A = A \cap B^C \\
(A \cap B)^C \cap A^C &= (A^C \cup B^C) \cap A^C = A^C,
\end{aligned}$$

and the third equality follows from the definition of \mathcal{C} and that $A \in \mathcal{C}$. This proves that for any $A, B \in \mathcal{C}$, $A \cap B \in \mathcal{C}$. Property (2) then implies that $A \cup B \in \mathcal{C}$.

To establish closure under countable unions, fix $E \subseteq \mathcal{X}$ and let $\{A_i\}_{i=1}^\infty \in \mathcal{C}$ be arbitrary with $B = \bigcup_{i \in \mathbb{N}} A_i$ and define

$$B_n := \bigcup_{i=1}^n A_i$$

where $B_n \in \mathcal{C}$ by our result on closure under finite unions. Without loss of generality, we can assume that the $\{A_i\}$ are pairwise disjoint (since we could otherwise replace A_i with $C_i := A_i \setminus \bigcup_{j=1}^{i-1} A_j$ which are disjoint such that $\bigcup_{i=1}^\infty A_i = \bigcup_{i=1}^\infty C_i$). Then, we have that

$$\begin{aligned}
\mu^*(E) &= \mu^*(E \cap B_n^C) + \mu^*(E \cap B_n) \\
&= \mu^*(E \cap B_n^C) + \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^C) \\
&= \mu^*(E \cap B_n^C) + \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1})
\end{aligned}$$

where we used the fact that $A_n \in \mathcal{C}$ for the second equality and the disjointness of A_i for the third equality. Observe that the equality $\mu^*(E \cap B_n) = \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1})$ is a recurrence relation that can be expanded as

$$\mu^*(E \cap B_n) = \sum_{i=1}^n \mu^*(E \cap A_i)$$

and so

$$\begin{aligned}
\mu^*(E) &= \mu^*(E \cap B_n^C) + \sum_{i=1}^n \mu^*(E \cap A_i) \\
&\geq \mu^*(E \cap B^C) + \sum_{i=1}^n \mu^*(E \cap A_i)
\end{aligned}$$

for every $n \in \mathbb{N}$ where the inequality is due to the fact that $B^C \subseteq B_n^C$ and the monotonicity of outer measures. After taking limits, we have that

$$\begin{aligned}
\mu^*(E) &\geq \mu^*(E \cap B^C) + \sum_{i=1}^\infty \mu^*(E \cap A_i) \\
&\geq \mu^*(E \cap B^C) + \mu^*\left(\bigcup_{i \in \mathbb{N}} (E \cap A_i)\right) \\
&= \mu^*(E \cap B^C) + \mu^*(E \cap B)
\end{aligned}$$

where the second inequality follows by countable subadditivity. Another application of countable subadditivity yields

$$\mu^*(E) \leq \mu^*(E \cap B^C) + \mu^*(E \cap B)$$

and together the two inequalities establish that $B \in \mathcal{C}$, finishing the proof of (ii).

Finally, in order to show that $\mu^*|_{\mathcal{C}}$ is indeed a countably additive measure on \mathcal{C} , let $\{A_i\}_{i=1}^{\infty} \in \mathcal{C}$ be pairwise disjoint, and observe that for $B := \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{C}$ and any $E \subseteq \mathcal{X}$

$$\mu^*(E) \geq \mu^*(E \cap B^C) + \sum_{i=1}^{\infty} \mu^*(E \cap A_i)$$

due to our previous work. Letting $E = B$, we have

$$\begin{aligned} \mu^*(B) &\geq \sum_{i=1}^{\infty} \mu^*(B \cap A_i) \\ &= \sum_{i=1}^{\infty} \mu^*(A_i). \end{aligned}$$

Since the reverse inequality follows by the subadditivity of the outer measure, our proof is complete. \square

Note that in general such an extension may not be unique and we provide sufficient conditions for uniqueness in Chapter XX.

1.2.5. Defining the Lebesgue measure.

1.3. Abstract measure spaces

DEFINITION 1.9. A pair $(\mathcal{X}, \mathcal{F})$, where \mathcal{X} is an arbitrary set and \mathcal{F} is a σ -algebra on \mathcal{X} , is called a *measurable space*.

Although we had implicitly defined a measure in the previous section, it's appropriate to write down a formal definition in this section.

DEFINITION 1.10. Let $(\mathcal{X}, \mathcal{F})$ be a measurable space. A function $\mu : \mathcal{F} \rightarrow [0, \infty]$ is a *measure* on \mathcal{X} if

- (i) $\mu(\emptyset) = 0$
- (ii) For disjoint $\{A_i\}_{i \in \mathbb{N}} \in \mathcal{F}$

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

The triple $(\mathcal{X}, \mathcal{F}, \mu)$ is called a *measure space*. If $\mu(\mathcal{X}) = 1$ then μ is called a *probability measure* and $(\mathcal{X}, \mathcal{F}, \mu)$ is called a *probability space*.

DEFINITION 1.11. Given a measurable space $(\mathcal{X}, \mathcal{F})$, any set $A \in \mathcal{F}$ is called a *measurable set*. Conversely, any set $A \subset \mathcal{X}$ such that $A \notin \mathcal{F}$ is referred to as a *non-measurable set*.

CHAPTER 2

Measurable functions

TALK ABOUT EQUIVALENCE BETWEEN INDICATOR FUNCTIONS AND SETS

Part 2

Probability