These are my notes on basic measure theory. For each claim at least a sketch of proof is given unless: the claim can be established as a brute force exercise of applying definitions, or it has a long or otherwise "difficult" proof. The latter case is indicated by an asterisk.

Amir Kadivar, March 2016

Contents

0.0.0.1 Notation

The set of non-negative reals is denoted by $\mathbb{R}^+$ and the extended non-negative reals is denoted by $[0, \infty]$. The length of an interval (or otherwise the "standard" measure of a set) I is denoted by $|I|$ and the number of elements in a finite set A is denoted by $\#(A)$. The power set of X is denoted by $\mathscr{P}(X)$ and the complement of a subset $A \subset X$ is denoted by A^c. The closure and interior of A is denoted by $\bar{A}$ and A° respectively. Subset relation is indicated by $\subset$ and $\subseteq$ is avoided. Disjointness of unions are expressed in words and $\coprod$ is avoided. The indicator function of a set A is denoted by 1_A and χ_A is avoided. The preimage of a function is denoted by $f^{-1}(\cdot)$, that is for any X we mean $f^{-1}(X) = \{x : f(x) \in X\}$. "Quotes" are used for abuses of notation or informal definitions.

---

1 Measures and Integration

1.1 Motivation

Here are four problems we wish to solve with the machinery of measure theory:

1. If two functions agree except at a countable number of points (e.g. f and $f + 1_{\mathbb{Q}}$ for any f), they will not agree on Riemann-integrability, let alone the value of their integrals, despite the fact that they agree on a "much larger" set than the set on which they disagree. The "Lebesgue philosophy" is to ignore "negligible sets" (see below).
2. Additionally, we seek a "consistent" way of assigning measures to subsets of $\mathbb{R}^d$ [1]. Let us demand: **i)** the measure of "ordinary" sets should coincide with what we ordinarily assign to them as length, area, volume, **ii)** the measure of a set must be non-negative and zero for the empty set, **iii)** the measure of a countable disjoint union should equal the sum of its components' measures [2], and **iv)** the assignment of measure must be translation invariant, i.e "$|A| = |x + A|$" for any set A and any point $x \in \mathbb{R}^d$.

[1] This is a separate issue from 1 which is concerned with negligbility of countable subsets of $\mathbb{R}^d$. For instance, as we will see it is not obvious how to assign measures such that the Cantor sets gets a measure of 0 and the fat Cantor set gets a measure of $\frac{1}{2}$, as intuitively expected (these two sets are both uncountable).

[2] Note that we do not wish this to be true of uncountable disjoint unions since any interval can be written as such.

3. It turns out that, under the axiom of choice, it is not possible to assign to *all* subsets of $\mathbb{R}$ a "consistent" measure [3] as demonstrated by the *Vitali Theorem*: Under ZFC, $\mathbb{R}$ has nonmeasurable sets (cf. *Vitali sets*).
4. As a consequence of 3, we also need to demand that the set of measurable sets be "well behaved" in the sense that it is closed under reasonable set operations. Let us demand that the set of measurable sets be closed under all basic set operations, i.e $\cap, \cup, \Delta, \setminus, ^c$. This would imply, by induction, that it must also be closed under finitely many applications of set operations. We will see that demanding only this much would correspond to the limitations of Riemann integration. In order to fix 1 we would need to allow also countably infinite applications of set operations [4].

1.1.0.1 Negligible sets

We can define negligible sets consistently without invoking any abstract measure theory. A set $A \subset \mathbb{R}$ is *null* or *negligible* (in the Lebesgue sense [5]) if it can be covered by a countable union of intervals whose sum of lengths can be made arbitrarily small. Specifically, $M \subset \mathbb{R}$ is negligible if for any $\epsilon > 0$ there exists a sequence of intervals $I_k \subset \mathbb{R}$ such that $M \subset \bigcup_k I_k$ and $\sum_k |I_k| < \epsilon$.

1.1.1

Our goal in measure theory is to find a "large enough" family of subsets, namely *measurable sets*, over which a "consistent" measure can be defined. Note that within the scope of these notes there is no ambiguity on assigning measures to building block sets (e.g. intervals). It is the set of measurable sets which we have to carefully craft to make sure all the desirable requirements are satisfied while avoiding problems with the axiom of choice. In this sense, when there is an obvious choice of measure for building block sets we consider a description of measurable sets under a measure to uniquely define the measure itself.

In this section we mostly follow [1] and [2]. We first introduce, as grounding examples, simple "measures" than do not satisfy all the requirements, hence the quotes around elementary "measure" and Jordan "measure" [6].

1.2 Elementary "measure"

In this section we make concrete the idea that: elementary sets are finite collections of boxes and their intuitive assignment of measure is well-defined.

1.2.0.1 Definition

[3] The axiom of choice is in fact necessary for this issue to arise as there are no constructive examples of nonmeasurable sets. In fact, the *Solovay model* shows that ZF together with an *axiom of dependent chocie* can be consistent with the demand that all subsets of $\mathbb{R}^d$ be measurable if ZFC were to be consistent with the existence of an inaccessible cardinal.

[4] Note that we do not wish this to be true of uncountable applications of set operations as that would lead to nonmeasurable sets yet again.

[5] By this we mean, this is the notion of negligibility due to Lebesgue (the person) and not the property of having Lebesgue-measure 0, although the two are equivalent.

[6] These are both examples of what is technically called a *content*.

An *interval* in $\mathbb{R}$ has either of the forms $[a, b)$, $[a, b]$, (a, b), $(a, b]$ (note that degenerate single point intervals are allowed). A *box* in $\mathbb{R}^d$ is a set of the form $I_1 \times I_2 \times \ldots I_d$ where I_k are all intervals. An *elementary set* in $\mathbb{R}^d$ is one that can be expressed as a finite union of (potentially overlapping) boxes. For an interval I with endpoints $a \leq b$ its *length* is defined to be $|I| = b - a$. For a box $B = I_1 \times I_2 \times \ldots \times I_d$ its *volume* is defined to be $|B| = |I_1| \times |I_2| \times \ldots |I_d|$.

1.2.0.2 Proposition

Closure properties of elementary sets: Elementary sets in $\mathbb{R}^d$ are closed under all basic binary set operations $\cap, \cup, \setminus, \Delta$ and under translation (i.e if E is elementary $x + E$ is elementary for any $x \in \mathbb{R}^d$).

1.2.0.3 Proposition

Discretization lemma: For any interval $I \subset \mathbb{R}$ we have:

$$|I| = \lim_{n\to\infty} \frac{1}{n} \# \left(I \cap \frac{1}{n}\mathbb{Z} \right)$$

Similarly, for any box B in $\mathbb{R}^d$ we have:

$$|B| = \lim_{n\to\infty} \frac{1}{n^d} \# \left(B \cap \frac{1}{n^d}\mathbb{Z}^d \right)$$

proof: Let the end points of I be $a < b$ (when $a = b$ both the RHS and LHS are trivially zero) and define $i_n = \#(I \cap \frac{1}{n}\mathbb{Z})$. We must have $\frac{1}{n}(i_n - 1) \leq b - a$ and $\frac{1}{n}(i_n + 1) \geq b - a$ and thus $(b-a) - \frac{1}{n} \leq \frac{1}{n} i_n \leq (b-a)n + \frac{1}{n}$. The squeeze theorem now establishes the result for intervals. The corresponding claim for a box can be established by decomposing the box into the Cartesian product of d intervals. ■

1.2.0.4 Proposition

Well-definedness of elementary measure: Any elementary set $E = B_1 \cup B_2 \ldots \cup B_n$ can be written as a disjoint union $E = B'_1 \cup B'_2 \ldots \cup B'_{n'}$. Furthermore, the value of $\sum_k |B'_k|$ is independent of the choice of the B'_k.

proof: The first claim can be established by induction on n: the union of any n elementary sets can be written as a disjoint elementary union. The second claim follows from the discretization lemma. ■

1.2.0.5 Definition

For any elementary set $E = \bigcup_k B_k$ its *elementary "measure"* $m(E)$ is defined to be $\sum_k |B'_k|$ where the B'_k are disjoint and satisfy $E = \bigcup_k B'_k$. We drop the quotes henceforth for convenience.

1.2.0.6 Proposition

Properties of elementary measure: The elementary measure as defined above is:

i. *finitely additive*, i.e for any n disjoint elementary sets $E_1, \ldots, E_n$ we have $m(\bigcup_k E_k) = \sum_k m(E_k)$,
ii. *finitely subadditive*, i.e for any n elementary sets (possibly overlapping) $E_1, \ldots, E_n$ we have $m(\bigcup_k E_k) \leq \sum_k m(E_k)$, and
iii. *translation invariant*, i.e for any $x \in \mathbb{R}^d$ and any elementary set E we have $m(x + E) = m(E)$.

1.2.0.7 Proposition

Uniqueness of elementary measure: Any non-negative function $m'(\cdot)$ over elementary sets in $\mathbb{R}^d$ satisfying the three properties of the above proposition (finite additivity, finite subadditivity, and translation invariance) agrees everywhere with the elementary measure $m(\cdot)$ everywhere upto a scaling factor.

proof: Let $\alpha = m'\left([0,1)^d\right)$. By finite additivity and translation invariance, for any $n \in \mathbb{N}$ we have $m'\left([0,\frac{1}{n})^d\right) = n^{-d}\alpha$. Now applying the discretization lemma shows $m'(E) = \alpha m(E)$ for any elementary set. ■

1.3 Jordan "measure"

Elementary sets are all box-like and rather restrictive, e.g. neither a triangle nor a circle are elementary sets. However, both these examples can be approximated by elementary sets with arbitrary precision. The Jordan "measure" formalizes this *approximation*. As always, we only need to specify the set of Jordan-measurable sets and note those set operations under which this set is closed. It is, however, more intuitive to define the measure function directly as an extension of the underlying measure (i.e it must agree with elementary measure over elementary sets) and see what additional subsets of $\mathbb{R}^d$ are recovered as measurable:

1.3.0.1 Definition

For any $M \in \mathbb{R}^d$ define the *Jordan inner measure* $m_{*,J}$ and the *Jordan outer measure* $m^{*,J}$ as follows [7]:

$$m_{*,J}(M) = \sup_{E \subset M} m(E) \quad \text{and} \quad m^{*,J}(M) = \inf_{E \supset M} m(E)$$

where the supremum and infimum are over elementary sets E and m is the elementary measure. The set M is *Jordan-measurable* if the inner and outer Jordan measures agree.

1.3.0.2 Proposition

Closure properties of Jordan-measurable sets: The set of Jordan measurable sets in $\mathbb{R}^d$ is closed under all basic binary set operations $\cap, \cup, \setminus, \Delta$.

1.3.0.3 Remarks

1) If we extend the real line with infinity then the inner and outer measures are well defined for *every* set in $\mathbb{R}^d$ and not just for Jordan-measurable ones.

2) Clearly, the Jordan measure is an extension of the elementary measure in the sense that any elementary sets is Jordan measurable with the same measure (this follows from finite additivity of the elementary measure applied to the RHS of the definition of inner and outer measures). We can therefore safely let $m(\cdot)$ also represent the Jordan measure in unambiguous contexts.

3) One obvious case where a set is not Jordan-measurable is when either the inner and outer Jordan measure are ∞ which happens iff M is unbounded. However, we will see that there are bounded sets in $\mathbb{R}^d$ that are not Jordan-measurable.

[7] Again, technically, "content" is the right word, not "measure".

4) If we were to allow ∞ to be a valid Jordan measure (with the criteria that both inner and outer measures are ∞) then the compliment of a Jordan-measurable set is also Jordan-measurable.

1.3.0.4 Proposition

Properties of Jordan measure: The Jordan measure is *finitely additive* (and thus *finitely subadditive*) and *translation invariant*. Additionally, for any $M \subset \mathbb{R}^d$ we have:

$$m^{*,J}\left(\overline{M}\right) = m^{*,J}(M) \quad \text{ and } \quad m_{*,J}(M^\circ) = m_{*,J}(M)$$

In words: any set M has the same Jordan outer measure as its closure $\overline{M}$ and the same Jordan inner measure as its interior M°. As a corollary, M is Jordan-measurble iff its topological boundary ∂M has outer Jordan measure 0.

proof: The steps are simpler versions of the proof of a stronger proposition for the Lebesgue outer measure (see below). ■

1.3.0.5 Proposition

Uniqueness of Jordan measure: Any non-negative function $m'(\cdot)$ over Jordan-measurable sets in $\mathbb{R}^d$ satisfying the three properties of the above proposition (finite additivity, finite subadditivity, and translation invariance) agrees everywhere with the Jordan measure $m(\cdot)$ everywhere upto a scaling factor.

proof: The same procedure we used to prove the uniqness of elementary measure applies. ■.

1.3.0.6 Proposition

Examples of (non-)Jordan-measurable sets: **i)** The following are Jordan-measurable: ordinary Cantor sets (cf. appendix on Cantor sets), compact convex polytopes in $\mathbb{R}^d$, and open or closed balls in $\mathbb{R}^d$. **ii)** The following are *not* Jordan-measurable: fat Cantor sets, $[0,1]^d \cap \mathbb{Q}^d$, and $\bigcup_{n\in\mathbb{N}}[n, n+2^{-n}]$. Specifically, the countable union of Jordan-measurable sets is not necessarily Jordan-measurable even if bounded.

proof: The ordinary Cantor set has vanishing Jordan inner and outer measure and is thus Jordan measurable with measure 0. Measurability of "regular" objects like compact convex polytopes (e.g. a triangle) and spheres can be shown using dyadic meshes (cf. appendix on dyadic-intervals) and the alternative characterization of Jordan-measurabile sets as approximated by sequences of elementary sets (see below). For $[0,1]^d \cap \mathbb{Q}^d$ we use the property that any set has the same Jordan outer measure as its closure and the same Jordan inner measure as its interior. The closure of $[0,1]^d \cap \mathbb{Q}^d$ is $[0,1]^d$ with Jordan measure 1 and the interior is $\emptyset$ with Jordan measure 0. For the fat Cantor set C^f_β we note that its Jordan inner measure vanishes (as it has empty interior) and its Jordan outer measure can not be smaller than $1-\beta/(1-2\beta) > 0$. For $\bigcup_{n\in\mathbb{N}}[n, n+2^{-n}]$ we note that the Jordan outer measure is ∞ and the Jordan inner measure is 1, the latter being the intuitive "length" one would have wished for. ■

1.3.0.7 Proposition

Approximating Jordan-measurable sets by elementary sets: A bounded set $M \subset \mathbb{R}^d$ is Jordan-measurable **i)** iff for any $\epsilon > 0$ there are elementary sets $E_* \subset M \subset E^*$, such that $m(E^* \setminus E_*) < \epsilon$, or alternatively, **ii)** iff for any $\epsilon > 0$ there is an elementary set E such that $m^{*,J}(E\Delta M) < \epsilon$.

1.3.0.8 Proposition

Discretization lemma for Jordan-measurable sets: If $M \subset \mathbb{R}^d$ is Jordan-measurable then:

$$m(M) = \lim_{n \to \infty} \frac{1}{n^d} \# \left(M \cap \frac{1}{n} \mathbb{Z}^d \right)$$

1.3.0.9 Remark

It is possible for the limit in the RHS of the above to exist and M not be Jordan-measurable. For instance, take $M = [0,1] \cap \mathbb{Q}$ which gives the limit 1 or take $M = \pi + [0,1] \cap \mathbb{Q}$ which gives the limit 0.

1.3.0.10 Proposition

Connections between Jordan measure and Riemann integral:

i) If $M \subset [a,b]$ is measurable then 1_M is Riemann-integrable over $[a,b]$ and:

$$\int_a^b 1_M(x) dx = m_1(M)$$

where $m_1(\cdot)$ is the Jordan measure over $\mathbb{R}^1$.

ii) If $f : [a,b] \to \mathbb{R}^+$ is such that $M = \{(x,y) : x \in [a,b], 0 \leq y \leq f(x)\}$ is Jordan-measurable then f is Riemann-integrable over $[a,b]$ and:

$$\int_a^b f(x) dx = m_2(M)$$

where $m_2(\cdot)$ is the Jordan measure over $\mathbb{R}^2$.

iii) If $f : [a,b] \to \mathbb{R}^d$ is piecewise-continuous (Note that Riemann-integrability is a weaker condition than piecewise-continuity, cf. appendix on Riemann and Darboux integration) then $M = \{(x, f(x) : x \in [a,b]\}$ is Jordan-measurable in $\mathbb{R}^{d+1}$ with Jordan measure 0. Furthermore, $M = \{(x,y) : x \in [a,b], 0 \leq y \leq f(x)\}$ is Jordan-measurable in $\mathbb{R}^{d+1}$.

1.4 Lebesgue measure

We now wish to extend the Jordan measure to a larger family of sets such that pathological cases mentioned above become measurable. This is done by implicitly extending the building blocks (thus far elementary sets) to those that are countable (finite or countably infinite) unions of boxes.

1.4.0.1 Definition

Let $M \subset \mathbb{R}^d$. The *Lebesgue outer measure* m^* of M is defined to be:

$$m^*(M) = \inf_{\bigcup_n B_n \supset M} \sum_{n=1}^{\infty} |B_n|$$

where the infimum is over all *countable* collections of boxes $(B_n)_{n \in \mathbb{N}}$ and $|\cdot|$ is the volume of a box (i.e Jordan or elementary measure). It follows, by definition, that all countable sets (including $\emptyset$) have Lebesgue outer measure zero and that $m^*(M) \leq m^{*,J}(M)$.

1.4.0.2 Remarks

1) If we extend the real line with infinity then the Lebesgue outer measure is well defined for *every* set in $\mathbb{R}^d$.
2) Naturally, we would like to define a Lebesgue *inner* measure in analagy to the Jordan inner measure as follows (we will call this object $\widetilde{m}_*$ as we will see that it is not a particularly useful way of defining the Lebesgue inner measure):
$$\widetilde{m}_*(M) = \sup_{\bigcup_n B_n \subset M} \sum_{n=1}^{\infty} |B_n|$$
However, two difficulties arise. First, the corresponding bound by the Jordan inner measure is useless, i.e $m_{*,J}(M) \leq \widetilde{m}_*(M)$ is never strict [8]. Second, this definition would not even help us define measurability fruitfully. Consider the fat Cantor set to which we would like to assign measure $\frac{1}{2}$ but for which $\widetilde{m}_*$ evaluates to 0.
3) Therefore, we have an inherent asymmetry here. This leads to two possible, yet equivalent, definitions of Lebesgue-measurability. The first option (the Carathéodory method) is to build an alternative formulation for the inner measure that works properly and proceed as usual by defining Lebesgue-measurable sets to be those with identical outer and inner measures. Here we follow [1] which gives an equivalent but less confusing construction, namely *Littlewood's first principle* (i.e "Lebesgue-measurable sets are almost open") relying solely on Lebesgue outer measure.

1.4.0.3 Proposition

Properties of Lebesgue outer measure: The Lebesgue outer measure $m^*(\cdot)$ is *non-negative*, i.e $m^*(M) \geq 0$ with equality for $M = \emptyset$, *monotonic*, i.e $M \subset N \Rightarrow m^*(M) \leq m^*(N)$, and *$\sigma$-subadditive* [9].

proof: The first two properties follow from definition. We here prove σ-additivity using a useful and generic trick (called the "ϵ of room" trick as per [1]): instead of showing the desired inequality we prove it within a margin of error ϵ in such a way that ϵ can be brought to zero without affecting the proof. We want to show:
$$m^*\left(\cup_n A_n\right) \leq \sum_n m^*(A_n)$$
for any countable collection $\{A_n\}_{n\in\mathbb{N}} \subset \mathscr{P}(\Omega)$. This is, by definition:
$$\inf_{\cup_n B_n \supset \cup_n A_n} \sum_n |B_n| \leq \sum_n \inf_{\cup_k B_k^n \supset A_n} \sum_k |B_k^n|$$
where the infimums are over countable collections of boxes. Now take any $\epsilon > 0$. Applying the properties of the infimum to all n terms on the RHS allows to claim there exists a collection $\{C_k^n\}_{n,k\in\mathbb{N}}$ of boxes such that for every $n \in \mathbb{N}$ we have:
$$\sum_k |C_k^n| \leq \inf_{\cup_k B_k^n \supset A_n} \sum_k |B_k^n| + \epsilon/n = m^*(A_n) + \epsilon/n$$
Adding over all n gives:
$$\sum_{n,k} |C_k^n| \leq \sum_n m^*(A_n) + \epsilon$$

[8] This has to do with the fact that finite sums of non-negatives can approach a limit, with arbitrary precision, from below and not from above. Thus, lifting the finiteness restriction from elementary covers does not lead to more power in the inner measure case. On a deeper level, this follows from the fact that elementary measure is subadditive, and not superadditive [1].

[9] The $\sigma-$ prefix generally stands instead of "countable union", e.g. as in σ-finite, σ-compact.

On the other hand, since $\bigcup_{n,k} C_k^n \supset \bigcup_n A_n$, we have:

$$\sum_{n,k} |C_k^n| \geq \inf_{\cup_n B_n \supset \cup_n A_n} \sum_n |B_n| = m^*(\cup_n A_n)$$

Thus:

$$m^*\left(\cup_n A_n\right) \leq \sum_{n,k} |C_k^n| \leq \sum_n m^*(A_n) + \epsilon$$

This completes the proof since we have established the desired inequality upto an ϵ of error for *any choice of* $\epsilon > 0$. ■

1.4.0.4 Definition

Lebesgue-measurable sets (defn. by approximation): A set $M \subset \mathbb{R}^d$ is Lebesgue-measurable if for every $\epsilon > 0$ there exists an open set $U \subset \mathbb{R}^d$ containing M such that $m^*(U \backslash M) < \epsilon$ [10] in which case the Lebesgue measure of M is $m(M) = m^*(M)$. Specifically, sets with outer measure zero are necessarily Lebesgue-measurable with measure 0.

1.4.0.5 Proposition

Closure properties of Lebesgue-measurable sets: The following sets are Lebesgue-measurable in $\mathbb{R}^d$: **i)** open sets, **ii)** complement of Lebesgue-measurable sets, **iii)** countable union of Lebesgue-measurable sets.

1.4.0.6 Remark

From **i** and **ii** it follows that all closed sets are Lebesgue-measurable. From **iii** and **ii** it follows that countable intersections of Lebesgue-measurable sets are Lebesgue-measurable.

1.4.0.7 Proposition

Alternative approximations for Lebesgue-measurable sets: The following approximation criteria provide equivalent conditions for a set $M \in \mathbb{R}^d$ to be Lebesgue-measurable:

1. $\forall \epsilon \; \exists$ open $U \supset M : m^*(U \setminus M) < \epsilon$.
2. $\forall \epsilon \; \exists$ closed $V \subset M : m^*(M \setminus V) < \epsilon$.
3. $\forall \epsilon \; \exists$ open $U : m^*(U \Delta M) < \epsilon$.
4. $\forall \epsilon \; \exists$ closed $V : m^*(M \Delta V) < \epsilon$.

1.4.0.8 Proposition

*Properties of Lebesgue measure**: The Lebesgue measure is σ-additive (and thus σ-subadditive) and translation invariant.

[10] This could have equivalently been written as $m^*(U \Delta M)$ since $M \subset U$. This brings to fore the connection between this definition and the approximation-by-elementaries formulation of Jordan-measurables.

1.4.1 Alternative definition of Lebesgue-measurable sets

1.4.1.1 Proposition

Lebesgue inner measure. For any bounded $M \subset \mathbb{R}^d$ let its Lebesgue inner measure be $m_*(M) = m(E) - m^*(E \setminus M)$ for *any* elementary set $E \supset M$ (the claim is that the choice of E is irrelevant). Furthermore, $m_*(M) \leq m^*(M)$ and equality holds iff M is Lebesgue-measurable.

1.4.1.2 Proposition

Carathéodory's criterion. A set $M \subset \mathbb{R}^d$ is Lebesgue-measurable iff for any elementary set $E \subset \mathbb{R}^d$ we have $m(A) = m^*(E \cap M) + m^*(E \setminus M)$.

1.5 σ-algebras and abstract measures

In this section we formalize the results of the previous section in abstract terms. In all that follows let Ω be the universe (instead of $\mathbb{R}^d$).

1.5.0.1 Definition

$\mathscr{F} \subset \mathscr{P}(\Omega)$ is an *algbera* [11] over Ω if $\mathscr{F}$ is closed under complements and unions.

1.5.0.2 Remarks

1) From this definition it follows that $\emptyset, \Omega \in \mathscr{F}$ and that $\mathscr{F}$ is closed finite applications of all basic binary set theoretic operations $\cap, \cup, \setminus, \Delta$. Alternatively, we could have demanded that $\mathscr{F}$ be closed under complements and either of $\cap, \cup, \setminus$ (but being closed under complement and Δ is not enough). **2)** Examples are: $\mathscr{P}(\Omega)$ for any Ω and the set of elementary, Jordan-measurable, or Lebesgue-measurable sets in $\mathbb{R}^d$. **3)** As usual a *subalgebra* of $\mathscr{F}$ is a subset of $\mathscr{F}$ which satisfies the axioms on its own, i.e is closed under complements and unions.

1.5.0.3 Definition

$\mathscr{F} \subset \mathscr{P}(\Omega)$ is a *σ-algbera* [12] over Ω if it is closed under complements and countable unions. The pair $(\Omega, \mathscr{F})$ is a *measurable space.*

1.5.0.4 Remarks

1) From this definition it follows that $\mathscr{F}$ is also an algebra over Ω and that it is closed under countable intersections. Equivalently, we could have demanded that $\mathscr{F}$ be closed under complements and countable intersections. **2)** For any Ω, $\mathscr{P}(\Omega)$ is a σ-algbera. Over $\mathbb{R}^d$, the set of Lebesgue-measurable sets is a σ-algebra. But the set of elementary sets and the set of Jordan-measurable sets are not σ-algebras over $\mathbb{R}^d$. **3)** As usual

[11] Alternatively, a *boolean algebra*, or a *field.*
[12] Alternatively, a *σ-field.*

a *sub-σ-algebra* (or subalgebra when the context is clear) of $\mathscr{F}$ is a subset of $\mathscr{F}$ which satisfies the axioms on its own, i.e is closed under complements and countable unions. In the section on stochastic processes we will see that sub-σ-algebras can be regarded as "experiments" whose possible outcomes are "events" (the sets in the subalgebra) and in this sense represent "partial information".

1.5.0.5 Definition

Let $\mathscr{F} \subset \mathscr{P}(X)$ be an arbitrary collection of subsets. The *σ-field generated by $\mathscr{F}$* is defined to be:

$$\sigma(\mathscr{F}) = \bigcap_{\mathscr{F} \subset \Sigma} \Sigma$$

where the intersection is over σ-algebras Σ over Ω. That any arbitrary intersection of σ-algebras is necessarily a σ-algebra is true but nontrivial (in fact, this is clearly not true for arbitrary unions of σ-algebras).

1.5.0.6 Definition

Let $\mathscr{F}$ be an algebra over Ω. A function $m : \mathscr{P}(\Omega) \to [0, \infty]$ is a *content* if it is finitely additive and $m(\emptyset) = 0$.

1.5.0.7 Remark

In $\mathbb{R}^d$ the elementary "measure" is a content over the algebra of elemtnary sets, and the Jordan "measure" is a content over the algebra of Jordan-measurable sets. Note that both these contents are merely restrictions of the Lebesgue measure to the corresponding σ-algebras.

1.5.0.8 Definition

Let $\mathscr{F}$ be a σ-algebra over Ω. A function $m^* : \mathscr{P}(\Omega) \to [0, \infty]$ is an *outer measure* if $m(\emptyset) = 0$, and m is σ-subadditive and monotone, i.e $A \subset B \Rightarrow m^*(A) \leq m^*(B)$.

1.5.0.9 Remarks

1) As shown earlier, in $\mathbb{R}^d$, both the Lebesgue and Jordan outer measures are outer measures. **2)** An important intuition to have in mind about outer measures is that they are easy to make. This is ultimately because demonstrating σ-subadditivity is much easier than demonstrating σ-additivity. For instance, the *counting measure* (i.e the measure of A is $|A| \in \mathbb{N}$ if A is finite and ∞ otherwise) is an outer measure over $\mathbb{R}^d$. Or in the definition of the Lebesgue outer measure, we could have as well used open or closed balls, half-spaces, or compact convex polytopes, instead of boxes and the proof of σ-subadditivity would have remained untouched. **3)** In our study of Lebesgue measure in $\mathbb{R}^d$ we followed the more geometric route and mentioned Carathéodory's method of defining the Lebesgue measure as a less intuitive alternative. In the abstract setting, things are more or less reversed thanks to the power of the Carathéodory extension theorem (see below).

1.5.0.10 Definition

Let $\mathscr{F}$ be an algebra over Ω. A function $m_0 : \mathscr{P}(\Omega) \to [0, \infty]$ is a *pre-measure* if $m_0(\emptyset) = 0$, and m_0 is σ-additive, i.e whenever $\bigcup_{n \in \mathbb{N}} A_n \in \mathscr{F}$ for disjoint A_n, we have $m_0(\bigcup A_n) = \sum_n m_0(A_n)$.

1.5.0.11 Remarks

1) What makes m_0 a pre-measure and not a measure is not its additivity properties (it is required to be σ-additive which is as strong as it gets) but the domain on which it is defined. If the domain is not a σ-algebra and merely an algebra, then m_0 is not a measure, but, as we shall see, a *bona fide* measure. **2)** A simple example is this: in $\mathbb{R}^d$ the elementary measure is a pre-measure on the elementary algebra (not by definition, since one has to prove σ-additivity and take care of coelementary sets).

1.5.0.12 Definition

Let $(\Omega, \mathscr{F})$ be a measurable space. A function $m : \mathscr{F} \to [0, \infty]$ is a *measure* over $(\Omega, \mathscr{F})$ if it is σ-additive and $m(\emptyset) = 0$. For any such measure, $(\Omega, \mathscr{F}, m)$ is a *measure space*. Exmaple: $(\mathbb{R}^d, \mathscr{L}, m)$ is a measure space where $m(\cdot)$ is the Lebesgue measure and $\mathscr{L}$ is the σ-algebra of Lebesgue-measurable sets in $\mathbb{R}^d$.

1.5.0.13 Definition

Let $(\Omega, \mathscr{F}, m)$ be a measure space. We say that m is *finite* if $m(\Omega) < \infty$. We say that m is *σ-finite* if Ω can be covered by a countable union of sets with finite measure.

1.5.0.14 Remarks

1) A special case of a finite measure is when $m(\Omega) = 1$ in which case m is a *probability measure* and $(\Omega, \mathscr{F}, m)$ is a *probability space*. Naturally, any finite measure gives a probability measure after proper scaling. **2)** Example: the counting measure is σ-finite over $\mathbb{N}$ (with the power set σ-algebra) but not σ-finite over $\mathbb{R}$. The Lebesgue measure, however, is σ-finite over $\mathbb{R}$ since $\mathbb{R} = \bigcup_n [-n, n]$. **3)** Although finiteness is a property of the measure only, σ-finiteness is a joint property of the measure and the σ-algebra. For instance, if we were to pick $\mathscr{F}$ such that Ω could not be covered by any countable collection of $\mathscr{F}$-sets then *no measure* could be σ-finite over $\mathscr{F}$.

1.5.0.15 Proposition

*Identification of measures over generator algebras** (cf. [2]): Let $\mathscr{A}$ be an arbitrary algebra over Ω and m_1 and m_2 be two σ-finite measures over $\sigma(\mathscr{A})$. If m_1 and m_2 agree everywhere in $\mathscr{A}$ then they agree everywhere in $\sigma(\mathscr{F})$, i.e they are the same measures.

1.5.0.16 Remark

1) This is not true if we remove the σ-finiteness requirement. For instance, the set of open half-lines $\mathscr{A} = \{(x, \infty) : x \in \mathbb{R}\}$ generates the Borel σ-algebra $\mathscr{B}(\mathbb{R})$ (see section on complete measures). The Lebesgue measure and the counting measure agree everywhere on $\mathscr{A}$ but they are not identical over $\mathscr{B}(\mathbb{R})$. **2)** The requirement that $\mathscr{A}$ be an algebra is two strong. In fact, it is enough to demand that $\mathscr{A}$ be a π-*system*, i.e that it is only closed under finite intersections. **3)** As a corollary, we can easily show that if m_1 and m_2 are finite (e.g. probability measures) and they agree on a π-system $\mathscr{A}$, *and* Ω can be covered by a countable union of $\mathscr{A}$-sets, then the agreement of m_1 and m_2 on $\mathscr{A}$ implies their agreement on $\sigma(\mathscr{A})$.

1.5.0.17 Proposition

*Carathéodory extension theorem**: Let m^* be an outer measure over Ω. Define a set $X \subset \Omega$ to be *Carathéodory-measurable* with respect to m^* if for every $A \subset \Omega$ we have:

$$m^*(A) = m^*(A \cap X) + m^*(A \setminus X)$$

Then the set $\mathscr{F}$ of Carathéodory-measurable subsets is a σ-algebra and $m : \mathscr{F} \to [0, \infty]$ the restriction of m^* to $\mathscr{F}$ is a measure.

1.5.0.18 Proposition

*Hahn-Kolmogorov theorem**: Let $m_0 : \mathscr{B}_0 \to [0, \infty]$ be a pre-measure over an algebra $\mathscr{F}_0$. There exists a measure $m : \sigma(\mathscr{F}_0) \to [0, \infty]$ which agrees with m_0 on $\mathscr{F}_0$.

1.5.0.19 Remark

The Carathéodory extension theorem formalizes the construction of the Lebesgue measure from the Lebesgue outer measure. Similarly, the Hahn-Kolmogorov theorem formalizes the construction of the Lebesgue measure from the elementary measure.

1.5.1 Excursion: equivalence with respect to algebras

The notion of equivalence with respect to an algebra has no analogue in our concrete developments for $\mathbb{R}^d$. Nevertheless, we mention it here since it has a deep connection to conditional probabilities. The core idea is similiar to that of the coarseness of a topology.

1.5.1.1 Definition

Let $\mathscr{F}$ be an arbitrary algebra over Ω. We say that $\mathscr{F}$ does *not distinguish between* $x, y \in \Omega$, or x, y are $\mathscr{F}$*-equivalent* if for every $A \in \mathscr{F}$ both x and y lie either in A or A^c, i.e $1_A(x) = 1_A(y)$. This is an equivalence relation and thus imposes a partition on Ω.

1.5.1.2 Proposition

Properties of algebra-equivalence: Let $\mathscr{F}$ be an arbitrary algebra over Ω. **i)** Any $C \subset \Omega$ is an equivalence class with respect to $\mathscr{F}$ iff $C \in \mathscr{F}$ and $\mathscr{F}$ contains no proper subset of C. **ii)** Any two elements $x, y \in \Omega$ are $\mathscr{F}$-equivalent iff they are $\sigma(\mathscr{F})$-equivalent, i.e the partitions induced by $\mathscr{F}$ and $\sigma(\mathscr{F})$ are identical.

1.6 Complete measure spaces

In this section we introduce the Borel σ-algebra and its corresponding measure, and note their relation to the Lebesgue measure.

1.6.0.1 Definition

Let $(X, \mathscr{T})$ be a topological space (e.g. $\mathbb{R}^d$ with the natural Euclidean topology). The *Borel* σ-algebra $\mathscr{B}(X)$ of X is defined to be $\sigma(\mathscr{T})$, i.e the σ-algebra generated by the open subsets of X. Elements of $\mathscr{B}(X)$ are called *Borel-measurable*.

1.6.0.2 Remark

We can easily show that all the following subsets of $\mathbb{R}^d$ equivalently generate $\mathscr{B}(\mathbb{R}^d)$: **i)** open subsets, **ii)** closed subsets, **iii)** compact subsets, **iv)** open balls, **v)** boxes, **vi)** elementary sets.

1.6.0.3 Definition

A measure space $(\Omega, \mathscr{F}, m)$ is *complete* if every subset of every null set is measurable.

1.6.0.4 Proposition

*Incompleteness of the Borel measure space**: The measure space $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d), m)$ is not complete.

proof: Here is the gist of the proof which needs some theory of cardinals to make concrete. Consider the middle thirds Cantor set C. We know that it is isomorphic to $\mathbb{R}$ and thus the cardinality of its power set is $2^{|\mathbb{R}|}$. On the other hand, one can show (using *transfinite induction*) that the cardinality of the Borel σ-algebra is identical to $|\mathbb{R}|$. It follows that the Cantor set (which is a Jordan and Lebesgue null set) must have nonmeasurable subsets.

1.6.0.5 Remarks

1) There is a natural way to complete an incomplete measure space by including all subsets of its null sets in the σ-algebra and assigning a measure 0 to them. In this sense the Lebesgue measure space over $\mathbb{R}^d$ is the *completion* of the Borel measure space. **2)** Incompleteness of the Borel measure space does not depend on the full axiom of choice (it can be proved using only the axiom of countable choice). However, it is difficult, but possible, to demonstrate *specific* examples of non-Borel sets in $\mathbb{R}^d$ that are subsets of null sets [1].

1.7 Lebesgue Integration

The goal of Lebesgue integration is to resolve the very first issue raised at the beginning of this chapter: if f and g agree everywhere except in a null set then we expect their integrability and the value of their integrals to coincide. The central idea can be visualized in the simplest case where $f : \mathbb{R} \to \mathbb{R}$. Instead of approximating the area under f with finer and finer vertical columns (as we do in Riemann integration), we approximate the area with finer and finer horizontal slabs. The increased "power" of Lebesgue integration, however, does not come from choice of vertical vs horizontal slabs (one could define the Riemann integral with horizontal slabs as well), but on that these slabs need not be erected on intervals but on arbitrary measurable sets. Accordingly, another group of functions, namely *simple functions*, take center in Lebesgue integration corresponding to piecewise constant functions in Riemann integration. Then, as with the Darboux formulation of Riemann integration (cf. appendix on Riemann and Darboux integration), we define the Lebesgue integral first on simple functions and generalize to an appropriately class of Lebesgue integrable functions.

1.7.0.1 Definition

A non-negative function $f : \mathbb{R}^d \to [0, \infty]$ is *simple* if it is a finite linear combination of indicator functions:

$$f = \alpha_1 1_{A_1} + \alpha_2 1_{A_2} + \ldots + 1_{A_n}$$

where the A_k are Lebesgue-measurable and the α_k are non-negative. Note that simple functions take only a finite number of values (the converse is not true).

1.7.1

Before we proceed with technicalities, we first address the subtle differences between Riemann and Lebesgue integration:

1.7.1.1 Remarks

1) Recall that the Riemann integral is first defined over bounded domains and improper integrals are defined as *limits* of proper integrals over arbitrarily large bounded domains. Lebesgue integration is *defined* over arbitrary measurable sets of $(\Omega, \mathscr{F})$ and Ω (or any other unbounded measurable domain of integration) is just another set. **2)** It is customary for the Riemann integral in a single variable to be interpreted with a *sense of direction*, that is $\int_a^b f dx = -\int_b^a f dx$. This sense is lost in multiple variables (e.g. consider $\int_\Omega f dx$ where Ω is the unit circle in $\mathbb{R}^2$). In Lebesgue integration this asymmetry is removed, again because integrals are defined primarily over sets and not on intervals. The appropriate notation for the Lebesgue integral over an interval $[a, b]$ would then be $\int_{[a,b]} f dx$. **3)** Riemann-integrability is defined with reference to the equality of the upper and lower approximations. Although a similar result is true for Lebesgue integration it is more common to demand, by way of definition, a property called "measurability" from the function before considering its Lebesgue-integrability.

1.7.1.2 Definition

Let $(\Omega, \mathscr{F}, \mu)$ be a measure space and let (S, Σ) be a measurable space. A function $f : \Omega \to S$ is $(\mathscr{F}, \Sigma)$-*measurable* (or simply measurable if the context is clear) if for every $A \in \Sigma$ we have $f^{-1}(A) \in \mathscr{F}$.

1.7.1.3 Remarks

1) This is exactly the definition of continuous functions over topological spaces (just replace measurable sets with open sets). The ease of manipulating preimages (e.g. under set theoretic operations) is the main reason we like such formulations of structural properties. **2)** The above definition has many equivalents for the special case of the Lebesgue measure space over $\mathbb{R}^d$ which are documented below. **3)** When $S = \mathbb{R}^d$ it is typically assumed that the target is the Borel measure space. That is, $f : \mathbb{R}^d \to \mathbb{R}$ is *Lebesgue-measurable* if it is $(\mathscr{L}(\mathbb{R}^d), \mathscr{B}(\mathbb{R}))$-measurable. Similarly, it is *Borel-measurable* if it is $(\mathscr{B}(\mathbb{R}^d), \mathscr{B}(\mathbb{R}))$-measurable. Furthermore, for most practical purposes (as we will see in stochastic processes), Borel-measurable functions are more desirable than Lebesgue-measurable functions. There are a few reasons for this preference for the Borel σ-algebra on either the domain or the target measure space: **i)** $(\mathscr{L}, \mathscr{L})$-measurable functions are not even necessarily continuous (e.g. Cantor's staircase function). **ii)** Composition of Borel-measurable functions is always Borel-measurable. The same is not true for Lebesgue-measurable functions (*proof*: draw a picture).

iii) The Lebesgue measure space is only defined for $\Omega = \mathbb{R}^d$ but the Borel measure space is defined for arbitrary topological spaces. Furthermore, since $\mathscr{B}(X)$ is, by definition, the smallest σ-algebra containing all open sets of X, $\mathscr{B}(X)$ plays nicely with the topology of X.

1.7.1.4 Proposition

Let $(\Omega, \mathscr{F}, \mu)$ be a measurable space and (S, Σ) be a measure space with $\mathscr{A} \subset \Sigma$ such that $\sigma(\mathscr{A}) = \Sigma$. A function $f : \Omega \to S$ is measurable iff $f^{-}1(A) \in \mathscr{F}$ for every $A \in \mathscr{A}$. In words: it always suffices to check the measurability condition only on a generating set of the destination σ-algebra.

1.7.1.5 Prosposition

Alternative definitions of Lebesgue-measurable functions in $\mathbb{R}^d$: Let $f : \mathbb{R}^d \to [0, \infty]$. The following are equivalent to f being Lebesgue-measurable:

1. f is the pointwise limit of a sequence of simple functions.
2. f is the pointwise almost everywhere limit of a sequence of simple functions.
3. The preimage, under f, of open/closed half-lines are Lebesgue-measurable in $\mathbb{R}^d$.
4. The preimage, under f, of open/closed intervals are Lebesgue-measurable in $\mathbb{R}^d$.

1.7.1.6 Proposition

We can consistently define the Lebesgue integral of a non-negative simple function $f = \sum_k \alpha_k 1_{A_k}$ to be:

$$\int_{\mathbb{R}^d} f dx = \sum_k \alpha_k m(A_k)$$

where $m(\cdot)$ is the Lebesgue measure. The claim here is that if a simple function has two representations as linear combinations of indicators, the corresponding RHS sums are identical.

1.7.1.7 Definition

Let $f : \mathbb{R}^d \to [0, \infty]$. Its *lower Lebesgue integral* is:

$$\underline{\int_{\mathbb{R}^d}} f dx = \sup_{0 \leq g \leq f} \int_{\mathbb{R}^d} g dx$$

If f is additionally Lebesgue-measurable, then the above is also defined to be its *Lebesgue integral*:

$$\int_{\mathbb{R}^d} f dx = \sup_{0 \leq g \leq f} \int_{\mathbb{R}^d} g dx$$

1.7.1.8 Remarks

1) Our formulation does not distinguish between Lebesgue-measurablity and Lebesgue-integrability of a function. Some authors distinguish the two by demanding that the lower Lebesgue integral be finite for a function to be called Lebesgue-integrable functions, e.g. they would *not* say $f = 1/x$ is Lebesgue-integrable with integral ∞. **2)** One coulde define an upper Lebesgue integral by:

$$\overline{\int_{\mathbb{R}^d}} f dx = \inf_{g \geq f} \int_{\mathbb{R}^d} g dx$$

where the infimum is over non-negative simple functions. This is not a useful construction because unless f is bounded it may not be bounded above by simple functions. However, if f is Lebesgue-measurable *and* is bounded then the upper and lower Lebesgue integrals will coincide. **3)** Thus far we have only considered non-negative functions. However, note that for any real function f we can write $f = f_+ - f_-$ where f_+ and f_- are both non-negative [13] and then we define:

$$\int_{\mathbb{R}^d} f dx = \int_{\mathbb{R}^d} f_+ dx - \int_{\mathbb{R}^d} f_- dx$$

demanding that at least on of the integrals on the RHS be finite.

1.7.2 Some important results

In all the following we consider an arbitrary measure space $(\Omega, \mathscr{F}, \mu)$ and focus on non-negative real functions $f : \Omega \to \mathbb{R}$, i.e by measurable we mean $(\mathscr{F}, \mathscr{B}(\mathbb{R}))$-measurable.

1.7.2.1 Proposition

Fatou's lemma: Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of non-negative measurable functions. We have:

$$\liminf_{n\to\infty} \int_\Omega f_n d\mu \geq \int_\Omega \liminf_{n\to\infty} f_n d\mu$$

1.7.2.2 Proposition

Monotone convergence theorem: Let $(f_n)_{n\in\mathbb{N}}$ be a pointwise non-decreasing sequence of non-negative measurable functions. We have:

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega \lim_{n\to\infty} f_n d\mu$$

1.7.2.3 Proposition

Borel-Cantelli Lemma: Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of $\mathscr{F}$-measurable sets. We have:

$$\sum_n \mu(A_n) < \infty \Rightarrow \mu(\limsup_n A_n) = 0$$

In words (for background, cf. appendix on Algebras and sequences of sets): if the A_n have a finite sum of measures then almost every $\omega \in \Omega$ appears in at most finitely many of the A_n.

proof: By defininition, we have:

$$\limsup_n A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} A_k \subset \bigcup_{k\geq n} A_k$$

By subadditivity of μ we have the following for every $n \in \mathbb{N}$:

$$\mu(\limsup_n A_n) \leq \sum_{k=n}^{\infty} \mu(A_k)$$

but the RHS tends to zero, by assumption, as $n \to \infty$. Since the inequality holds for any n the proof is complete. ■

[13] this move is not allowed in Riemann integration since it potentially splits the domain of f_+ and f_- into "countably-intricate" sets for instance take $f(x) = 2 \times 1_{\mathbb{Q}} - 1$

1.7.2.4 Proposition

1. *Markov's inequality*: for any $0 \leq \lambda \leq \infty$ we have:

$$\mu\left(f^{-1}\left([\lambda, \infty)\right)\right) \leq \frac{1}{\lambda} \int_{\Omega} f d\mu$$

2. *Chebyshev's inequality*: for any $0 \leq \lambda \leq \infty$ and $0 \leq p \leq \infty$ we have:

$$\mu\left(f^{-1}([\lambda, \infty)\right) \leq \frac{1}{\lambda^p} \int_{\Omega} |f|^p d\mu = \left(\frac{\|f\|_p}{\lambda}\right)^p$$

3. *Jensen's inequality*: if μ is finite, i.e $\mu(\Omega) < \infty$ and $\phi : \mathbb{R} \to \mathbb{R}$ is *convex* then:

$$\phi\left(\int_{\Omega} f d\mu\right) \leq \int_{\Omega} \phi \circ f d\mu$$

2 Appendix

In this section we review some objects and operations that serve mostly as building blocks or (counter) examples in previous sections. Elementary results in real analysis and point set topology are assumed.

2.1 Dyadic Intervals

2.1.0.1 Definition

The *dyadic interval* D_i^n is defined to be $\left[\frac{i}{2^n}, \frac{i+1}{2^n}\right)$ for any $i, n \in \mathbb{N}$ with $i \leq 2^n$. A *dyadic cube* in $\mathbb{R}^d$ is any set of the form

$$D_{i_1}^n \times D_{i_2}^n \ldots \times D_{i_k}^n = \left[\frac{i_1}{2^n}, \frac{i_1+1}{2^n}\right) \times \left[\frac{i_2}{2^n}, \frac{i_2+1}{2^n}\right) \ldots \times \left[\frac{i_d}{2^n}, \frac{i_d+1}{2^n}\right)$$

for some d integers $i_1, i_2, \ldots, i_d$ in $\{1, \ldots, 2^n\}$.

2.1.0.2 Proposition

Properties of dyadic intervals: **i)** For any $n \in \mathbb{N}$ there are a total of 2^n dyadic sets D_i^n each of length 2^{-n}. These 2^n sets are disjoint and provide a partition of $[0, 1)$. **ii)** (*nesting property*) For any $n \in \mathbb{N}$ any dyadic interval D_j^{n+1} is either contained within D_i^n or lies entirely outside it. It follows that any two dyadic intervals D_i^n and D_j^m are either disjoint or contained in one another.

2.1.0.3 Definition

The *dyadic expansion* functions $d_n : [0, 1) \to \{0, 1\}$ assign to any real $\omega \in [0, 1)$ the n-th digit of its binary representation; that is, if $[\omega]_2 = 0.\omega_1\omega_2 \ldots$ then $d_n(\omega) = \omega_n$.

2.1.0.4 Proposition

Properties of dyadic expansions: For any $\omega \in [0, 1)$ and any $n \in \mathbb{N}$ there exists a unique dyadic interval D_i^n that contains ω and we always have $d_n(\omega) \equiv i \mod 2$.

2.2 Cantor sets

We consider two families of Cantor sets:

2.2.0.1 Definition

Fix some $0 < \alpha < 1$ and let $A_1 = [0,1]$. For each $n \in \mathbb{N}$, let $A_{n+1} \subset A_n$ be the collection of intervals obtained by excluding from every interval in A_n, say with length ℓ, its centered subinterval of length $\alpha\ell$. We define "A_∞" to be C_α, the Cantor set corresponding to α. The parameter α is assumed to be $\frac{1}{3}$ when not specified in which case the Cantor set is known as the *middle thirds Cantor set*.

2.2.0.2 Definition

Fix some $0 < \beta < 1$ and let $B_1 = [0,1]$. For each $n \in \mathbb{N}$, let $B_{n+1} \subset B_n$ be the collection of intervals obtained by excluding from every interval in B_n, say with length ℓ, its centered subinterval of length $\beta^n\ell$. We define "B_∞" to be C^f_β, the *fat Cantor set* corresponding to β. The parameter β is assumed to be $\frac{1}{4}$ when not specified in which case the fat Cantor set is known as the *Smith-Volterra-Cantor* set.

2.2.0.3 Remarks

1) The quotes around "A_∞" and "B_∞" have to do with the fact that it is not immediately obvious that the construction leads to a nonempty set. This is established by Cantor's intersection theorem (see below). Furthermore, the limiting sets can be written explicitly as countably infinite intersections $\bigcap_n A_n$ and $\bigcap_n B_n$ which are preferable to getting involved in limits of sequences of sets.

2) The reason we need $\beta < \frac{1}{2}$ in the construction of fat Cantor sets is to ensure the subinterval to be removed is not larger than its containing interval, i.e $\beta^2 < 1 - 2\beta$.

3) Typically we are concerned with values of α and β that are of the form $1/n$ for some $n \in \mathbb{N}$. The results below are true for any choice but proofs are easier and more intuititve for rational values. Specifically, if one considers irrational α, β one is lead to the representation of numbers in $[0,1]$ in an irrational base which has some nonintuitive (and for our purposes irrelevant) properties.

4) For any $n \in \mathbb{N}$, the sets A_n and B_n are composed of exactly 2^n disjoint intervals. This leads to the intuition that every point in a Cantor (or fat Cantor) set corresponds to a sequence of binary digits. This establishes isomorphism of any Cantor set with $\mathbb{R}$ and provides a connection to dyadic intervals and infinite binary trees.

2.2.0.4 Proposition

Cantor's Intersection Theorem: Any decreasing nested sequence of compact subsets of a compact space has nonempty intersection.

proof: Denote the decreasing sequence by (C_n) and the ambient space by X. Define the increasing sequence of open sets (C'_n) by $C'_n = X \setminus C_n$. If the intersection $\bigcap_n C_n$ is empty then $\bigcup_n C'_n = X \setminus \bigcap_n C_n = X$, which by compactness of X, implies that there must be a finite subset of $\{C'_n\}$ that covers X. If C'_k is the largest of the sets in the finite cover, we must have $X = C'_k$ and thus $C_k = \emptyset$ which is a contradiction. ■

2.2.0.5 Proposition

Topological Properties of Cantor Sets: A Cantor set of either kind is **(i)** compact, **(ii)** totally disconnected, **(iii)** nowhere dense in $[0,1]$, and **(iv)** homeomorphic to any other Cantor set.

proof: Any Cantor set is an intersection of closed sets and is thus closed. Being a closed subset of a complete metric space it is also complete. Total boundedness is obvious, and thus the Hein-Borel theorem implies that any Cantor set is compact. Furthermore, having empty interior, any Cantor set has a closure (which is itself) with empty interior which means, by definition, that it is nowhere dense. Homeomorphism of arbitrary Cantor sets can be established by showing that they are all homeomorphic to some other set X. This can be established easily for $X = \prod_{n\in\mathbb{N}}\{0,1\}$ with the discrete topology (see the proof of next proposition). ∎

2.2.0.6 Proposition

Measure of Cantor Sets: Let C_α be a Cantor set and C^f_β a fat Cantor set. Both C_α and C^f_β are isomorphic to $[0,1]$ and are thus uncountable. Furthermore, C_α has measure 0 and C_β has measure $1-\beta/(1-2\beta)$.

proof: We show isomorphism of C_α with $[0,1)$; the exact same argument applies to C^f_β. The standard proof is to build an injection from C_α to $[0,1)$ and using the obvious inclusion injection from $[0,1)$ to C_α invoke the Schröder–Bernstein theorem. Note that any $x \in C_\alpha$ must lie in all of the intermediate sets A_n. Further, each A_n is a finite collection of disjoint intervals. Therefore, we can assign to every $x \in C_\alpha$ a binary sequence $(\omega_n)_{n\in\mathbb{N}}$ with the rule that ω_n is the counting position of the interval in A_n that contains x modulo 2. Since the set of binary sequences is isomorphic to $[0,1)$ we have an injection from C_α to $[0,1)$ and the proof of isomorphism to $[0,1)$ is complete. The lengths of the sets follow from routine calculations: For standard Cantor sets note that for all $n \in \mathbb{N}$, the set A_n has length $(1-\alpha)^{n-1}$ which converges to 0 as $n \to \infty$. For fat Cantor sets note that for all $n \in \mathbb{N}$, the set $B_n \setminus B_{n+1}$ has length $2^n\beta^{n-1}$ and thus $B_1 \setminus B_\infty$ has length $\sum_n 2^n\beta^{n-1} = \beta/(1-2\beta)$. ∎

2.2.0.7 Remarks

1) Here we are using the term “measure” loosely (but correctly). The only informal part of our usage is that we are *assuming* that there is a way to assign lengths to subsets of the real line such that the length of a countable union of disjoint sets is precisely the sum of their lengths (which is true in this case, but not obvious).

2) The last two propositions capture deep concepts embodied by Cantor sets: a Cantor set of either kind provides an example for a set which is compact despite being nowhere dense and totally disconnected. A fat Cantor set additionally provides an example for a set which has positive measure despite being nowhere dense (i.e containing no intervals).

3) The $C_\alpha \to [0,1)$ injection provided above can be stated somewhat differently: let $\alpha = 1/(2k+1)$ for some natural $k \in \mathbb{N}$. Then at each step of construction of C_α the set A_n is precisely the set of $[0,1)$ numbers for which the n-th digit of the base $2k+1$ representation is anything but k. It follows that C_α is the set of $[0,1)$ reals whose base $2k+1$ representation does not contain k. We can then shift down all digits larger than k to get the representation of another number in base $2k$ and this correspondance is injective. For instance, the middle thirds Cantor set consists of all points in $[0,1)$ whose ternary representation consists entirely of 0’s and 2’s. Swapping 2’s for 1’s gives an injection from the ternary reprsentation of $C_{\frac{1}{3}}$ to the binary representation of $[0,1)$.

2.3 Riemann and Darboux integration

These are two equivalent forms of integration that we refer to in building stochastic integrals.

2.3.0.1 Definition

For any $f : [a, b] \to \mathbb{R}$, the *upper Riemann integral* of f is

$$\overline{\int_a^b} f dx = \lim_{|P|\to 0} \sum_k \sup_{\tau\in[t_k,t_{k+1}]} f(\tau)(t_{k+1} - t_k)$$

and the *lower Riemann integral* of f is:

$$\underline{\int_a^b} f dx = \lim_{|P|\to 0} \sum_k \inf_{\tau\in[t_k,t_{k+1}]} f(\tau)(t_{k+1} - t_k)$$

where the limit is over partitions $P : a = t_1 < t_2 < \ldots < t_n = b$ with mesh size $|P| = \max_k(t_{k+1} - t_k)$. We say that f is *Riemann-integrable* if the two integrals coincide.

2.3.0.2 Definition

Let $f : [a, b] \to \mathbb{R}$. If f is piecewise constant, namely $f(\tau) = f_k$ for all $\tau \in [t_k, t_{k+1}]$, then define:

$$\int_a^b f dx = \sum_k f_k(t_{k+1} - t_k)$$

For arbitrary f define the *upper Darboux integral* of f to be:

$$\overline{\int_a^b} f dx = \sup_{g\geq f g \text{ piecewise const.}} \int_a^b g dx$$

and the *lower Darboux integral* of f is:

$$\underline{\int_a^b} f dx = \sup_{g\leq f g \text{ piecewise const.}} \int_a^b g dx$$

We say that f is *Darboux-integrable* if the two integrals coincide.

2.3.0.3 Proposition

Equivalence of Darboux and Riemann integration: Any real function is Riemann-integrable iff it is Darboux-integrable in which case the two integrals coincide in value.

2.3.0.4 Proposition

Lebesgue's Riemann-integrability criterion: Any real function is Riemann-integrable iff it is continuous almost everywhere.

2.3.0.5 Remarks

1) Note that this has nothing to do with Lebesgue integration and does not need the full machinery of measure theory. Negiligible sets can be defined without invoking measure theory (cf. Negligible sets). **2)** This theorem establishes the following surprising results: the indicator function of the Cantor set is Riemann-integrable but that of the fat Cantor set is not. This is because 1_C is discontinuous exactly at every point of C (since C is nowhere dense) but C is negligible. On the other hand 1_{C^f}, being also discontinuous exactly at every point of C^f, has a non-negligible set of discontinuities and is thus not Riemann-integrable.

2.4 Total Variation and Riemann-Stieltjes integration

Riemann-Stieltjes integration serves as an intuitive bridge between Riemann and Lebesgue integration. Furthermore, the notion of total variation used to prove the existence of the former integral is also central in the study of continuous time stochastic processes.

2.4.0.1 Definition

Let $f : \mathbb{R} \to \mathbb{R}$ be an arbitrary function. The *total variation* of f over an interval $[a, b]$ is defined to be:

$$V(f) = \sup_P \sum_k |f(t_{k+1}) - f(t_k)|$$

where the supremum is over all partitions P of $[a, b]$: $a = t_1 < t_2 < \ldots < t_n = b$. We say that f has *bounded variation* if $V(f) < \infty$.

2.4.0.2 Proposition

Properties of bounded variation functions: **i)** If f has bounded variation, it is not necessarily continuous, e.g. take staircase functions. **ii)** If f is C-Lipschitz then $V(f) \leq C(b - a)$, i.e it has bounded variation. The converse is not true (by i). **iii)** If f is monotone over $[a, b]$ then its total variation is $|f(a) - f(b)|$. **iv)** If f has bounded variation, it can be written as the difference of two monotone functions.

2.4.0.3 Proposition

Riemann-Stieltjes Integral: Let f be continuous and g have bounded variation over $[a, b]$. The integral of f with respect to g exists, i.e approximating sums converge as the partition mesh size tends to zero:

$$\int_a^b f dg = \lim_{|P| \to 0} \sum_k f(\tau_k) \left[g(t_{k+1} - g(t_k)\right]$$

where the mesh size of P is $|P| = \max_k(t_{k+1} - t_k)$ and the choice of $\tau_k \in [t_k, t_{k+1}]$ is arbitrary.

proof: The simpler case of montone g can be established routinely by constructing lower and upper *Darboux sums*. For the arbitrary bounded variation g use **iv** of the previous proposition to write the integral as the difference of two integrals with respect to monotone g. ■

2.5 Algebraic structures

The following is a list of algebraic structures of which only their definition and elementary properties are of interest.

2.5.1 Group-like structures

The main objects are semigroups, monoids and groups which have increasingly more structure. All of the below structures have *commutative* (or *abelian*) variants which simply demand that the group operator be commutative.

A *semigroup* is a pair $(G, \cdot)$ such that G is closed under $\cdot$ and $\cdot$ is associative, i.e $x \cdot (y \cdot z) = (x \cdot y) \cdot z$. An example is $(\mathbb{N}, +)$. A *monoid* is a semigroup with an identity element 1 satisfying $1 \cdot x = x \cdot 1 = x$. An example is the set of $n \times n$ matrices with usual matrix multiplication or the set of character strings with concatenation. It can be shown that identity elements are necessarily unique and that inverse elements, i.e x^{-1} such that $x \cdot x^{-1} = x^{-1} \cdot x = 1$, are unique if they exist. A *group* is a monoid with guaranteed inverses everywhere. An example is $(\mathbb{Z}, +)$.

2.5.1.1 Ring-like structures

A *semiring* is a triplet $(R, +, \cdot)$ such that $(R, +)$ and $(R, \cdot)$ are both semigroups [14] and $\cdot$ distributes from left and right over $+$. An example is $(\mathscr{P}(X), \cup, \cap)$ with $\emptyset$ as 0 and Ω as 1, see below. A *ring* is a semiring such that $(R, +)$ guarantees inverses, i.e it is a commutative group. A *commutative ring* is a ring such that $(R, \cdot)$ too is a commutative monoid. A *division ring* is a ring which guarantees inverses under both $+$ and $\cdot$ (excluding multiplicative inverse for 0), i.e both $(R, +)$ and $(R, \cdot)$ are commutative groups. An example is the set of invertible $n \times n$ matrices and the zero matrix together with matrix composisition and addition. Commutative division rings are *fields*, e.g. $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$.

2.6 Algebras and sequences of sets

2.6.1 Algebras of Sets

Much of measure theory is concerned with containing pathological cases where elementary set operations are applied infinitely many times (and as we saw drawing the line between countably many and uncountably many operations leads to desirable behavior while coverying most objects of interest). Here we present the simples algebraic structures over sets which are contained with only finite applications of set operations which is the point of contact to boolean logic. This connection is intuitively captured by the correspondance between $\cap, \cup$ of set theory and $\wedge, \vee$ of logic in various formulae.

[14] Some authors use a much stronger definition for semirings which is no longer analogous to semigroups: $(R, +)$ is a commutative monoid with identity 0, and $(R, \cdot)$ is a monoid with identity 1, i.e identity elements and commutativity of + are added to the axioms. In addition, this definition requires that $0 \cdot x = x \cdot 0 = 0$ which can be *proved* for rings but is required as axiom for this stronger definition of semirings. In the terminology we use, an *additive semiring* restores commutativity of +, a *rig* ("ring without *n*egative element") restores the latter and the existence of 0, and a *rig with unity* restores additionally the existence of 1. The latter would be equivalent to the stronger definition of a semiring. In either case, a *commutative semiring* would be one whose multiplication commutes.

2.6.1.1 Definition

Let Ω be any nonempty set and take any $S \subset \mathscr{P}(x)$ satisfying $\emptyset \in S$. Then $(S, \cup, \cap)$ is a commutative *rig of subsets* (not a typo, cf. footnote) over Ω, i.e commutative and additive semigroup with a 0 or equivalently a ring without a negative element, if S is closed under union and intersection. Additionally, $(S, \Delta, \cap)$ is a commutative *ring of subsets* if S is closed under union and symmetric set-theoretic difference Δ, i.e $x \Delta y = (x \setminus y) \cup (y \setminus x)$.

2.6.1.2 Remarks

1) Note the $\cap$ and $\cup$ in the defintion of ring of subsets: S is required to be closed under Δ and $\cup$ for $(S, \Delta, \cap)$ to be a ring of subsets. Further, note that the additive inverse of any element $x \in S$ is itself since $x \Delta x = \emptyset$. **2)** Rigs of subsets admit neither additive inverses (i.e $-x$ such that $x \cup -x = \emptyset$) nor multiplicative inverses (i.e x^{-1} such that $x \cap x^{-1} = \Omega$). However, the set complement operation is an *inverse-like* alternative which can endow them with more structure. These include: De Morgan laws, i.e $(x \cap y)^c = x^c \cup y^c$ and vice versa, inversion of partial order, i.e $x \subset y \Rightarrow y^c \subset x^c$, and double inverse, i.e $(x^c)^c = x$. **3)** There are additional set-theoretic properties that can impose further structure than those summarized in the ring-like structure. These include *idempotence*, i.e $x + x = x \cdot x = x$, *domination*, i.e $x + 1 = 1$ and $x \cdot 0 = 0$, and *absorption*, i.e $x \cdot (x + y) = x + (x \cdot y) = x$ [15]. These properties in addition to those of the complement (see above) are the link between ring-like structures of subsets and ring-like Boolean structures of logic where $\cap$ corresponds to $\wedge$ (meet) and $\cup$ corresponds to $\vee$ (join). This correspondence is characterized by *Stone's representation theorem*. **4)** There is a related but not identical notion of *semi-ring of sets* in measure theory which does not coincide with the algebraic semiring definition (of either kind, cf. footnote above). We have used neither those objects nor these defined here.

2.6.2 Sequences of Sets

In analogy to corresponding notions in analysis, and in light of the partial order imposed by $\subset$, we define corresponding concepts for limit sets of sequences of sets.

2.6.2.1 Definition

Let $(A_n)_{n \in \mathbb{N}}$ be a sequence of sets in Ω. Define $B_n = \bigcup_{k \geq n} A_k$, i.e the supremum of the n-tail, and $C_n = \bigcap_{k \geq n} A_k$, i.e the infimum of the n-tail. The *limit superior* and *limit inferior* of (A_n) are defined as:

$$\limsup_n A_n = \bigcap_n B_n = \bigcap_n \bigcup_{k \geq n} A_k$$

$$\liminf_n A_n = \bigcup_n C_n = \bigcup_n \bigcap_{k \geq n} A_k$$

and if the two sets coincide $\lim_n A_n$ is well defined.

2.6.2.2 Proposition

Properties of $\liminf$ *and* $\limsup$ *for sequences of sets*:

[15] Using set-theoretic operations these read: $x \cup x = x \cap x = x$, $x \cup \Omega = \Omega$, $x \cap \emptyset = \emptyset$, and $x \cap (x \cup y) = x \cup (x \cap y) = x$.

i) $B_n \downarrow \limsup_n A_n$ and $C_n \uparrow \liminf_n A_n$.
ii) $\liminf_n A_n \subset \limsup_n A_n$.
iii) $\omega \in \limsup_n A_n$ iff $\omega \in A_n$ for infinitely many n.
iv) $\omega \in \liminf_n A_n$ iff $\omega \in A_n$ for all but finitely many n.
v) if $A_n \in \mathscr{F}$ for all n and some σ-algebra $\mathscr{F}$ then $\limsup_n A_n \in \mathscr{F}$ and $\liminf_n A_n \in \mathscr{F}$.

References

[1] T. Tao, *An introduction to measure theory.* American Mathematical Society, 2011.

[2] P. Billingsley, *Probability and measure.* John Wiley & Sons, 2008.